



**Impact Assessment study on the list  
of High Value Datasets to be made  
available by the Member States  
under the Open Data Directive**

**Deloitte.**

## **Internal identification**

Contract number: SMART 2019/0025

Project number: 2020.5755

### **EUROPEAN COMMISSION**

Directorate-General for Communications Networks, Content and Technology

Directorate G — Data

Unit G.1 — Data Policy & Innovation

*Contact: [cnect-g1@ec.europa.eu](mailto:cnect-g1@ec.europa.eu)*

*European Commission*

*B-1049 Brussels*

# **Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive**

**EUROPE DIRECT is a service to help you find answers  
to your questions about the European Union**

Freephone number (\*):  
00 800 6 7 8 9 10 11

(\*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

**LEG  
AL  
NOTI  
CE**  
This

document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. The Commission does not guarantee the accuracy of the data included in this study. More information on the European Union is available on the Internet (<http://www.europa.eu>).

---

EN PDF

ISBN 978-92-76-25267-2

doi: 10.2759/493091

KK-02-20-017-EN-N

---

The European Commission is not liable for any consequence stemming from the reuse of this publication.

Luxembourg: Publications Office of the European Union, 2020

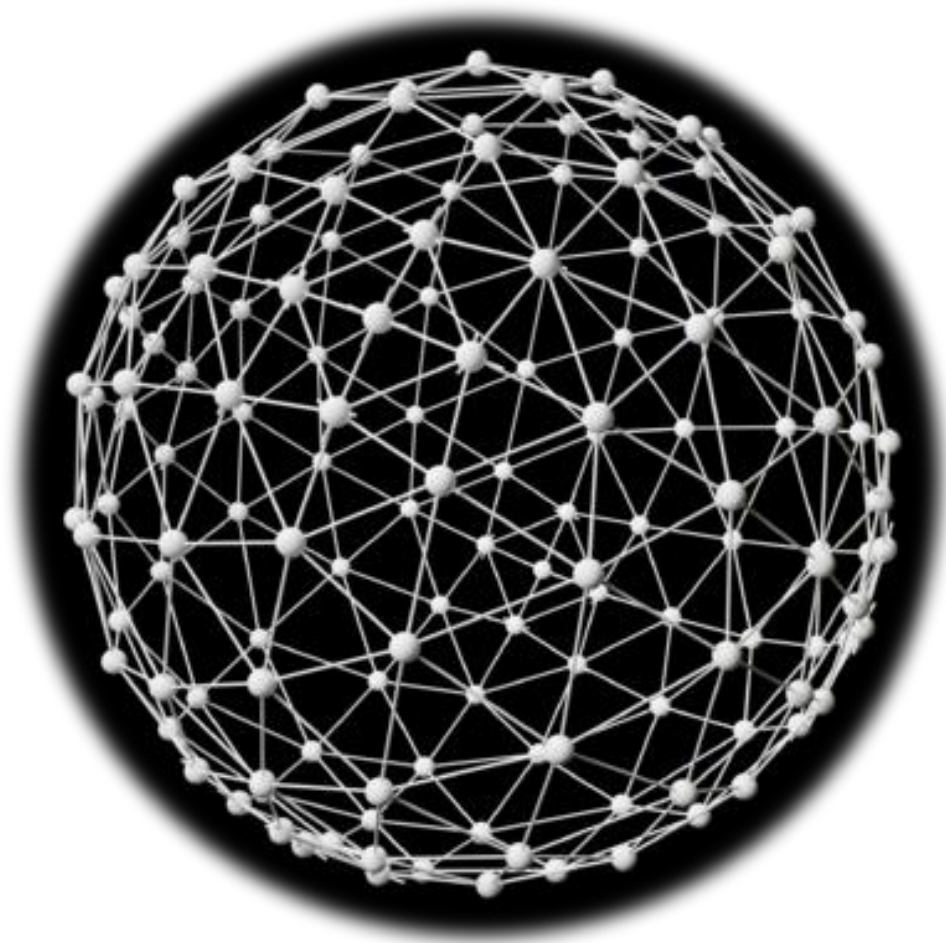
© European Union, 2020



The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

For any use or reproduction of elements that are not owned by the European Union, permission may need to be sought directly from the respective rightholders.





## **SMART 2019/0025 – Final Report**

Impact Assessment study on the list of High Value Datasets to be made available by the Member States under the Open Data Directive

# Table of Content

<b>1 INTRODUCTION .....</b>	<b>1</b>
1.1 Purpose and structure of the document.....	1
1.2 Context of the impact assessment study .....	2
<b>2 IDENTIFICATION OF HIGH-VALUE DATASETS IN SCOPE OF THE IMPACT ASSESSMENT STUDY.....</b>	<b>7</b>
2.1 Methodological framework for the identification of high-value datasets .....	7
2.2 Company and company ownership.....	16
2.3 Geospatial.....	33
2.4 Meteorological data .....	41
2.5 Earth observation and environment.....	54
2.6 Statistics.....	67
2.7 Mobility.....	74
2.8 Horizontal considerations from the Open Public Consultation .....	103
<b>3 MICRO-LEVEL ANALYSIS OF HIGH-VALUE DATASETS AND RECOMMENDED POLICY OPTIONS .....</b>	<b>109</b>
3.1 Company and company ownership.....	109
3.2 Geospatial.....	158
3.3 Meteorological data .....	210
3.4 Earth observation and environment.....	242
3.5 Statistics.....	274
3.6 Mobility.....	316
3.7 Horizontal considerations for the micro-level analysis .....	353
<b>4 MACRO-LEVEL IMPACTS OF POLICY OPTIONS FOR HIGH VALUE DATASETS.....</b>	<b>355</b>
4.1 Comparison of policy options.....	355

**4.2 Cost estimations of the policy options .....374**

**4.3 Modelling of the impacts of policy options .....381**

**4.4 Identification of policy intervention packages and their aggregated impacts ...409**

**ANNEXES .....415**

**Annex A – Used graphs for the analysis of the Open Public Consultation .....415**

## Table of Figures

Figure 1 – Overview of EU PSI re-use policy initiatives over the last 30 years.....3

Figure 2 - Ratio between price and number of reusers for Company House Datasets ..... 134

Figure 3 – Scale for the Rating of the Criteria of the MCA ..... 357

Figure 4 – Our method: Assumptions used and literature ..... 381

Figure 5 – Our method: Top-down approach ..... 383

Figure 6 – Baseline estimates..... 384

Figure 7 – Baseline estimates – market shares..... 384

Figure 8 – Direct economic impact per Policy Option by thematic area – company and company ownership ..... 387

Figure 9 - Direct economic impact - geospatial ..... 388

Figure 10 - Direct economic impact - meteorological data ..... 388

Figure 11 - Direct economic impact - earth observation and environment ..... 389

Figure 12 - Direct economic impact statistics ..... 389

Figure 13 - Direct economic impact - mobility ..... 390

Figure 14 – Overview of direct, indirect, and induced impacts ..... 390

Figure 15 – Indirect economic impact – Company and company ownership ..... 392

Figure 16 - Indirect economic impact - Geospatial ..... 392

Figure 17 - Indirect economic impact - Meteorological Data ..... 393

Figure 18 - Indirect economic impact - Earth observation and environment ..... 393

Figure 19 - Indirect economic impact - Statistics ..... 394

Figure 20 - Indirect economic impact - Mobility ..... 394

Figure 21 – Total economic impact – Company and company ownership..... 395

Figure 22 - Total economic impact - Geospatial ..... 396

Figure 23 - Total economic impact - Meteorological Data ..... 396

Figure 24 - Total economic impact - Earth observation and environment ..... 397

Figure 25 - Total economic impact - Statistics ..... 397

Figure 26 - Total economic impact - Mobility ..... 398

Figure 27 – Employment impact – Company and company ownership. .... 399

Figure 28 - Employment impact - geospatial ..... 399

Figure 29 - Employment impact - meteorological data..... 400

Figure 30 - Employment impact - earth observation and environment..... 400

Figure 31 - Employment impact - Statistics..... 401

Figure 32 - Employment impact - Mobility.....	401
Figure 33 – Governmental revenues impact – Company and company ownership .....	403
Figure 34 - Governmental revenues impact - Geospatial.....	403
Figure 35 - Governmental revenues impact - Meteorological data .....	404
Figure 36 - Governmental revenues impact - Earth observation and environment.....	404
Figure 37 - Governmental revenues impact - Statistics.....	405
Figure 38 - Governmental revenues impact - Mobility.....	405
Figure 39 – Enterprises number impact – Company and company ownership .....	406
Figure 40 - Enterprises number impact - Geospatial.....	407
Figure 41 - Enterprises number impact - Meteorological data .....	407
Figure 42 - Enterprises number impact - Earth observation and environment.....	408
Figure 43 - Enterprises number impact - Statistics.....	408
Figure 44 - Enterprises number impact - Mobility.....	409
Figure 45 – Economic Value of PSI/HVD per policy Package in m€ and compared to EU27 GDP .....	410
Figure 46 – Incremental impacts policy package 1 (lower intensity intervention) .....	412
Figure 47 – Incremental impacts policy package 2 (higher intensity intervention) .....	413
Figure 48 – Incremental impacts policy package 3 (mixed intensity intervention) .....	414

## Table of Tables

Table 1 – HVD Framework: macro characteristics and categories of value.....	9
Table 2 - Possible indicators for measuring innovation and AI .....	13
Table 3 – Possible indicators for measuring value for public services and public administrations .....	13
Table 4 – Possible indicators for measuring value for reuse.....	14
Table 5 – Possible indicators for measuring social value.....	15
Table 6 – Relevant EU legislation for company and company ownership .....	17
Table 7 – Initial lists of HVDs under company and company ownership thematic area .....	22
Table 8 - List of INSPIRE data themes.....	34
Table 9 – Geospatial thematic area – Datasets in scope.....	35
Table 10 - Legislative scope of meteorological legislation .....	41
Table 11 - Meteorological legislation .....	43
Table 12 - Earth observation and Environment legislation .....	56
Table 13 – Data themes and datasets covered by the ITS Directive and Delegated Acts .....	76
Table 14 – Data themes and datasets in scope of the analysis for PSI mobility thematic area.....	78
Table 15 - High-Value Datasets in the mobility thematic area, inland waterways .....	79
Table 16 – List of INSPIRE transport network features .....	82
Table 17 – Notices to Skippers and Inland ECDIS Systems, and their benefits.....	87
Table 18 – Assessment of value for inland waterway and river infrastructure data .....	90
Table 19 – Assessment of value for transport networks data .....	99
Table 20 – Yearly revenue linked to company information charging per country .....	118
Table 21 – Recommended measures for publication of company datasets .....	120
Table 22 – Cost drivers for company datasets .....	123
Table 23 – Recommended measure for publication for company datasets .....	125
Table 24 – Expected implementation costs of the recommended measures .....	127
Table 25 – Benefits assessment (Finland) .....	140
Table 26 – Benefits and costs (Finland) .....	141

Table 27 – Benefits assessment (Slovenia) .....	141
Table 28 – Benefits and costs assessment (Slovenia) .....	141
Table 29 – Company and company ownership – Scope of the low intensity intervention .....	148
Table 30 – Recommended measures for publication of company datasets- Low intensity intervention ...	150
Table 31 – Company and company ownership – Scope of the higher intensity intervention .....	154
Table 32 - Recommended measures for publication of company datasets- Higher intensity intervention	156
Table 33 - Current modalities for provision of geospatial data .....	160
Table 34 – Recommended measures for publication for geospatial data .....	194
Table 35 – Expected cost categories for geospatial data .....	197
Table 36 – Main budgetary implication of the recommended measures for publication .....	197
Table 37 – Magnitude of costs .....	198
Table 38 – Geospatial - Scope of the lower intensity intervention .....	206
Table 39 – Geospatial - Scope of the high intensity intervention .....	209
Table 40 – Statistics – Datasets overview .....	311
Table 41 - Statistics - Scope of the low intensity intervention.....	313
Table 42 - Statistics - Scope of the high intensity intervention .....	315
Table 43 – Current modalities for provision of inland waterway and river infrastructure data .....	318
Table 44 – Current modalities for provision of transport networks data (1/2) .....	323
Table 45 – Current modalities for provision of transport networks data (2/2) .....	328
Table 46 – Links to national access points to geospatial data and transport networks datasets .....	329
Table 47 – Recommended modalities for publication of inland waterway infrastructure data .....	331
Table 48 – Recommended modalities for publication of transport networks data .....	333
Table 49 – Estimation of costs related to fairway and infrastructure data provision under VisuRIS COMEX .....	336
Table 50 – Indicative comparison of impact of VisuRIS COMEX operational costs on Member States’ annual budgets.....	337
Table 51 – Current costs borne by DB Netz AG for the provision of rail transport networks data .....	338
Table 52 – WMS requests and downloads of transport networks and infrastructure datasets from Finnish transport open data portal (2018) .....	341
Table 53 – high-value datasets in the mobility thematic area, inland waterways .....	345
Table 54 – Recommended modalities for publication of inland waterway infrastructure data – Lower intensity intervention .....	346
Table 55 – high-value datasets in the mobility thematic area, INSPIRE transport network features .....	348
Table 56 – Recommended modalities for publication of INSPIRE transport networks data – Lower intensity intervention .....	349
Table 57 – Recommended modalities for publication of other national transport networks data – High intensity intervention .....	351
Table 58 – Input Matrix: Weight, direction and performance value of the assessment criteria .....	357
Table 59 – Explanation of the performance values: Company and company ownership .....	359
Table 60 – Explanation of the performance values: Geospatial .....	360
Table 61 – Explanation of the performance values: Meteorological data .....	363
Table 62 – Explanation of the performance values: Earth observation and environment .....	365
Table 63 – Explanation of the performance values: Statistics .....	366
Table 64 – Explanation of the performance values: Mobility .....	369
Table 65 – Input matrix: performance values for the four intervention options .....	371
Table 66 – Outranking matrix.....	372

Table 67 – Possible Permutations of the policy options .....	373
Table 68 – Possible Policy Pairings .....	373
Table 69 – Outcome of MCA: preferred policy options .....	374
Table 70 - Revenues for selected Member States in the thematic area company and company ownership .....	376
Table 71 - Multipliers .....	391
Table 72 – Policy Packages composition .....	410
Table 73 - Economic Value of PSI/HVD per policy package in mEUR and compared to EU27 GDP .....	411

## Table of Boxes

Box 1 – Validation workshop results: company and company ownership, overall appreciation of policy intervention options .....	147
Box 2 – Validation workshop results: geospatial, overall appreciation of policy intervention options .....	204
Box 3 – Validation workshop results: meteorological data, overall appreciation of policy intervention options .....	236
Box 4 – Validation workshop results: earth observation and environment, overall appreciation of policy intervention options .....	270
Box 5 – Validation workshop results: statistics, overall appreciation of policy intervention options .....	312
Box 6 – Validation workshop results: mobility, overall appreciation of policy intervention options .....	344



# 1 Introduction

## 1.1 Purpose and structure of the document

This deliverable represents the Final Study Report of the “*Impact Assessment study on the list of high-value datasets to be made available by the Member States under the PSI Directive*”. The overarching goal of this deliverable is to present the overall findings and final results of the study conducted between January and August 2020. In essence, this Final Study Report integrates the results of five interim deliverables: D1, D2, D3, D4, D5.

This deliverable is structured as follows:

- Chapter 1 provides a short introduction to the document, including an overview of the policy context and the objectives and scope of the study;
- Chapter 2 presents the findings related to the identification of high-value datasets in scope of the study, including the methodological framework developed and applied within the six thematic areas to assess the value of these datasets (as per D1, D2);
- Chapter 3 presents the findings of the micro-level analysis of the high-value datasets in scope of the study, as well as the policy intervention options proposed for each thematic area on this basis (as per D3, D4)
- Chapter 4 presents the findings of the macro-level analysis of the policy intervention options identified for the high-value datasets in the six thematic areas, and provides the grounds for the selection of the preferred policy intervention options and resulting packages (as per D4,D5).

This report also includes one Annex:

- Annex A, including the graphs pertaining to the analysis of the feedback collected through the Open Public Consultation (OPC).



## 1.2 Context of the impact assessment study

This section provides an overview of the overall context of the study including the policy background, its objectives and its scope.

### 1.2.1 Policy background

**European Public Sector Information (PSI) policies celebrate thirty years in 2019.** Thirty years ago, in 1989, the European Commission produced the first document concerning public sector information in Europe: the Guidelines for improving the synergy between public and private sectors in the information market<sup>1</sup>. This first policy initiative described “already the potential resulting from the combination of information from a variety of government sources in view of producing and distributing information products oriented to the needs of the market. It also called for “a positive initiative [...] from governments, to encourage the use and exploitation of public sector data and information” so that an information industry could be developed and strengthened”<sup>2</sup>. Three decades later, the message brought by these first guidelines remain relevant and, today, public sector information are widely recognised as an “economic asset and an important resource for economic growth and competitiveness”<sup>3</sup>. Furthermore, this role of PSI as economic motor has been strengthened by the tremendous growth of the data economy and the take up of Artificial Intelligence (AI) technologies and applications.

Given the ever increasing important of this topic, the last thirty years have seen a wide array of policy initiatives being developed at the national, European and international level. The graph below gives an historic overview of the main European policy initiatives - policy documents, legislation, studies and overarching programs - surrounding PSI re-use since the 1989 Guidelines<sup>4</sup>.

---

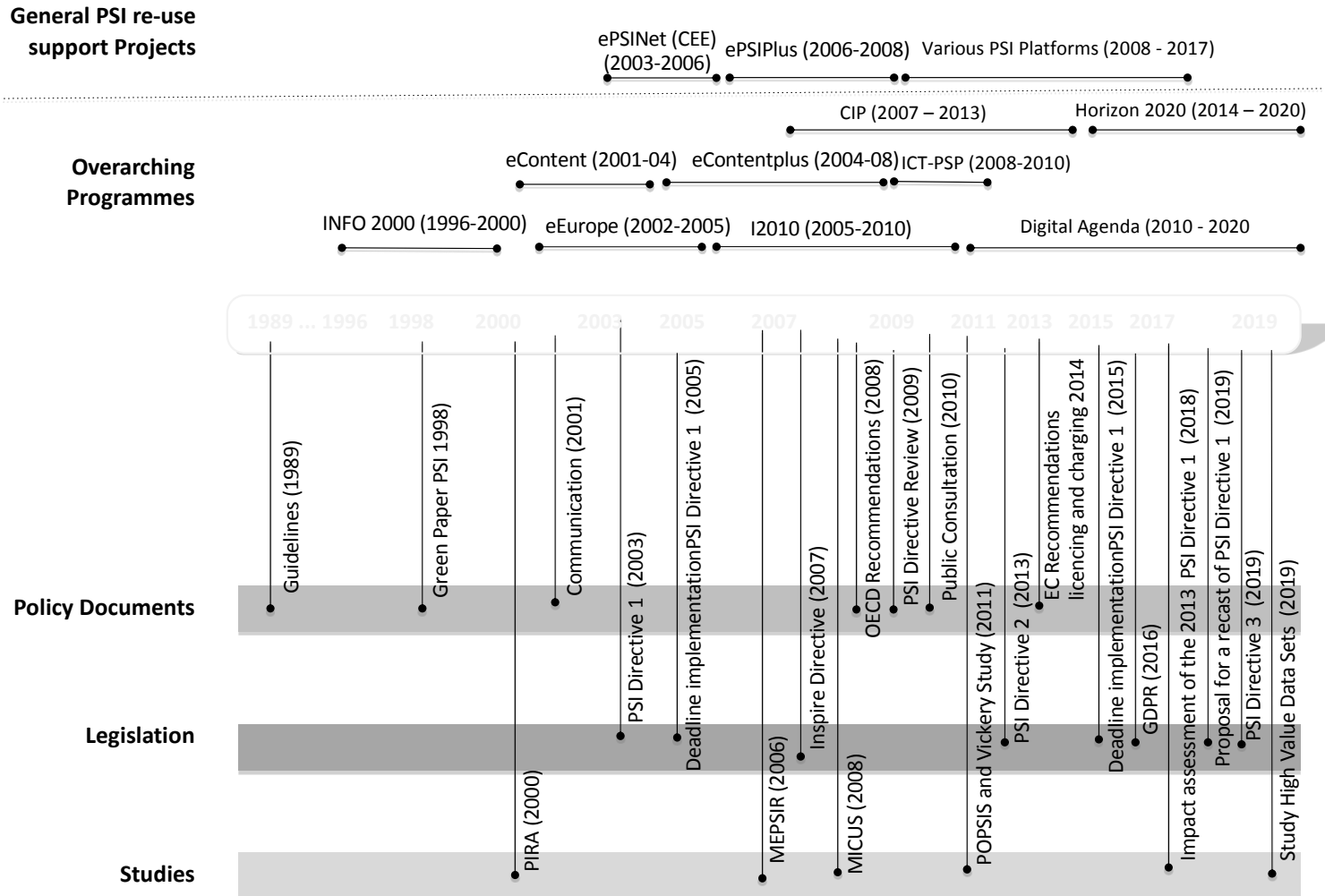
<sup>1</sup> Guidelines for improving the synergy between public and private sectors in the information market, <https://ec.europa.eu/digital-single-market/en/news/guidelines-improving-synergy-between-public-and-private-sectors-information-market>

<sup>2</sup> Guidelines for improving the synergy between public and private sectors in the information market, <https://ec.europa.eu/digital-single-market/en/news/guidelines-improving-synergy-between-public-and-private-sectors-information-market>

<sup>3</sup> Commission Staff Working Document Evaluation, Accompanying the document, Proposal for a Directive of the European Parliament and of the Council on the re-use of public sector information, {COM(2018) 234 final} - {SWD(2018) 129 final} <https://eur-lex.europa.eu/legal-content/EN/PIN/?uri=SWD:2018:129:FIN>

<sup>4</sup> Guidelines for improving the synergy between public and private sectors in the information market, <https://ec.europa.eu/digital-single-market/en/news/guidelines-improving-synergy-between-public-and-private-sectors-information-market>

Figure 1 – Overview of EU PSI re-use policy initiatives over the last 30 years



Source: Marc De Vries, The Green Land, 2019

Paramount in this chain of events have been the successive adoption of the three PSI Directives. The first version of this text, adopted in 2003<sup>5</sup> aimed to facilitate the re-use of PSI throughout the Union by harmonising the basic conditions for making PSI available to re-users, to foster Community-wide products and services based on PSI, and to avoid distortions of competition. The huge economic impact it had was demonstrated by a number of subsequent studies (in particular, at EU level, MEPSIR, MICUS<sup>6</sup>, Vickery<sup>7</sup> and POPSIS<sup>8</sup>).

Appreciating the further potential due to rapid technological changes and being confronted with the inherent limitations of the first PSI Directive, the Commission launched the initiative for a revised PSI Directive, which was adopted in 2013.<sup>9</sup> In essence, this second PSI Directive: (1) extended the scope to cultural PSI, while at the same time creating a sort of special re-use regime for this sector, (2) linked the right of re-to access rights, (3) further limited the room for charging inter alia by imposing transparency obligations and (4) introduced a set of practical measures (machine readable formats, central repositories) to facilitate the discovery and re-use of public sector information.

Again, being confronted with rapid and unprecedented technological developments, the extensive review (under Article 13 of the 2013 Directive) that was undertaken by the EC in 2018, concluded that the Directive still contributed to the achievement of its main policy objectives, but a number of issues were affecting the full exploitation of the potential of PSI for the European economy and society. Accordingly, in April 2018, so five years after the adoption of the second PSI Directive, the April 2018, the EC submitted a proposal for a recast of the Directive 2003/98/EC<sup>10</sup>. Recently, this proposal was adopted by both Parliament and the Council<sup>11</sup>. Briefly put, the 2019 PSI Directive aims to: (1) reduce market entry barriers by lowering charges for the re-use (2) increase the availability of data by bringing new types of public and publicly funded data into the scope of the Directive (3) minimise the risk of excessive first-mover advantage by requiring a more transparent process for the establishment of public-private arrangements and (4) increase business opportunities by encouraging the publication of dynamic data and the uptake of APIs.

Underlining the importance of concrete harmonization at European level, Article 14 of the new Directive empowers the Commission to adopt implementing acts laying down a list of specific high-value datasets (HVDs) belonging to six thematic categories to be made available for re-use for free, machine-readable, provided via APIs and, where relevant, as a bulk download. Furthermore, these implementing acts may also specify the arrangements for the publication and re-use of high-value datasets, which shall be compatible with open standard licenses, including terms applicable to re-use, formats of data and metadata and technical arrangements for dissemination.

This short policy anthology clearly demonstrates not only the paramount role the Commission has played in taking this dossier further, it also shows the logic of the next steps to be taken: identification and subsequent freeing of high-value datasets across Europe will likely spark a next chain of benefits. This

---

<sup>5</sup> Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32003L0098>

<sup>6</sup> Assessment of the Re-use of Public Sector Information in the Geographical Information, Meteorological Information and Legal Information sectors, MICUS, December 2008.

<sup>7</sup> Graham Vickery, Review of recent studies on PSI reuse and related market developments, 2011, <https://ec.europa.eu/digital-single-market/en/news/review-recent-studies-psi-reuse-and-related-market-developments>

<sup>8</sup> Pricing of Public Sector Information Study (POPSIS) - Models of supply and charging for public sector information (ABC) - final report, 2011, <https://ec.europa.eu/digital-single-market/en/news/pricing-public-sector-information-study-popsis-models-supply-and-charging-public-sector>

<sup>9</sup> Directive 2013/37/EU of the European Parliament and of the Council and amending Directive 2003/98/EC on the re-use of public sector information, <https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:32013L0037>

<sup>10</sup> <https://ec.europa.eu/digital-single-market/en/public-sector-information-psi-directive-open-data-directive>

<sup>11</sup> <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>

notion turns this study into an important piece of the ongoing PSI re-use puzzle and pivotal piece of work for DG CNECT.

### 1.2.2 Objectives

In line with the policy background summarised above, the objective of the study is to support the impact assessment process underpinning the adoption of implementing acts and *“to define concrete high-value datasets that fall under the thematic categories included in the Annex I of the revised Directive, based on an iterative process involving a number of cycles, by which an initial broad range of datasets would be narrowed down”*.

This assignment finds its origin in the revised PSI Directive itself<sup>12</sup>, which, in its Article 14<sup>13</sup>, mandates the European Commission to conduct an impact assessment for identifying high-value datasets falling under the list of thematic categories included in Annex I and that should be covered by the Directive.

This assignment is hence aimed at supporting the impact assessment process of the European Commission by providing evidence for this analysis. Besides this, the study team also assisted the European Commission with the organisation of the public consultation activities (public consultation and public hearing), which are also requested by the revised Directive<sup>14</sup> and by the Better Regulation Guidelines<sup>15</sup>.

### 1.2.3 Scope

The revised PSI Directive lists six thematic areas for which high-value datasets must be identified:

- Geospatial;
- Earth Observation and Environment;
- Meteorological;
- Statistical;
- Companies and company ownership; and,
- Mobility.

For each thematic area, the exact scope of the analysis, including relevant EU legislation and datasets pertaining thereto, are discussed in sections 2.2 to 2.7. Within the remits of each thematic area’s scope, the study analyses from micro to macro-level the impacts of a potential inclusion of given datasets as high-value datasets under the revised PSI Directive for the EU27.

It should be noted that:

- The study is dependent to a very large extent on the available data and information from data re-users and holders. Especially concerning the latter, data provision has been scarce, resulting therefore to findings being mostly anecdotic examples of possible effects in given Member States, rather than a fully-fledged assessment scaled-up at EU-27 level.

---

<sup>12</sup> Directive 2019/0024 of the European Parliament and the Council, of 20 June 2019, on open data and re-use of public sector information (recast), <https://eur-lex.europa.eu/eli/dir/2019/1024/oj>

<sup>13</sup> Directive 2019/0024 of the European Parliament and the Council, of 20 June 2019, on open data and re-use of public sector information (recast), <https://eur-lex.europa.eu/eli/dir/2019/1024/oj>

<sup>14</sup> Article 14 of the Directive establishes that: *“for the purpose of identifying such specific datasets, the Commission shall carry out appropriate consultations, including at expert level”*. Directive 2019/0024 of the European Parliament and the Council, of 20 June 2019, on open data and re-use of public sector information (recast), <https://eur-lex.europa.eu/eli/dir/2019/1024/oj>

<sup>15</sup> Commission Staff Working Document, Better Regulation Guidelines, Brussels, 7 July 2017, <https://ec.europa.eu/info/sites/info/files/better-regulation-guidelines.pdf>

- The study does not seek to provide recommendations on technical aspects regarding the implementation of the proposed policy intervention options, e.g. on the type of APIs that should be considered. As such, recommendations are of general nature and aim to be technology-neutral.
- The study does not ambition to provide indications in terms of timing and sequencing of roll-out and/or implementation of the proposed policy intervention options, should these be eventually adopted by the Commission.

# 2 Identification of High-value Datasets in Scope of the Impact Assessment Study

## 2.1 Methodological framework for the identification of high-value datasets

Establishing a framework for the definition of HVDs was paramount for the successful completion of the study as this has provided the study team with a means to identify potential HVDs to consider in the scope of the analysis, and discuss those with stakeholders based on commonly understood value drivers and criteria. A literature review was conducted to collate categories of value related to the six macro characteristics of potential value derived from open data and described in the Open Data/Public Sector Information Directive<sup>16</sup>: economic benefits; environmental benefits; generation of innovative services and innovation (innovation and artificial intelligence); reuse; and the improving, strengthening, and supporting of public authorities in carrying out their missions (public services and public administration, social).<sup>17</sup>

The review carried out by the study team generated 32 categories of value, supported by 126 quantitative and qualitative indicators, within the six characteristics, though it should be noted that this methodology does not require this or other studies to cover all indicators and produce exact value figures for the relevant datasets.

The realisation of value with open data can be considered through two lenses: informational and economic.<sup>18</sup> The former is about the information content of a dataset, with its generality - how many decisions it is useful for - affected by its quality; sensitivity; interoperability and linkability; excludability; and accessibility.<sup>19</sup> Economic characteristics include whether a dataset has diminishing or increasing marginal returns; externalities; optionality; consequences; and costs.<sup>20</sup> The interaction of these features is what makes identifying high-value a complex task.<sup>21</sup>

To clarify the search for value and indicators of it, the definition of value within the given categories was guided by the directive and supported by open data literature. This meant the adoption of the following definitions:

- **Climate change and environment:** open data exploited for understanding and improving environmental conditions and addressing climate change.<sup>22</sup>

---

<sup>16</sup> See Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>

<sup>17</sup> See Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>

<sup>18</sup> Bennett Institute for Public Policy and Open Data Institute (forthcoming) The valuation of data: policy implications

<sup>19</sup> Bennett Institute for Public Policy and Open Data Institute (forthcoming) The valuation of data: policy implications

<sup>20</sup> Bennett Institute for Public Policy and Open Data Institute (forthcoming) The valuation of data: policy implications

<sup>21</sup> European Data Portal (2020) Analytical report 15: high-value datasets: understanding the perspective of data providers, [https://www.europeandataportal.eu/sites/default/files/analytical\\_report\\_15\\_high\\_value\\_datasets.pdf](https://www.europeandataportal.eu/sites/default/files/analytical_report_15_high_value_datasets.pdf)

<sup>22</sup> Blended from Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>

- **Economic:** national income attributable to industries and activities that are based on the exploitation of open data - the value added of open data with respect to the economy as a whole and private sector expenditure on public sector information - and the creation of quality and decent jobs<sup>23</sup>.
- **Innovation and Artificial Intelligence (AI):** new services and applications, specifically including those related to algorithmic decision-making, created through the use of open data in the private sector - in particular by small and medium-size enterprises - or the public sector.<sup>24</sup>
- **Public services and public administration:** open data exploited for the access, efficiency, and quality improvements of public service delivery.<sup>25</sup>
- **Re-use:** open data exploited by intermediaries for a high number of users, with potential for being combined with other datasets, and potential beneficiaries of value-added services and applications based on these datasets.<sup>26</sup>
- **Social:** open data exploited for improving transparency and accountability and creating other important benefits for society.<sup>27</sup>

For each of these macro characteristics, the team identified a number of value drivers and related indicators which guided the discussion and data collection for the assignment and notably for the identification and characterisation of high-value datasets.

The table below provides an overview of the six macro characteristics and the corresponding categories of value identified in the literature. The full framework is described in the next sub-sections.

---

[lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN](https://www.europeandataportal.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN); European Data Portal (2018) The importance of open environment data on the European Data Portal, <https://www.europeandataportal.eu/en/highlights/importance-open-environment-data-european-data-portal>

<sup>23</sup> Blended from Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>; European Data Portal (2018) The importance of open environment data on the European Data Portal,

<sup>24</sup> Blended from Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>; Deloitte (2012) Open data: driving growth, ingenuity and innovation, <https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf>; Open Data Institute (2018) Using open data to deliver public services, <https://theodi.org/article/using-open-data-for-public-services-report-2/>

<sup>25</sup> Blended from Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>; European Data Portal (2012) Analytical report 9: the economic benefits of open data, [https://www.europeandataportal.eu/sites/default/files/analytical\\_report\\_n9\\_economic\\_benefits\\_of\\_open\\_data.pdf](https://www.europeandataportal.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf); Open Data Institute (2018) Using open data to deliver public services, <https://theodi.org/article/using-open-data-for-public-services-report-2/>

<sup>26</sup> Blended from Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>; European Commission (2011) Review of recent studies on PSI reuse and related market developments, <https://ec.europa.eu/digital-single-market/en/news/review-recent-studies-psi-reuse-and-related-market-developments>

<sup>27</sup> Blended from Official Journal of the European Union (2019) Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024&from=EN>; European Data Portal (2012) Analytical report 9: the economic benefits of open data, [https://www.europeandataportal.eu/sites/default/files/analytical\\_report\\_n9\\_economic\\_benefits\\_of\\_open\\_data.pdf](https://www.europeandataportal.eu/sites/default/files/analytical_report_n9_economic_benefits_of_open_data.pdf)

Table 1 – HVD Framework: macro characteristics and categories of value

<b>Summary value framework</b>	
<b>Macro characteristics</b>	<b>Macro characteristics description, and value categories</b>
<b>Climate change and environment</b>	<b>Definition: Open data exploited for understanding and improving environmental conditions and addressing climate change</b>
<i>Categories of value</i>	Citizen engagement in addressing climate change Climate change Energy management and efficiency Environment management
<b>Economic</b>	<b>Definition: National income attributable to industries and activities that are based on the exploitation of open data - the value added of open data with respect to the economy as a whole and private sector expenditure on public sector information - and the creation of quality and decent jobs</b>
<i>Categories of value</i>	Competition Consumer benefits Economic output Economy monitoring Employment International competitiveness Product market dynamism Productivity and commercialisation Public-private coordination Trust and transaction costs
<b>Innovation &amp; AI</b>	<b>Definition: New services and applications, specifically including those related to algorithmic decision-making, created through the use of open data in the private sector - in particular by small and medium-size enterprises - or the public sector</b>
<i>Categories of value</i>	Citizen innovation Entrepreneurialism and private sector innovation Public sector innovation Public-private coordination
<b>Public services and public administration</b>	<b>Definition: Open data exploited for the access, efficiency, and quality improvements of public service delivery</b>
<i>Categories of value</i>	Access to public services Public administration transparency, accountability & engagement



## Summary value framework

Macro characteristics	Macro characteristics description, and value categories
	Public sector procurement
	Public sector revenue
	Public services management
	Public services performance
<b>Re-use</b>	<b>Definition: Open data exploited by intermediaries for a high number of users, with potential for being combined with other datasets, and potential beneficiaries of value-added services and applications based on these datasets</b>
<i>Categories of value</i>	Demand for information
	Trust and confidence in information
	Volume and range of information
<b>Social</b>	<b>Definition: Open data exploited for improving transparency and accountability, and creating other important benefits for society</b>
<i>Categories of value</i>	Crime and justice
	Disease prediction and prevention
	Mobility access
	Mobility efficiency
	Mobility planning
	Mobility systems planning

As the table suggests there are several categories of value for each macro characteristics. Each of them can be measured through several qualitative and quantitative indicators. The tables presented in the sub-sections below indicate which indicators have been found to assess each category of value for each macro characteristic of the HVD.

It is important to note here that the 32 categories of value and 126 possible indicators have been identified with the purpose of guiding the data collection activities and fostering a common thinking about HVD. As expected, the study team was not able to find evidence for each HVD in all these areas nor was it needed, as no HVD tick all these "value" boxes. Different HVDs have different values, for instance some bearing more potential for economic benefits (i.e. see section on company and company ownership) and others for environmental benefits (i.e. see mobility section). For this assignment, we consider that HVDs will be those having either a value in many of the categories of the framework (breadth) or a very strong value in particular categories (depth).

### 2.1.1 Environment and climate change

Table 2 – Possible indicators for measuring climate change and environmental value

Value	Indicators
Citizen engagement in addressing climate change	Citizen understanding of climate change issues
	Number of citizen-led environmental initiatives
Climate change	CO2 emissions
Energy management and efficiency	Levels of energy use
Environment management	Air quality
	Citizen use of waste disposal sites and schemes
	Number of flood prevention schemes
	Responsiveness to environmental emergencies

### 2.1.2 Economic

Table 3 – Possible indicators for measuring economic value

Value	Indicators
Competition	Exit of inefficient firms
	Improvements in performance of old firms
	Lower firm input costs
	Number and level of entry of firms
Consumer benefits	Attractiveness of products to consumers
	Purchasing power of consumers
Economic output	Economic output from digital sectors
	Market size
	Rate of sector growth
	Value added
Economy monitoring	Use of imports and exports information
	Use of inflation and GDP information
Employment	Number of countries in which new jobs have been created
	Number of new high value jobs created
	Number of new jobs created
	Number of new jobs in high value sectors created
	Number of sectors in which jobs created

<b>Value</b>	<b>Indicators</b>
	Rate of increase in the creation of new jobs
	Total number of jobs
International competitiveness	Level, range, and value of exports
Product market dynamism	Attractiveness of products to consumers
	Diversity of products
	Number of products bought by consumers
	Prices optimization
	Product improvement
Productivity and commercialisation	Citizen and firm ease of paying taxes
	Efficiency of logistics
	Level of business costs
	Level of fraud
	Level of market insight
	Level of reporting demands on business
	Level of sales by firms
	Number and types of new business model
	Number of new products and services
	Optimization of prices
	Quality of data-driven decision-making
	Revenue of firms
	Time saved
	Time searching for information
	Turnover of firms
Public-private coordination	Quality of coordination between public and private organisations
Trust and transaction costs	Efficiency of market transactions
	Reliability and transparency of transactions in the real property sector
	Reliability of market transactions

### 2.1.4 Innovation & AI

Table 2 - Possible indicators for measuring innovation and AI

Value	Indicators
Citizen innovation	Level of self-servicing by citizens
Entrepreneurialism and private sector innovation	Commercial use of public sector technology
	Growth of firms using public data
	Number of firms using public sector data
	Number of new products and services
Public sector innovation	Number of registered developers of apps using public sector information
Public-private coordination	Number of hackathons and similar initiatives for private sector collaboration

### 2.1.5 Public services and public administration

Table 3 – Possible indicators for measuring value for public services and public administrations

Value	Indicators
Access to public services	Ease of access by citizens to healthcare facilities
	Efficiency of school choices made by citizens
	Quality of comparisons made by citizens between schools
	Use of libraries
Public administration, transparency accountability & engagement	Access to voting locations
	Public participation in political and social activities
	Public understanding of legislative processes
	Public understanding of political processes
	Public understanding of public administration and spending
Public sector procurement	Level of competition in bidding
Public sector revenue	Level of tax revenue
Public services management	Cost of maintaining information on companies
	Cost of publishing open data relative to previous costs, and returns
	Efficiency of public spending allocation
	Level of citizen requests for information from public sources
	Location accuracy of public services provision
	Number of links between datasets

<b>Value</b>	<b>Indicators</b>
	Number of similar datasets managed by different organisations
	Number of staff employed in the creation, maintenance, and management of public sector information
	Range of data available to public sector decision-makers
	Savings made from moving from SMS alerts
	Size of advertising budget
	Spending on in-house app development
	Total cost savings for the public sector
Public services performance	Ability of public health services providers to predict risks to health
	Improvements in school performance
	Quality of healthcare services
	Responsiveness of emergency services

### 2.1.6 Re-use

Table 4 – Possible indicators for measuring value for reuse

<b>Value</b>	<b>Indicators</b>
Demand for information	Level of demand by country
	Number of companies using information
	Number of customer groups using public sector information portals
	Number of established and returning users
	Number of licences issued, delivered, sold
	Number of new users
	Number of online subscribers to information
	Number of re-users
	Number of requests received by PSI holders for data
	Rate of change in the level of re-use
	Requests for information
	Total income received for the information
	Traffic on public sector information portals

<b>Value</b>	<b>Indicators</b>
Trust and confidence in information	User confidence in the market for re-use of public sector information
Volume and range of information	Download volume
	Number of types of public sector information used

### 2.1.7 Social

Table 5 – Possible indicators for measuring social value

<b>Value</b>	<b>Indicators</b>
Crime and justice	Compliance with the law
	Crime monitoring
	Legal system efficiency
	Public knowledge of legislation
Disease prediction and prevention	Disease detection
	Disease prevention
Mobility access	Number of public transport users
	Number of public transport users from marginal groups
	Use of foot and cycle paths
	Use of healthy mobility schemes by citizens
Mobility efficiency	Access to apps by public transport users
	Awareness of mobility delays
	Hours saved when searching for parking
	Savings made from moving from SMS alerts
	Time saved from avoiding congestion and delays
	Time spent on public transport
Mobility planning	Transport cost savings
	Ease of citizen transport planning
	Numbers of people walking or cycling
	Transport user satisfaction
Mobility systems planning	Use of travel information data
	Mobility planning
	Transport system integration

## 2.2 Company and company ownership

In this section, the report provides a detailed overview of the datasets in scope of the analysis for this thematic area, and maps the main benefits and use cases of these data fields.

### 2.2.1 Identification of the datasets in scope

For the definition of the datasets in scope of the analysis for the company and company ownership thematic area, the study team started from mapping all the relevant legislation at the European level, in order to identify which data fields must already be made available by Member States, and therefore exist all across the European Union.

From this EU level legislative perspective, there are several pieces of legislation that help understanding which data fields on company and company ownership need to be provided by all Member States, and these are:

- The Company Law Directive (2017/1132/EU)<sup>28</sup>;
- The Accounting Directive (2013/34/EU)<sup>29</sup>;
- The (Fifth) Anti-Money Laundering Directive (2018/843/EU)<sup>30</sup>;
- The Transparency Directive (2004/109/EC)<sup>31</sup>;
- Regulation (EU) 2015/848 on insolvency proceedings<sup>32</sup>.

The table below summarises the most important provisions of these five legislative acts in terms of establishing obligations for Member States to make available certain data fields on company and company ownership. It is important to mention here that, concerning the information on insolvency, these are considered as falling in the scope of the company and company ownership thematic area when they refer to companies (leaving out of scope individuals which are nonetheless also covered by the register).

---

<sup>28</sup> Directive (EU) 2017/1132 of the European Parliament and of the Council of 14 June 2017 relating to certain aspects of company law, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02017L1132-20200101>

<sup>29</sup> Directive 2013/34/EU of the European Parliament and of the Council of 26 June 2013 on the annual financial statements, consolidated financial statements and related reports of certain types of undertakings, amending Directive 2006/43/EC of the European Parliament and of the Council and repealing Council Directives 78/660/EEC and 83/349/EEC, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013L0034&from=EN>

<sup>30</sup> Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L0843>

<sup>31</sup> Directive 2004/109/EC of the European Parliament and of the Council of 15 December 2004 on the harmonization of transparency requirements in relation to information about issuers whose securities are admitted to trading on a regulated market and amending Directive 2001/34/EC, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32004L0109&from=EN>

<sup>32</sup> Regulation (EU) 2015/848 of the European Parliament and of the Council of 20 May 2015 on insolvency proceedings, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R0848>

Table 6 – Relevant EU legislation for company and company ownership

Company Law Directive	Accounting Directive	5 <sup>th</sup> Anti-Money Laundering Directive (AMLD)	Transparency Directive	Regulation on insolvency proceedings
<p>Article 14 of this Directive requires Member States to take the measures required to ensure compulsory disclosure by companies of at least the documents and particulars listed there in. This includes, among others, the instrument of constitution, the amendments thereof, the appointment, termination of office and particulars of the persons who either are authorised to represent the company, and other information.</p> <p>The company law directive was recently modified by Directive (EU) 2019/1151 as regards the use of digital tools and processes in company law.</p>	<p>This Directive defines and explains what the accounting documents must comprise depending on the type of company, size, or situation.</p> <p>Accounting Documents may contain a variety of documents:</p> <ul style="list-style-type: none"> <li>• financial statements;</li> <li>• management report, audit report (Art. 30);</li> <li>• the country by country report (Art 45);</li> <li>• the governance report if separated from the management report (Art 20(2));</li> <li>• the non-financial statement is separated from the management report (Art 19(4), Art. 29(4)).</li> </ul>	<p>The Fourth Anti-Money Laundering Directive already required Member States to set up a “beneficial ownership” register<sup>33</sup> but the access to this information was restricted to those who could prove “an interest”. The revised, fifth AML Directive makes access unrestricted and stipulates that Member States may charge a fee for access to beneficial ownership data, but that it shall not exceed the <i>administrative costs associated with developing and maintaining the register (marginal costs)</i>. Specifically, AMLD 5 requires the following beneficial ownership information to be made available by Member States:</p>	<p>The Transparency Directive lists the information companies with securities on public regulated capital markets must make available:</p> <ul style="list-style-type: none"> <li>• Periodic information: annual financial reports, and half-early financial reports;</li> <li>• Ongoing information: major holding of voting rights, and information requirements for issuers.</li> </ul>	<p>This regulation establishes in its article 24 that Member States must set up insolvency registers containing a number of information to be made publicly available (for free) and notably:</p> <ul style="list-style-type: none"> <li>• the date of the opening of insolvency proceedings;</li> <li>• the court opening insolvency proceedings and the case reference number, if any;</li> <li>• the type of insolvency proceedings;</li> <li>• whether jurisdiction for opening proceedings is based on Article 3(1), 3(2) or 3(4);</li> <li>• if the debtor is a company or a legal person, the debtor's name, registration number, registered office</li> </ul>

<sup>33</sup> “Beneficial ownership refers to the person or persons who ultimately own or control an asset (for example, a property or a company) and benefit from it. The concept of beneficial ownership exists because the direct legal owner of an asset is not necessarily the person ultimately controlling and benefitting from the asset”. Briefing Paper, Number 8259, 7 August 2019 Registers of beneficial ownership, <https://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-8259#fullreport>



		<ul style="list-style-type: none"> <li>• Name of the beneficial owner;</li> <li>• Month and Year of Birth;</li> <li>• Country of Residence;</li> <li>• Nationality; and</li> <li>• Extent of Beneficial Interest held.</li> </ul>		<p>or, if different, postal address;</p> <ul style="list-style-type: none"> <li>• if the debtor is an individual whether or not exercising an independent business or professional activity, the debtor's name, registration number, if any, and postal address or, where the address is protected, the debtor's place and date of birth;</li> <li>• the name, postal address or e-mail address of the insolvency practitioner, if any, appointed in the proceedings;</li> <li>• the time limit for lodging claims, if any, or a reference to the criteria for calculating that time limit;</li> <li>• the date of closing main insolvency proceedings, if any;</li> <li>• the court before which and, where applicable, the time limit within which a challenge of the decision opening insolvency proceedings is to be lodged in accordance with Article 5, or a reference to the criteria for calculating that time limit.</li> </ul>
--	--	---	--	--

				The Regulation however does "not preclude Member States from charging a reasonable fee for access to the documents or additional information.
--	--	--	--	---

The analysis of EU level legislation therefore already allows to identify a number of data fields, which all Member States should possess and are required to be made public, although not necessarily in an open data format. However, there is not a unanimous definition on the concept of 'company'. Member States' legislative frameworks include a wide range of legal forms (e.g. public limited companies, partnerships, trust owners, individual entrepreneurs, foundations, NGOs), and varying provisions on whose data has to be included in their companies' registers. This fragmented landscape poses the question regarding the material scope of the PSI Directive, and more concretely which "types" of companies, or even legal entities (i.e. public trusts) should be included. From the reusers' perspective, any economic operator is considered as a company/legal entity, and should thus be subject to the PSI Directive. In this sense, according to a joint paper submitted by Febis, PSI Alliance, and GFII, the scope of this category should be broad, and include all information made available through any type of business register, regardless the legal form of the business. Therefore, they suggest to consider that company and company information refers to all businesses, in the sense of Art 54 TFEU.<sup>34</sup> This ambitious vision however is not shared by data holders who tend to consider in scope as "companies" only whatever is already provided by their own registers.

Furthermore, a few stakeholders suggested including in the scope of the PSI Directive also publicly available information related to "specific sector companies" as for instance legal firms and lawyers or transport companies. In the case of the former, national law bar associations possess additional data on these companies and professionals which are already publicly available and might be of value for LegTech (i.e. Doctrine.fr<sup>35</sup>). For transport companies, there is an EU level regulation<sup>36</sup> mandating the establishment of national electronic registers containing basic data like fleet name and size, managers' names and legal form for each company. These are only a couple of examples of additional company data which might be available at the Member States level and are considered of value by reusers but not necessarily provided through general business registers.

Despite these differences on what "companies" mean and what is collected and made available in terms of different types of companies, fundamentally, when available, the information covered by EU level legislation concern four aspects:

- **Basic information on company:** this category includes all fundamental information characterising companies and can be split according to the personal or non-personal nature of the information. In a vast majority of countries, basic non personal information is available for free and in a machine readable format although there are exceptions (i.e. the capital of companies, or the number of employees are not available in all countries). Personal information, on the other hand, is less available (for free) across EU Member States, and the data holders have raised concerns about sharing these data points more as it could lead to privacy issues.
- **Company documents and accounts:** this category includes all documents/reports that companies are obliged/encouraged to provide to public authorities/business registers (regularly or not). Amongst the accounting documents<sup>37</sup>, Balance sheets and financial statements are very often made available but charged for. Annual accounts (and especially management reports and audit

---

<sup>34</sup> Art. 54 TFEU: '*Companies or firms*' means companies or firms constituted under civil or commercial law, including cooperative societies, and other legal persons governed by public or private law, save for those which are non-profit-making.

<sup>35</sup> <https://www.doctrine.fr/>

<sup>36</sup> Article 16 of EC 1071/2009 lists the minimum data that transport companies must contribute to the register; the register "shall be accessible to all the competent authorities of the Member State in question." <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009R1071>

<sup>37</sup> As defined in the Company Law and Accounting Documents Directive, accounting documents refer to: (a) Financial statements, whether individual or consolidated, including balance sheets, P&L, notes, as well as possibly a cash flow statements and other statements; (b) Management report; (c) The auditor's report; (d) The non-financial statement (ESG), unless it is included in the management report; (e) The corporate governance report, unless it is included in the management report; (f) The Country-by-country report (extractive / logging industries).

reports) are provided in a smaller number of countries but are becoming increasingly interesting for reuse. In certain countries, other documents can be added to this list (i.e. minutes of management board meetings in France)<sup>38</sup>.

- **Company ownership information:** this category includes the data on shareholders/associates but also ultimate beneficial owner(s) of companies, which needs to be provided by Member States according to the Anti-Money Laundering Directive<sup>39</sup>. The establishment of beneficial ownership registers across Member States is still ongoing and a majority of countries foresees to charge for these data according to DG FISMA.
- **Company insolvency status:** this category includes the data on the insolvency status of the company as described by Regulation (EU) 2015/848 on insolvency proceedings<sup>40</sup> described above.

This list of required publicly available datasets based on EU legislation constituted a possible starting point for the scoping discussions. Furthermore, interviews with stakeholders allowed to go one step further and develop a **preliminary “wish list”** of datasets which are considered to be high value from an economic, societal and reuse perspective. Generally speaking, this list includes data fields falling under the four abovementioned aspects but is more detailed in terms of data points considered of high value, as shown in the table below.

---

<sup>38</sup> Second online focus group

<sup>39</sup> Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L0843>

<sup>40</sup> Regulation (EU) 2015/848 of the European Parliament and of the Council of 20 May 2015 on insolvency proceedings, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R0848>

Table 7 – Initial lists of HVDs under company and company ownership thematic area

Basic information		Company documents and accounts	Company ownership		Company insolvency status
<i>Non personal</i>	<i>Personal</i>		<i>Non personal</i>	<i>Personal</i>	
<ul style="list-style-type: none"> <li>- Name of the company (full version; in different languages when applicable)</li> <li>- Company status (active, resolved, in liquidation, reconstruction, merger ...)</li> <li>- Founding date</li> <li>- Cessation date (if applicable);</li> <li>- Historical names</li> <li>- Addresses (i.e. legal, visiting postal)</li> <li>- Legal form</li> <li>- Identifiers (registration number / company identifier / the valid VAT identification number / phone number / e-mail address)</li> <li>- Data from VIES</li> <li>- Member State where registered</li> <li>- NACE code (of the predominant and secondary activities and the code's source)</li> <li>- Number of employees</li> <li>- Turnover</li> <li>- Capital</li> <li>- Detailed information on branches (including the features presented elsewhere under</li> </ul>	<ul style="list-style-type: none"> <li>- (Name(s) of company legal representative(s)</li> <li>- Name of company's directors)</li> <li>- The appointment, termination of office and particulars of the persons who either as a body constituted pursuant to law or as members of any such body: are authorised to represent the company in dealings with third parties and in legal proceedings; it shall be apparent from the disclosure whether the persons authorised to represent the company may do so alone or are required to act jointly; take part in the administration, supervision or control of the company</li> <li>- All changes (to individual companies and list of companies dissolved), and date of the last update</li> </ul>	<ul style="list-style-type: none"> <li>- Legal entities</li> <li>- Accounting documents, which include:                             <ul style="list-style-type: none"> <li>o consolidated financial statements (incl. the list of resident and foreign affiliates and subsidiaries, their countries, and unique identifiers),</li> <li>o non-financial statements, management reports,</li> <li>o transfer prices reports (e.g. as in the country-by country reports of BEPS Directive (2016/1164)); and</li> <li>o other reports (e.g. financial reports, audit reports,</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- Share (percentage) of ownership, and nature and extent of Beneficial Interest held (in shareholding and/or voting rights) as well as legal ownership</li> <li>- Capital links between companies</li> <li>- All changes, and date of the last update</li> </ul>	<ul style="list-style-type: none"> <li>- Name of the owner</li> <li>- Month and Year of birth</li> <li>- Nationality</li> <li>- Owner identifier</li> <li>- Names of shareholders</li> <li>- Country of residence of the shareholders/owners</li> </ul>	<ul style="list-style-type: none"> <li>- Type of insolvency proceeding</li> <li>- Time limit for lodging claims</li> <li>- Date of closing main insolvency proceedings</li> <li>- The court before which the decision opening insolvency proceedings is to be lodged</li> <li>- All changes (i.e. to individual companies and list of companies dissolved), and date of the last update</li> </ul>

<p>“basic information”)</p> <ul style="list-style-type: none"> <li>- All changes (to individual companies and list of companies dissolved), and date of the last update</li> </ul>		<p>corporate governance reports);</p> <ul style="list-style-type: none"> <li>- Detailed data on branches (including the features presented elsewhere under “company documents and accounts”)</li> <li>- Intra-group transactions</li> <li>- Date of the last update of the reports</li> <li>- Other companies documents which are provided to the authority (i.e. companies’ meeting minutes)</li> </ul>			
--	--	--	--	--	--

The first category of data included in the table above concerns *basic information on companies* and can be split up into non personal and personal data. The distinction between the two categories relies on the different nature of the data, and the diverging types of measures applicable to each of them. Concerning the use of personal data, Member States have different traditions and approaches to it. Data holders in some Member States (e.g. France, Italy, Malta, and the Netherlands) shared their concerns about sharing personal data for free, and fear a misuse of it. On the other hand, reusers also indicated the inconsistency of using this argument while selling the data to those stakeholders willing to pay a fee. In their view, if personal data can be purchased, it should be available publicly for the sake of consistency. Moreover, reusers also highlighted the access to non-personal data without personal data decreases significantly the utility of the data overall. Therefore, the two categories should go hand in hand from the reusers' perspective.

The non-personal data sub-category refers to the basic information of companies. As indicated above, this type of information is available in the vast majority of EU Member States. However, there are some exceptions for specific data points, such as the capital, or the number of employees.

The second pillar displayed in the table above refers to *company documents and accounts*. This pillar includes the key documents related to the company (e.g. legal entities, balance sheet, financial statements). In most of the countries, data holders provide these data fields against a fee. Similarly to the previous pillar, there are also standardisation issues concerning the type of documents to be included. For example, in some given Member States the meeting minutes are provided as part of the company documents (i.e. in France), while in others this type of document is not included.

The information on *companies ownership* listed in the table matches to a large extent what is made already available under the AMLD 5. However, this Directive currently does not fulfil all "PSI" users' requirements as 1) it allows charging practices (of administrative costs, currently varying between 1 euro per search and 10 euro per search, but Member States are still working on the implementation of the Directive<sup>41</sup>), and 2) is conceived for searching individual beneficial owners and does not foresee APIs or bulk download options<sup>42</sup>. Furthermore, there are countries in which the license attached to this information forbids reuse: in Austria, the Ministry of Economy in charge of the beneficial ownership registers developed a specific license which prevents reuse of the information on the ground of data protection. Finally, stakeholders stressed the need to obtain access to sufficiently "detailed" information when it comes to beneficial and legal ownership in order to clearly and easily determine the power of control, and concerns were raised that the Anti-Money Laundering Directive was not precise enough in terms of data attributes and data quality<sup>43</sup>. In this sense, reusers also highlighted<sup>44</sup> the need to include in this category the *capital links* to smoothly identify any existing dependency between companies. For these reasons the inclusion of these information as HVD must be carefully considered in terms of costs and political implications.

Finally, concerning *insolvency information*, it must be highlighted that, although a preliminary analysis seemed to indicate that the category was less important<sup>44</sup>, its relevance has been confirmed during our second online focus group. Mostly, these datasets are important for reusers working on Know Your Customers (KYC) and business clearance types of applications. Although the insolvency registers are available in all countries and for free, they are similar to beneficial ownership registers in the sense that

---

<sup>41</sup> According to DG FISMA a strong majority of Member States has already put in place a beneficial ownership register for legal entities and half of them are charging/planning to charge for individual searches in the register.

<sup>42</sup> Interview with DG FISMA

<sup>43</sup> In this domain, Open Ownership has developed a data model which specifies the type of information required for the entity, person and ownership or control statement concerned: <http://standard.openownership.org/en/latest/schema/concepts.html#data-model-overview>

<sup>44</sup> See Deliverable 1 of this assignment

they focus on individual searches due to its intended functionality (and the privacy concerns related) rather than bulk download and APIs access. As such, they are not really in line with the spirit of the PSI Directive in terms of reusability.

The next section describes the overall benefits and value which can be attributed to the datasets discussed above, building on our value framework developed under Task 1.

### 2.2.2 Analysis of the value of these datasets

Company, insolvency and beneficial ownership datasets are unanimously considered of high value by the literature and the reusers and many use cases have already been mapped for this type of information. The EU Business Graph project<sup>45</sup> for instance created a categorisation of business use cases which is useful to help conceptualise the re-use of company data by private stakeholders and NGOs. The project identified three main use case sectors: *the business information sector, the marketing and sales sector, and the business publishing sector.*

- The business information sector consists of organisations that re-use company and company ownership data, insolvency and beneficial ownership information to review and evaluate companies – for example by performing know your customer (KYC) activities, credit checks (to approve credits for all types of firms, including SMEs), check for trade credits, or evaluating tax compliance. Due to an increasing attention to business clearance and KYC, especially in the banking sectors, this type of use cases has a very high economic value according to stakeholders. The COVID-19 crisis might further increase the reuse of this type of information: on the one hand, the public sector is implementing recovery programmes for businesses, which require a very good knowledge of companies' situations. In certain countries, recovery programmes are characterised by conditionalities<sup>46</sup>, which can only be checked if company information are available to public authorities, but also to banks. In fact, the private sector, and especially banks and insurances, but also the suppliers (for trade credits), which are working on extraordinary loans and guarantee schemes also require company information to run their checks on applicants. Therefore, the importance of company and company ownership data for the business information sector has increased in a post COVID-19 business environment, as argued by many consulted stakeholders.
- The *marketing and sales sector* utilises company information to perform tasks like market research and provide various business analytics services to clients. Some historical data platforms such as D&B Hoovers<sup>47</sup>, Factiva<sup>48</sup>, Altares<sup>49</sup> or Refinitiv<sup>50</sup> are strong reusers and key intermediaries for company and company ownership data. Their services include providing intelligence for sales and marketing prospects and gathering contacts for marketing activities. It is sometimes difficult to draw a strong line between marketing/sales services and the business information sector as many players are active in both markets: for instance Altares' services include both data provision for KYC, but also sharing of personal information and business contacts for marketing purposes. Nonetheless, with regards to marketing and sales, some stakeholders noted that the importance of this use case is currently decreasing, especially when compared with the previous use case and due to stricter General Data Protection practices. This seems to be confirmed by the analysis carried out by Company House on its reusers' base: only 6% of reusers report leveraging Company House information for marketing and sales purposes while 71% use the data for compliance purposes for instance<sup>51</sup>.

---

<sup>45</sup> <https://www.eubusinessgraph.eu/business-cases/>

<sup>46</sup> <https://www.bdo.dk/en-gb/covid-19-and-the-danish-business/covid-19-and-danish-business>

<sup>47</sup> <https://www.dnb.com/products/marketing-sales/dnb-hoovers.html>

<sup>48</sup> <http://factiva.com/sources/facts.asp>

<sup>49</sup> <https://www.altares.com/fr/>

<sup>50</sup> <https://www.refinitiv.com/en>

<sup>51</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015



- The *business publishing sector* can be understood to mean organisations that may publish and report on company data for the purpose of improving transparency or promote ethical practices. Social, environmental and governance data reported by companies are also essential for NGOs, and represent important public sector use cases. This is the case for instance of many NGOs such as Transparency International<sup>52</sup>, Open State Foundation<sup>53</sup> or Global Witness<sup>54</sup>. Once again, it can be argued that the increased financial support provided by governments to companies in the context of the COVID-19 outbreak might call for a strengthened monitoring of companies behaviours and therefore an increasing importance of company information for NGOs and civil society players. On top of NGOs working on business transparency, the public sector also works extensively with business data from a fraud analytics perspective and in areas such as customs, taxation and procurement. In this respect, procurement data, were also often mentioned as potential high value dataset and some reusers from the business information sector also argued that having public contracts awarded to private companies is very valuable. However, it is uncertain whether procurement data would fit in the overall category of company and company ownership, and are they considered as out of scope for the PSI Directive at this stage. Nonetheless, they bring additional value when reused together with company and company ownership information: examples of how the combination of these datasets create value are included in the study and in the analysis of benefits provided in the table below.

Besides these main and “traditional” use cases identified by the EU Business Graph project<sup>55</sup>, there are a number of other use cases including for instance:

- The reuse of company and company ownership information for the provision of new Business to Citizens (B2C) or Business to Business (B2B) services: for instance, B2B services such as provision of SMEs friendly reporting services in Denmark or business location intelligence on where to start a business<sup>56</sup> are very often based on company datasets.
- The development of LegTech applications of different kind: LegTech are using company information more and more and for different purposes and services. Companies such as Doctrine.fr<sup>57</sup> for instance index companies on their website, providing information on their ongoing litigation and using company ownership data to be able to feed those pages.
- Academic research: it is worth reminding that company data and statistics are considered key datasets by researchers and economists in particular. In many different countries (i.e. Greece, France for the datasets of the INSEE), academics and researchers are counted amongst the most intensive reusers of aggregated company datasets.

As the high amount of possible use cases suggests, company and company ownership data are valuable in many different contexts and for many different reusers including the private and public sector but also civil society organisations and researchers. In this respect, their value is not concentrated on a few specific categories but rather spread across several. The following table provides a tentative assessment of the value of the “company characteristics” according to the value framework. The table only includes value categories for which there is very **substantial evidence** supporting the high (potential) value of company and company ownership data. It is worth mentioning, however, that the value of these information for other categories not included in this table should not be excluded.

<sup>52</sup> <https://www.transparency.org/>

<sup>53</sup> <https://openstate.eu/nl/>

<sup>54</sup> <https://www.globalwitness.org/fr/?p=2>

<sup>55</sup> <https://www.eubusinessgraph.eu/business-cases/>

<sup>56</sup> See for instance the Locator project : [http://www.the-locator.eu/72EMR\\_Frontend/](http://www.the-locator.eu/72EMR_Frontend/)

<sup>57</sup> <https://www.doctrine.fr/>

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
Economic	Competition		There is strong evidence that the reuse of company information improves firms' performance through their optimisation of business partners' clearance processes (i.e. information included under the pillars basic information, but also company documents and accounts, and companies ownership) and decrease in investments risks <sup>58</sup> . Inefficient firms might also exit the market as a result of better clearance processes <sup>59</sup> . Furthermore, reuse of company information and procurement data can also streamline companies' procurement processes <sup>60</sup> . Overall, availability of company information supports good competition in the market and allow healthy firms to strive <sup>61</sup> .
	Consumer benefits		There is a very strong literature linking benefits for consumers to higher business transparency (this refers to the activity of the company, i.e. second pillar: company documents and accounts, specifically the account documents, and other company documents). Consumers today are increasingly concerned about the environmental and social impact of their choices, and the development of a market of applications for ethical/environmentally friendly shopping is a reflection of this trend. There is a very long list such "ethical shopping apps", which reuse company and company data for assessing companies or products. Examples include GoodGuide for choosing healthier and environmentally better food products or Ethical Barcode, for any type of product, including clothing <sup>62</sup> . Most of these apps integrate data from different sources, including from business registers when it comes to information related to annual accounts (incl. balance sheets) and environmental footprint.
	Economic output		Evidence on the value of company characteristics (i.e. basic information) for economic output is largely available. One of the abovementioned use cases for company data is compliance and KYC services. The market for these services will amount to up to 1 015.36 Million Dollars globally by 2026 <sup>63</sup> . The anticipated CAGR for the e-KYC market is around 22% from 2020 to 2026 <sup>64</sup> . In Europe, the KYC market is very well developed, although precise figures on its size and growth are more difficult to find. Company data are at the heart of this use case and the increased provision of these information in open data (especially in France) has favoured the emergence of many smaller players on the KYC market as reiterated by several stakeholders.

<sup>58</sup> As the Report on Valuing Company House data suggests "The most common reasons for using the company search services were to confirm and check the consistency of information provided by companies (e.g. suppliers and/or customers), or as part of more detailed due diligence research into a company". See: "Valuing the user benefits of companies house data Report 2: Direct Users" BEIS Research Paper Number 2019/015

<sup>59</sup> Stakeholders interviews

<sup>60</sup> <https://www.mckinsey.com/business-functions/operations/our-insights/the-era-of-advanced-analytics-in-procurement-has-begun>

<sup>61</sup> <https://slackhq.com/transparency-in-business-company-evolution>

<sup>62</sup> For a longer list see <https://shopethica.com/blogs/latest-stories/9-ethical-shopping-apps-plug-ins>

<sup>63</sup> <https://www.marketwatch.com/press-release/global-e-kyc-market-size-projected-to-reach-usd-101536-million-by-2026-facts-factors-2020-02-13>

<sup>64</sup> <https://www.marketwatch.com/press-release/global-e-kyc-market-size-projected-to-reach-usd-101536-million-by-2026-facts-factors-2020-02-13>

Macro characteristics	Value category	Assessment of value for the datasets in scope	Examples/references
			<p>Besides this, the IMF has indicated that up-to-date is critical. Investors look for reliable audits and accounting documents<sup>65</sup>. Thus, further data integration at the EU level would benefit very deeply capital markets investors and companies looking for public funding.</p> <p>Looking beyond specific use cases, Company House reports that basic company information, like address, nature of business, and incorporation date accounts for 41% of all company data direct usage benefits – generating a total of 0.4 to 1.2 billion GBP per year in the United Kingdom alone.<sup>66</sup> In total, “the overall aggregate benefit for UK users of CHS in 2018 is estimated to be within the approximate range of £1 billion to £3 billion<sup>67</sup>”. In Italy, a study carried out two years ago suggests that the value of the company data market amounts to 600 million euro in 2020<sup>68</sup>. This value is increasing over time at a 4% growth rate<sup>69</sup>.</p>
	Employment		<p>If increased company data availability allows more players to enter markets – whether that would be companies who re-use company data, for example in the KYC market (i.e. using data fields under the first pillar basic information, and third pillar companies ownership), or companies in the registries themselves who can now more easily prove their reputation, and enter for example public procurement (using data fields under the second and third pillars, particularly the accounting documents, and the percentage ownership) – then competition improves in those areas and, presumably, employment is positively impacted. Given the estimates of economic output available (for the United Kingdom) and the high number of reusers, it is logical to assume a very positive impact on employment although no specific studies on this aspect could be identified across EU countries.</p>
	Productivity and commercialisation		<p>There is strong evidence suggesting that availability of company characteristics (i.e. data fields in the first pillar basic information) and data fosters the development of many new products and services, allows the development of new business models, increase turnover of firms and reduces risks of frauds. Concerning the development of products and services, the many use cases identified including some “innovative” use case suggests that reusers are still exploring what can be done with company data and that there is still a vast potential for new applications to be developed. Stakeholders agree that opening these datasets, for instance in France, has brought to higher levels of competition on the market and the emergence of new business models and new players, especially in the banking and finance sector<sup>70</sup>. While this has to a certain</p>

<sup>65</sup> See : <https://www.imf.org/en/Publications/Staff-Discussion-Notes/Issues/2019/09/06/A-Capital-Market-Union-For-Europe-46856>

<sup>66</sup> BEIS Research Paper 2019/015, pg. 16.

<sup>67</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>68</sup> Stakeholders interview

<sup>69</sup> Stakeholders interview

<sup>70</sup> Stakeholders interview

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			extent disrupted the business model or traditional data intermediaries it has also opened up a whole new realm of possibilities for businesses and created new markets. Finally, availability of this data also increases time saved and levels of market insights while reducing frauds and time searching for information <sup>71</sup> .
	Trust and transaction costs		Performance of due-diligence checks on potential contractors, or due diligence services provided by third-party organisations rely partially on company characteristics data (i.e. data fields such as the basic information, the accounting documents, and the percentage ownership) to draw conclusions on the reliability and trustworthiness of a company <sup>72</sup> . The study mandated by Company House suggests that a high percentage of reusers (in between 28% and 38%) leverage company data to carry out due diligence on buyers and suppliers <sup>73</sup> . The importance of these data for due diligence has been confirmed for a vast majority of the Member States interviewed. The availability of easily accessible companies' information makes due diligence easier, increases business trust and reduces transaction costs.
Innovation & AI	Citizen innovation		Freely available company data can improve citizen self-serving by allowing citizens to directly research the status (i.e. company status, and data fields under the company insolvency status pillar) and general information (i.e. basic information, and accounting documents) about a company with whom they have hired or contracted. Several apps already exist providing such types of services <sup>74</sup> . The usefulness of company data for credibility-checking depends, of course, on the granularity of the available data.
	Public sector innovation		Company and company ownership data (i.e. name of the owner, percentage ownership, capital links between companies) are reused by high number of public sector applications and developers across countries. It is actually one of the domain showing the highest potential for public sector innovation according to stakeholders, as it can be reused in many different ways and in combination with many different public services. Once again, the COVID 19 crisis could further highlight the innovative potential of these information and especially in the context of the delivery of recovery plans.
	Entrepreneurialism and private sector innovation		As noted elsewhere – a vast majority of stakeholders have expressed their belief that increased availability and accessibility of company data (i.e. basic information, name of the owner, and percentage ownership) will boost competition among re-user firms which may, in turn, encourage innovation (i.e. new/improved products and services) <sup>75</sup> .
Public sector	Public sector procurement		Company information (i.e. basic information, accounting documents, name of the owner, percentage ownership, and company status), analysed in combination with

<sup>71</sup> See for instance : Valuing the user benefits of companies house data, Policy Summary BEIS Research Paper Number 2019/015

<sup>72</sup> [http://www3.weforum.org/docs/WEF\\_PACI\\_ConductingThirdPartyDueDiligence\\_Guidelines\\_2013.pdf](http://www3.weforum.org/docs/WEF_PACI_ConductingThirdPartyDueDiligence_Guidelines_2013.pdf)

<sup>73</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>74</sup> Apps like DueDil offer a platform to search and explore the characteristics of companies. <https://www.duedil.com/about-duedil/our-technology>

<sup>75</sup> Stakeholders interviews

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			procurement data, can support the improvement of public sector procurement practices. Public procurement represents in between 13% and 20% of GDP worldwide <sup>76</sup> and there is solid evidence on the importance of company data from a procurement perspective: for public sector bodies, getting to know the bidders and carrying out background checks on their providers is absolutely crucial as the study of Company House suggests: around 50% of public sector reusers of Company House data leverage them for "making better decisions about suppliers and customers" <sup>77</sup> . Carrying out analysis of procurement and company data combines allows public sector to perform due diligence faster but also to feel more secure and to open up the market to new players. This aspect is particularly emphasised by stakeholders as the most positive consequence of the availability of company data <sup>78</sup> .
	Public services management		Companies and company ownership data (i.e. name of the owner, percentage ownership, capital links between companies, name of the shareholders) are reused by public sector bodies in the delivery of many different policies and services, ranging from investment policies to anti-corruption and social policies. They are the basis for most of the risk analytics applications developed in the areas of customs and taxation. There is strong evidence on the fact that company data greater availability allows public sector to manage policies and services better. As a recent study suggests, these data "helps to inform and support good policy decisions... (this data) is being used to inform and influence policy decisions in a number of ways including: providing contextual data to inform government reports, policy papers, impact assessments and ministerial briefings; helping to validate other government data; and providing information, statistics and reports to support decision-making at a local level" <sup>79</sup> . The same study also mentions that company data have always been extensively used by the public sector for public service management. However, increased availability of data for free in the UK also triggered an increase in the use of the data by the public sector <sup>80</sup> . Facilitating access to these datasets therefore also favours the development of better public services.
Re-use	Demand for information		There is strong evidence suggesting that the number of users and of requests related to company characteristics (i.e. basic information) increases exponentially when the information is free and publicly available <sup>81</sup> and that, in general, demand for this type of information is growing. In the United Kingdom alone, Company House has 1.5 million of unique visitors yearly and this number increased very significantly when data started to be provided for free <sup>82</sup> . Although data on the number of users is not available for

<sup>76</sup> OECD, "Fighting Corruption in the Public Sector": [www.oecd.org/gov/ethics/meetingofleadingpractitionersonpublicprocurement.htm](http://www.oecd.org/gov/ethics/meetingofleadingpractitionersonpublicprocurement.htm), 2013

<sup>77</sup> Valuing the user benefits of companies house data Report 4: Providers of Public Goods BEIS Research Paper Number 2019/015

<sup>78</sup> Stakeholder interviews

<sup>79</sup> Valuing the user benefits of companies house data Report 4: Providers of Public Goods BEIS Research Paper Number 2019/015

<sup>80</sup> Valuing the user benefits of companies house data Report 4: Providers of Public Goods BEIS Research Paper Number 2019/015

<sup>81</sup> Valuing the user benefits of companies house data, Policy Summary BEIS Research Paper Number 2019/015

<sup>82</sup> Valuing the user benefits of companies house data Report 4: Providers of Public Goods BEIS Research Paper Number 2019/015

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			many other countries, all stakeholders agree that, when data is available, figures reach the million users and searches <sup>83</sup> . In Italy, the number of reusers of company data increased steadily over time: the business register counts today around 300 000 individual users and 260 large reusers (+25 large reusers since 2018) <sup>84</sup> . Through the large reusers, another 90 000 organisations have access to company information provided by the business register <sup>85</sup> . Furthermore, these datasets are also directly accessed by more than 60 000 public authorities in Italy <sup>86</sup> . Finally, in Poland the number of reusers registered to the API Regon which provides company data is of 322 state entities and 5137 other users today <sup>87</sup> .
	Trust and confidence in information		According to reusers, company data (all data fields under basic information, company documents and accounts, company ownership, and company insolvency status) are amongst the most valuable and the size of the reuse market for these datasets testifies the benefits for them but also the level of trust and confidence of reusers in these information. It is worth noting here however that, for some large reusers, the trust in the quality of the data is declining when the datasets are provided in open data only <sup>88</sup> . The reason is linked to the loss of revenue for data holders, which are not able, as a consequence, to maintain high standards in the quality of the data and service provided. Some of these stakeholders show a willingness to pay for services, which are at higher standards and suggest to maintain "premium" services in parallel to the open data services.
	Volume and range of information		Amongst public sector's data, these datasets (all data fields under basic information, company documents and accounts, company ownership, and company insolvency status) are on the top of the list of the most downloaded. Company House counts more than 74 million visits to its portal every year: "for a weekday, the estimate is approximately 278 000 uses; and approximately 49 500 uses for a weekend (i.e. roughly 18% of weekday usage)" <sup>89</sup> . In Ireland, stakeholders mention that the business register must be able to cope with millions of searches to be considered as sufficiently solid <sup>90</sup> . These numbers gives an idea of the volume of company information downloaded and consumed by reusers.
Social	Crime and justice		Open company data (e.g. accounting documents, and company ownership) are very frequently used to identify illegal practices, both by the public and the private/civil society sector. For example, a stakeholder shared a case study outlining how company

<sup>83</sup> Stakeholders interviews

<sup>84</sup> Stakeholders interviews

<sup>85</sup> Stakeholders interviews

<sup>86</sup> Stakeholders interviews

<sup>87</sup> Stakeholders input

<sup>88</sup> Stakeholders interviews

<sup>89</sup> Valuing the user benefits of companies house data Report 4: Providers of Public Goods BEIS Research Paper Number 2019/015

<sup>90</sup> Stakeholders interviews

Macro characteristics	Value category	Assessment of value for the datasets in scope	Examples/references
		<p>data, available on public registries in a machine-readable and easily-combinable format, was used to uncover corruption in Myanmar's jade industry<sup>91</sup>. Using company data for exposing corruption and fraud within the public and private sector is the core business of several NGOs such as Transparency International or Open State Foundation. Furthermore, as mentioned above, public sector uses extensively company data to "to inform law enforcement investigations and/or inform court proceedings"<sup>92</sup>. Company House's study explains that this data is "a key source for law enforcement and 'the starting point' for all investigations of fraudulent activity"<sup>93</sup>. Finally, private companies using company data in their due diligence activities also contribute to exposing fraud and fighting crime. Therefore, the social benefits of these datasets should not be underestimated.</p>	
	Public engagement and understanding, and government accountability	<p>Linked to the value of these information from the procurement perspective, company datasets (i.e. accounting documents) also have a high value for citizens to understand how public administrations are spending tax payers' money and who the recipients of governments' contracts are<sup>94</sup>. This is particularly important from an anti-corruption perspective and company data are essential for journalists and NGOs to carry out detailed investigations.</p>	
Climate change	Environment management	<p>As suggested by DG FISMA in particular, the provision of companies' management reports, i.e. accounting documents (which should include contents on the environmental sustainability strategies and outlooks of firms) will allow citizens and society to monitor and recognise the efforts of businesses investing in sustainability and reward those. In a context where climate change and environmental concerns are increasingly sensible for citizens and as mentioned above, there is a market of applications providing background information on social and environmental behaviours of firms or on the environmental footprint of products, thus allowing customers to make informed choice.</p>	

<sup>91</sup> *White Paper: Open company data & the Global Witness Myanmar jade investigation, October 2015.*

<sup>92</sup> Valuing the user benefits of companies house data Report 4: Providers of Public Goods BEIS Research Paper Number 2019/015

<sup>93</sup> Valuing the user benefits of companies house data Report 4: Providers of Public Goods BEIS Research Paper Number 2019/015

<sup>94</sup> For a review of use cases and stories concerning the use of companies and procurement data see for instance the Open Contracting Partnership stories: <https://www.open-contracting.org/impact-stories/>

As the table suggests, and coherently with the use cases analysed, most of the value and benefits linked to the reuse of company and company ownership information are concentrated in the economic and reuse categories, although the social value is also very high and links to their potential for crime and fraud detection. When combined with procurement data in particular, company information also help understanding governments' expenditure and increase citizens trust and governments' accountability. Despite the very high potential of these datasets, and especially when they are made available for free and in accessible formats, the provision of companies and company ownership information across European countries remains suboptimal, as further described in the next section.

## 2.3 Geospatial

In this section, the report provides a detailed overview of the datasets in scope of the analysis for this thematic area, and maps the main benefits and use cases of these data fields.

### 2.3.1 Identification of the datasets in scope

For the definition of the datasets in scope of the analysis for the geospatial thematic areas, the study team started from mapping relevant legislation at the European level, in order to identify which datasets must be made available by Member States and therefore exist all across the European Union. At the EU level, there are several legislative acts that help understanding which of the geospatial datasets need to be made available by all Member States:

- *Infrastructure for Spatial Information in the European Community – INSPIRE Directive*<sup>95</sup> - 2007/2/EC
- *Regulation on the alignment of reporting obligations in the field of legislation related to the environment*<sup>96</sup>.

A number of relevant EC Regulations and EU Decisions (Implementing Rules) that form the EU legislative framework in the wider Geospatial domain, oblige Member States to provide geospatial information to the public and the European Commission. The implementation of these legal acts in Member States helps the provision of EU wide data.

The analysis of EU level legislation therefore already allows to identify a number of data points, which all Member States should possess and are required to be made public, although not in an open data format.

The INSPIRE Directive is a major effort to harmonise and share data between administrations, and also across borders. Spatial data in scope of the Directive are grouped into 34 data themes and include, among others, postcodes, national and local maps (cadastral, topographic, marine, administrative boundaries). According to the Directive, a European Spatial Data Infrastructure (SDI) has to be implemented to "enable the sharing of environmental spatial information among public sector organisations and facilitate public access to spatial information across Europe and assist in policy-making across boundaries". As the EU SDI is built on the national SDIs, the aspect of full re-use of national SDIs is very strongly embedded in the

---

<sup>95</sup> Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32007L0002>

<sup>96</sup> Regulation (EU) 2019/1010 of the European Parliament and of the Council of 5 June 2019 on the alignment of reporting obligations in the field of legislation related to the environment, and amending Regulations (EC) No 166/2006 and (EU) No 995/2010 of the European Parliament and of the Council, Directives 2002/49/EC, 2004/35/EC, 2007/2/EC, 2009/147/EC and 2010/63/EU of the European Parliament and of the Council, Council Regulations (EC) No 338/97 and (EC) No 2173/2005, and Council Directive 86/278/EEC, <https://eur-lex.europa.eu/legal-content/EN/TXT/?toc=OJ%3AL%3A2019%3A170%3ATOC&uri=uriserv%3A0J.L.2019.170.01.0115.01.ENG>



Directive. The Directive required member states to set up coordination structures to facilitate and monitor the progress of implementation<sup>97</sup>.

The table below illustrates the data themes as organized in three annexes. Annex I-II represent the datasets that create the base of the spatial infrastructure and serve as spatial reference to the real objects, whereas in Annex III datasets have clear role in supporting wider environmental policy processes.

Table 8 - List of INSPIRE data themes

<b>List of INSPIRE themes</b>		
<b>ANNEX I</b>	<b>ANNEX II</b>	<b>ANNEX III</b>
Coordinate reference systems	Elevation	Statistical units
Geographical grid systems	Land cover	Buildings
Geographical names	Ortho-imagery	Soil
Administrative units	Geology	Land use
Addresses		Human health and safety
Cadastral parcels		Utility and governmental services
Transport networks		Environmental monitoring facilities
Hydrography		Production and industrial facilities
Protected sites		Agricultural and aquaculture facilities
		Population distribution – demography
		Area management/restriction/regulation zones and reporting units
		Natural risk zones
		Atmospheric conditions
		Meteorological geographical features
		Oceanographic geographical features
		Sea regions
		Bio-geographical regions
		Habitats and biotopes
		Species distribution
		Energy resources
		Mineral resources

<sup>97</sup> See the Joint EEA-JRC Mid-term evaluation report on INSPIRE implementation by Christian Ansoerge, Max Craglia, Freddy Fierens, Paul Hasenohr, Stefan Jensen, Darja Lihteneger, Michael Lutz, Michel Millot, Maria Nunes de Lima, Elena Roglia, P Smits, and Robert Tomas. 2014.

While the Directive has become a fundamental reference for geospatial information, it has encountered challenges in its implementation, both in terms of standard adoption and data sharing. In other words, there is scope for additional intervention to achieve its goal. The mid-term evaluation of the Directive exposes delays in implementation and “based on the evaluation results, it is clear that greater effort at all levels by all actors is needed in the future.” Among the recommended actions, a prioritisation of datasets is mentioned, and in particular “to set clear priorities, i.e. to identify the most important datasets for end-user applications amongst the data themes.”<sup>98</sup>

This list of required publicly available datasets based on EU legislation constitutes a possible starting point for the scoping discussions. While no single data set or theme within the scope of Geospatial data stands out as most important, or as singularly valuable for re-use on its own, there are obviously some differences in utility of these data sets for specific use cases. The clearest example of this serendipity is the recent Covid-19 crisis, which suddenly made location data and their reusability essential to tracking the emergency and the opening up. Moreover, there seems to be no strict ranking possible, considering the thematic diversity in the data sets that individual MS suggested to be taken into account as HVD. Therefore, the interviews with stakeholders allowed to go one step further and tried to narrow down a preliminary “wish list” of datasets considered to be of high value from an economic, societal and reuse perspective and which could be seen as the initial datasets to be publicly released as high value. Our initial investigation considered the INSPIRE Annex I data themes (Coordinate reference systems, Geographical grid systems, Geographical names, Administrative units, Addresses, Cadastral parcels, Transport networks, Hydrography, Protected sites).

The table below includes the list of datasets most widely indicated for the initial release, as well as some generic use cases for these datasets. The descriptions provided in the table are based on the Inspire theme register<sup>99</sup>.

Table 9 – Geospatial thematic area – Datasets in scope

<b>Dataset</b>	<b>Short description</b>	<b>Use Cases</b>
<b>Administrative Units</b>	Units of administration, dividing areas where Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries. Land Administrative Units and Maritime Units are the basic units. Land Administrative Units are covering mostly land surface, while Maritime Units are covering territorial waters.	Mapping or use as statistical units; manage emergency rescue; waste management plans; protect water ecosystems, find responsible party for policy implementation and administration; forest management, subsidies for farmers, forecast agricultural production, spatial planning, monitoring of regional and urban policy implementation using territorial typologies based on administrative units, maritime spatial planning, integrated coastal management.
<b>Place Names</b>	Geographical names or place names (or toponyms) are the proper nouns applied to topographical features and settled (and used) places and spaces on the earth’s surface. Toponyms represent an important reference	Emergency response; economic, social and environmental analysis; cultural identity and heritage; mapping and navigation; providing a link index function to other spatial and aspatial data.

<sup>98</sup> European Commission, Report from the Commission to the Council and the European Parliament on the Implementation of Directive 2007/2/EC of March 2007 Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) Pursuant to Article 23, SWD (2016) 273, 2016.

<sup>99</sup> See the Inspire theme register available on line: <http://inspire.ec.europa.eu/theme>.

Dataset	Short description	Use Cases
	system used by individuals and societies throughout the world.	
<b>Addresses</b>	Location of properties based on address identifiers, usually by road name, house number, postal code. The basic unit of addressing is a building; a permanent construction, intended or used for the shelter of people, having at least one entrance from publicly accessible space.	Geocoding of statistical surveys, manage emergency rescue, locate where people are, accessibility studies, manage incidents; locate economic activities in ecosystem accounting.
<b>Buildings</b>	Geographical location of constructions above and/or underground, intended or used for the shelter of humans, animals, things, the production of economic goods or the delivery of services that refer to any structure permanently constructed or erected on its site. <sup>100</sup>	These data are required for serving citizens (e.g. school, hospital), assessments for air and noise pollution or risk assessments to various kinds of risks (earthquake, fire, flood etc.), monitoring of land consumption, population concentration and access to services, and are crucial for the emergencies.
<b>Cadastral Parcels</b>	Single areas of Earth surface (land and/or water), under homogeneous real property rights and unique ownership, where national law defines the real property rights and ownership.	Protect state lands; reduce land disputes, facilitate land reform, agriculture, land management, taxation, disaster management: real Estate Market; Taxation; LPIS (Agriculture); Land consolidation; Infrastructure Management; Spatial Planning, Protection of Soil and Water; Statistics.

This list was developed considering the feedback received during the stakeholders' interviews, the focus groups with the users and the inputs received from the data holders at national level. According to them, this selection of datasets represents a good starting point. The geospatial datasets in scope are meant as topographical and cadastral data. It is important to note that in this section, data as Digital Elevation Model (DEM), Digital Surface Model (DSM) and Digital Terrain Model (DTM)<sup>101</sup>, transport networks, hydrography and Land Use/Land Cover are not covered by this list, since they are included in the Earth Observation and Environment and Mobility thematic areas. Our findings have shown that the inclusion of LIDAR point cloud data (made with technologies like LIDAR) could be important, considering the related high value use-cases. Moreover, 3D data such as DTM and DSM are often sought for and LIDAR technologies offer highly accurate 3-dimensional points and component attributes. Transport networks (road networks, etc.) data is another key dataset that is geospatial by nature, however in this context it falls under the mobility thematic area. In particular, the road networks data is extremely important as, in connection with the addresses, they can be used for geo-referencing, in cases where house numbers are

<sup>100</sup> See the INSPIRE theme register: <https://inspire.ec.europa.eu/id/document/tg/bu>

<sup>101</sup> Regarding Ortho-photography, Elevation and Surface Models, interviewed stakeholders found surprising that geospatial data like DSM and DEM are included in Earth Observation or Environmental data, as these are clearly geospatial data. As they are often active in earth observation, for them such data are geospatial by their very nature.

missing. Later in the report, several examples of geospatial HVD-related use cases are included that will underline why it is important to pay more attention to more modern and open data formats (e.g. geopackage). At the same time, users request data with the highest spatial resolution and level of scale, whenever possible. Also, from a reusability perspective, users fully support and recommend that the data is provided under an open data license less restrictive as much as possible (e.g. CC0 or CC-BY).<sup>102</sup>

The national territories are divided into **administrative units**. These units delimit administrative areas where the Member States have and/or exercise jurisdictional rights, for local, regional and national governance, separated by administrative boundaries. They are considered as reference data through inclusion in Annex I, i.e. data that constitute the spatial frame for linking to and/or pointing at other information that belong to specific thematic fields such as the environment and socio-economic statistics, alongside many others. INSPIRE<sup>103</sup> excludes from administrative units the ones at the cadastral parcel-level, as well as territorial waters. These are assigned to the INSPIRE 'Cadastral parcels', 'Hydrography' (Annex I) and/or 'Sea regions' (Annex III). Related systems such as census districts, post office regions, and other sector-specific regions are not included, but they will contain a reference to national statistical units at the local level (LAU) and the Nomenclature of Territorial Units for Statistics (NUTS) established by Eurostat.

**Place names (or toponyms)** are an important reference framework used by individuals and societies throughout the world. Toponyms cover names of areas, regions, localities, cities, suburbs, towns or settlements, or any geographical or topographical feature of public or historical interest. Also, all these features can have different names in one or several languages. These names are often provided together with appropriate information on the feature in different products, like maps and gazetteers, and services. Due to their nature as location identifiers for cultural and physical features of the real world, the place names are often used as a proxy for other thematic areas, such as Buildings and Settlements. In some cases, names can be applied as attributes of appropriately modelled spatial objects in other INSPIRE themes. However, often the definition, classification, geometry and other attributes of these objects do not correspond with the respective named places in this definition sense, as in this case it focuses on names aspects. Besides, commonly named features such as elevations, islands, natural shoreline features and stretches of water bodies are seldom modelled as spatial objects in other themes, while they are considered as place names in here.<sup>104</sup>

**Addresses** are structured labels, usually containing a property number, a street name and a locality name that provide identification of the fixed location of properties. The basic unit of addressing is a building (a permanent construction) intended or used for the shelter of people, having at least one entrance from publicly accessible space. Together with coordinates indicating the geographic position, an address is used to identify a plot of land, a building or part of a building, or some other construction. The full address is a hierarchy consisting of components such as geographic names, with an increasing level of detail, e.g.: town, then street name, and then house number or name. It may also include a postcode or other postal descriptors, and sometimes could also include a path of access, depending on its function. Addresses are often used as a proxy for other data themes such as Land Parcels.<sup>105</sup> Different object types can be related to property. The most commonly recognised types with addresses are land parcels and buildings (including flats or apartments). However, other types of objects, such as street furniture, water pumping stations, parking lots, sports grounds, foothold, mooring places and agricultural buildings, can also have addresses. Although they do not receive post, they may need to have an address for other functions. Collectively, objects that can have addresses are referred to as addressable objects. The address is often defined so that it characterises the location of the related addressable object.<sup>106</sup> Although all national or local address systems share similar concepts and general properties, differences exist in formal and informal standards, rules, schemas and data models within Europe (some differences might appear in the extent of the address system, such as a more simplified version for rural areas, for example).

---

<sup>102</sup> Stakeholders' interviews and focus groups' discussions.

<sup>103</sup> <https://inspire.ec.europa.eu/theme/au:2>

<sup>104</sup> <https://inspire.ec.europa.eu/id/document/tq/gn>

<sup>105</sup> Stakeholders' interviews.

<sup>106</sup> <https://inspire.ec.europa.eu/id/document/tq/ad>

**Buildings** are covered facilities, usable for the protection of humans, animals, things or the production of economic goods and they refer to any structure permanently constructed or erected on the sites.<sup>107</sup> Settlements are collections of buildings and associated features where a community carries out socio-economic activities. Information on location of buildings may be supplied as points or with the actual basic form of the building itself. Usually buildings are part of cadastre. On the local level, buildings are available within the large-scale cadastral maps or cadastral data sets and are geometrically represented as surfaces. Most buildings can be identified (geo-coded) by addresses, as mentioned previously when discussed this dataset.

In the scope of INSPIRE Directive, the **Cadastral Parcels** focus on geographical part of cadastral data. Often, the cadastral parcels are mainly used as locators for geo-information, including within environmental data context. As much as possible, cadastral parcels should be forming a partition of national territory. However, this does not apply to specific rights as servitudes, which may only affect part of the parcel. In general, cadastre or land registry data are complemented with data gathered by other public agencies and institutions that register parts of the Earth's surface such as special domains, urban cadastres, public lands, which spatially complement the registrations by the main cadastre or land registry. The cadastral parcels have significant importance when it comes to protection of state lands, reducing land disputes and facilitating land reform. Other areas that use cadastral parcels are agriculture, land management, taxation, disaster management, real estate market and statistics. The datasets bring important contribution in land consolidation, spatial planning and infrastructure management, protection of soil and water.

The next section describes the overall benefits and value which can be attributed to the datasets discussed above, building on our value framework developed under Task 1.

### 2.3.2 Analysis of the value of these datasets

Geospatial datasets are infrastructural, as they serve a whole variety of purposes, typically in conjunction with other datasets. They are essential to a wide variety of activities, both private and public.

Geospatial datasets are used by the public sector for public tasks such as policy making, spatial planning, flood prediction and relief, emergency services, environmental assessments and many other applications. Robust geospatial data are essential to delivering public services.

The Covid-19 crisis highlighted the importance of geospatial data to address emergencies. Geospatial data have underpinned all data initiatives to monitor the virus spreading as well as the lockdown, from detecting hotspots to monitoring mobility. It is likely to play an even more important role in the reopening, to ensure that any surge in local infections is promptly detected and acted upon, and to adequately organise service provision to maintain social distancing.

Geospatial data play a fundamental role in the private sector too. Geospatial industries account for 400 billion dollars of revenues and play a key role in industries representing 75% of the global GDP. Indirect impact is way bigger: Boston Consulting Group estimated that revenues driven by geospatial services in the United States were around US\$1.6 trillion.<sup>108</sup>

The following table defines the specific benefits and aims to qualify and quantify their impact across the conceptual model developed by the project, based on the available evidence.

---

<sup>107</sup> <https://inspire.ec.europa.eu/id/document/tg/bu>

<sup>108</sup> Boston Consulting Group (2012), "Geospatial Services: A \$1.6 trillion growth engine for the U.S. Economy" Available at: <https://www.bcg.com/documents/file109372.pdf>

Macro characteristics	Value category	Assessment of value for the datasets in scope	Examples/references										
Economic	Competition		According to a study by the Aalto School of Engineering <sup>109</sup> on the effects of the opening of topographic data by the National Land Survey in Finland, SMEs are the user group with largest growth after data were opened.										
	Consumer benefits		A recent study estimates the value of digital maps at 105 USD per user per year, for a total benefit of about 420 billion euros per year in Europe. In addition, digital maps reduce travel time by 12% on average. <sup>110</sup> The Swedish meta-study mentioned by the interviewee finds that 4 different data categories/datasets could bring benefits to the Swedish society of approximately 11.1 billion SEK yearly.										
	Economic output		In Spain, geographic information sector is the most important one regarding infomediary sector. There are 166 companies (22% of the total), almost 550 M sales (26.6%) and more than 6 800 employees (30%). <sup>111</sup>										
	Employment												
	Productivity and commercialisation		Swedish meta-study (reported by interviewee) identified potential yearly benefits in these sectors together of about 10 billion to 21 billion SEK: <table border="0"> <tr> <td>Agricultural Sciences</td> <td>1.1 – 2.9 billion</td> </tr> <tr> <td>Information and Tech</td> <td>2.6 – 6.4</td> </tr> <tr> <td>Finance and Insurance</td> <td>2.0 – 3.3</td> </tr> <tr> <td>Spatial Planning</td> <td>3.0 – 6.2</td> </tr> <tr> <td>Public Sector</td> <td>1.1 – 1.7</td> </tr> </table> <p>Geospatial services are deemed to have a significant productivity impact in sectors covering 75% of global GDP.<sup>112</sup> In agriculture, net benefit of \$75 per hectare in the cropping industry from use of precise positioning and GIS in the grains industry.<sup>113</sup></p>	Agricultural Sciences	1.1 – 2.9 billion	Information and Tech	2.6 – 6.4	Finance and Insurance	2.0 – 3.3	Spatial Planning	3.0 – 6.2	Public Sector	1.1 – 1.7
Agricultural Sciences	1.1 – 2.9 billion												
Information and Tech	2.6 – 6.4												
Finance and Insurance	2.0 – 3.3												
Spatial Planning	3.0 – 6.2												
Public Sector	1.1 – 1.7												
Innovation & AI	Citizen innovation		Citizens are the segment with the largest increase in users (+150%) of topographic data, once they were opened up in Finland. Services for instance in the field of tourism are, for example Wikiloc, Komoot.										
	Public sector innovation		Public sector reuses topographic data mostly as base map for internal usage. <sup>114</sup>										
	Entrepreneurialism and private sector innovation		29% of SMEs and 26% of large companies use data to refine products and services, the rest being mostly for internal use.										
	AI		The Swedish meta study shows that datasets that are included in the Swedish proposal have good potential for being used in AI applications. The datasets have all the components that are required when it comes to AI. The datasets are updated, they are possible to link with other										

<sup>109</sup> Jaana Mäkelä, Paula Ahonen-Rainio and Kirsi Virrantaus "Effects of open topographic data in Finland, A user study one and half years after the opening" Dept of Real Estate, Planning and Geoinformatics, Aalto School of Engineering, 2014.

<sup>110</sup> AlphaBeta, 2017. The Economic impact of geospatial services: How consumers, businesses and society benefit from location-based information.

<sup>111</sup> ASEDIE, '2020 Infomediary Sector', 2020.

<sup>112</sup> AlphaBeta, 2017. The Economic impact of geospatial services: How consumers, businesses and society benefit from location-based information

<sup>113</sup> Acil Allen Consulting and CRC for Spatial Information, 2017. Economic Value of Spatial Information in NSW.

<sup>114</sup> According to a study by the Aalto School of Engineering on the effects of the opening of topographic data by the National Land Survey in Finland.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			datasets, they have a high level of details, some of the dataset are in time series, the datasets are well structured and have metadata, and finally most of the datasets are available via APIs AI can detect changes in the territory and improve satellite images resolution, as in this Spanish example: <a href="https://www.grafcan.es/ia_territorial_en">https://www.grafcan.es/ia_territorial_en</a>
Public sector	Public services management		26M EUR savings per year in improved management of public assets. 7M EUR savings in planning in municipalities. <sup>115</sup>
	COVID		The Technical University of Cyprus has released recently an interesting and successful platform related to COVID-19 <sup>116</sup> . It uses spatial data to report the status of the pandemic in Cyprus on a continuous basis, providing analytical numbers, graphs interactive maps, and population concentrations, working from country to community level. The platform is being updated and upgraded on a regular basis. The emergency due to the spread of Covid-19 Pandemic has increased the need of having epidemiological data available in real time and in geocoded format.
Re-use	Demand for information		Geospatial data underpin reuse and integration with many different datasets – they enable wider reuse.
Climate change	Environment management		Crucial to disaster management and planning, for instance for flood, when timeliness is essential, with the possibility to reduce human and economic costs by 40/50% with improved and more timely information. <sup>117</sup> Experts at the National Land Survey of Finland have identified a short list of use cases for LOD2 3D building data: flood risk analysis is the most important application domain in this area. Another study attributes a reduction in response time of 20% thanks to improved geo-services, especially on flood risks. Geo-based services are a fundamental tool for more efficient fleet management and reduced congestion.

<sup>115</sup> Acil Allen Consulting and CRC for Spatial Information, 2017. Economic Value of Spatial Information in NSW.

<sup>116</sup> Stakeholders' interview.

<sup>117</sup> United Nations Office for Outer Space Affairs (UNOOSA), 2013. The Value of Geoinformation for Disaster and Risk Management.

## 2.4 Meteorological data

In this section, the report provides a detailed overview of the datasets in scope of the analysis for this thematic area, and maps the main benefits and use cases of these data fields.

### 2.4.1 Identification of the datasets in scope

Meteorology relies upon weather and climate information. There is one main EU level piece of legislation regulating access to data in this area and this is the INSPIRE Directive (Directive 2007/2/EC)<sup>118</sup>, which defines **meteorological geographical features** as one of the 34 spatial data themes needed for environmental applications and to be made available by public sector organisations.

The table below summarises the most important provisions of this legislative act in terms of obligation for Member States to make available certain datasets and services.

Table 10 - Legislative scope of meteorological legislation

---

#### INSPIRE Directive

---

Article 7 requires Member States to adopt “implementing rules laying down technical arrangements for the interoperability and...harmonisation of spatial data sets and services”. Furthermore, they must “ensure that all newly collected and extensively reconstructed spatial data sets and corresponding spatial data services are available with the implementing rules”.

---

In accordance with article 10(1), Member States must “ensure that any information... needed for compliance with the implementing rules is made available to public authorities or third parties”

---

Article 11(1) requires Member States to establish and operate a network of certain services, which must take into account relevant user requirements and be easy to use, available free of charge (article 14) and accessible to the public. Article 13 provides valid reasons for public access limitation.

---

Article 17 states that Member States must enable public authorities to gain access, exchange and use spatial data sets and services.

---

The EU legislator has only regulated spatial data sets and services within this theme. However, in addition to INSPIRE, specific entities and organisations working in this domain have made their own agreements and resolutions which help defining the scope of this thematic area. At this point it is important to bear in mind that this Impact Assessment study is only considering high-value datasets made available by Member States. The table below summarises the most important provisions of agreements made by entities active in the meteorological thematic area, namely:

- EUMETNET (the grouping of 31 European National Meteorological Services)<sup>119</sup>;

---

<sup>118</sup> Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1581591451478&uri=CELEX:32007L0002>

<sup>119</sup> <https://www.eumetnet.eu/>



- Copernicus Services<sup>120</sup>
- European Centre for Medium-Range Weather Forecasts (ECMWF)<sup>121</sup>
- EUMETSAT<sup>122</sup>
- European Environment Agency (EEA)
- World Meteorological Organization (WMO)<sup>123</sup>

---

<sup>120</sup> <https://www.copernicus.eu/en/services>

<sup>121</sup> <https://www.ecmwf.int/>

<sup>122</sup> <https://www.eumetsat.int/website/home/index.html>

<sup>123</sup> <https://public.wmo.int/en>

Table 11 - Meteorological legislation

<b>License agreement for the use of data and/or products for the Copernicus Services<sup>124</sup></b>	<b>License agreement: License to Use Copernicus Products by ECMWF<sup>125</sup></b>	<b>WMO Resolution 40 (Cg-XII): WMO policy and practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities<sup>126</sup></b>	<b>EUMETSAT Council Resolution on amendments to the data policy implementing rules in reconciliation with the objectives of the pathfinder projects<sup>127</sup></b>
<p>This agreement was signed between EUMETNET (the licensor) and The EEA and Copernicus Service Operators (the licensee)</p>	<p>This agreement was signed between ECMWF (the licensor) and Copernicus Services</p>	<p>This resolution was established at the twelfth conference of the WMO</p>	<p>This resolution was agreed upon by the participating states at the EUMETSAT Council.</p>
<p>According to articles 2 and 3, the licensee is authorised to use on a non-exclusive basis, the following data and/or products:</p> <ul style="list-style-type: none"> <li>• meteorological</li> <li>• hydrological</li> <li>• climatological</li> </ul> <p>in order to help fulfil the terms of their respective delegated</p>	<p>Article 4.2 states that: "Access to Copernicus Products is given for any purpose in so far as it is lawful, whereas use may include, but is not limited to: reproduction; distribution; communication to the public; adaptation, modification and combination with other data and information; or any combination of the foregoing."</p>	<p>EUMETSAT "Adopts the following practice on the international exchange of meteorological and related data and products:</p> <p>(1) Members shall provide on a free and unrestricted basis essential data and products which are necessary for the provision of services in support of the protection of life and property and the well-being of all nations, particularly those basic data</p>	<p>Article 4 states that: EUMETSAT shall make its Hourly Meteosat Data, all Derived Products and Advanced Image Products available to all users world-wide on a free and unrestricted basis, regardless of when and how these are made available to the user, as "Essential" Data and Products in accordance with WMO Resolution 40 (Cg-XII).</p>

<sup>124</sup> Copernicus Services Operators, "EEA and EUMETNET sign agreement on the provision of hydrological, meteorological and climatological data for the Copernicus Services", available at <https://insitu.copernicus.eu/news/eea-and-eumetnet-sign-public-duty-license-agreement-for-data-provision-to-copernicus> (last visited on 13 February 2020).

<sup>125</sup> ECMWF, "License to Use Copernicus Product", available at <https://apps.ecmwf.int/datasets/licences/copernicus/> (last visited on 13 February 2020).

<sup>126</sup> World Meteorological Organization, "Twelfth World Meteorological Congress", Geneva 30 May-21 June 1995 (WMO-No 827).

<sup>127</sup> EUMETSAT, "Resolution on amendments to the data policy implementing rules in reconciliation with the objectives of the pathfinder projects", January 2019 (Council Resolution EUM/C/89/18/Res. I).

---

agreements.

and products, as, at a minimum, described in Annex 1 to this resolution, required to describe and forecast accurately weather and climate, and support WMO Programmes;

(2) Members should also provide the additional data and products which are required to sustain WMO Programmes at the global, regional, and national levels and, further, as agreed, to assist other Members in the provision of meteorological services in their countries. While increasing the volume of data and products available to all Members by providing these additional data and products, it is understood that WMO Members may be justified in placing conditions on their re-export for commercial purposes outside of the receiving country or group of countries forming a single economic group, for reasons such as national laws or costs of production;

(3) Members should provide to the research and education communities, for their non-commercial activities, free and unrestricted access to all data and products exchanged under the auspices of WMO with the

---

understanding that their commercial activities are subject to the same conditions identified above”

---

Article 6 provides the limitation of use. The following uses are prohibited:

commercial exploitation  
business for-profit use  
resale  
reconstruction through reverse engineering or other techniques  
assignment of rights  
sub-licensing

According to Annex I: “The meteorological and related data and products which are essential to support WMO Programmes include, in general, the data from the RBSNs and as many data as possible that will assist in defining the state of the atmosphere at least on a scale of the order of 200 km in the horizontal and six to 12 hours in time.”

Articles 5 to 11 provide a list of conditions for access to non-essential meteosat data for:

NMS,  
Research Projects and Educational Users,  
European Centre for Medium-Range Weather Forecasts (ECMWF)  
Commercial Users

The long list of potential high value data that emerged from the discussions held with the stakeholders and the initial analysis of the domain presented above was relatively clear: diverse stakeholders point to the same types of data practically in consensus. Both data holders and re-users see a defined value chain for meteorological data, stretching from observation data to modelling, computing, forecasting to delivery of services.

The list of HVD which was suggested for further analysis includes:

- Observations data, these are the many different variables measured by (automated) weather stations.
- Climate data, both historic climate data, and the daily quality controlled/validated hourly and daily observations data that are added and become part of the permanent record
- Radar data, which is used to e.g. detect precipitation and wind
- Weather alert messages, which are part of the core public task of national weather services, that go out to both the general public and specific user groups (such as aviation)
- Numerical weather prediction (NWP) model output, which is calculated data from mathematical models (with a.o. observations as input), to be further used to create weather forecasts
- Lightning strike data, which provides insight into the frequency, location and intensity of lightning strikes

This list is roughly in the order in which various stakeholders expressed a preference. Those preferences are more or less based on the order of elements in the perceived value chain: it all starts with measurements and observations. Logically this means that observations data, and its validated form as climate data, are of prime interest to most re-users and sectors. Model data in some contrast, is very much seen as highly valuable but also as of use to a somewhat more limited group of professional meteorological re-users.

Currently Finland, Germany, Iceland, Ireland, Luxembourg, Netherlands, Norway, Slovenia and Sweden publish open meteorological data. Charges apply to observations data for instance in Austria, Belgium, Croatia, Cyprus, Czech Republic, Denmark, France, Greece, Hungary, Latvia, Poland, Portugal, Romania (and Switzerland), according to EUMETNET, the collaborative network of European national weather services. Some countries are in ongoing discussions to move to open data or are in the starting phase (e.g. Hungary, Belgium, Czech Republic). Denmark is currently in a four year transition process towards providing the data collected for re-use for free at the point of use. Within the list of EU countries, Malta has a different situation. Its meteorological services are provided by a publicly traded private company, the Malta Airport Metoffice on behest of the Maltese government.

Concerning re-use therefore, re-users see a big variation in current availability of observational data for re-use. At the same time complicated licensing structures play a role (e.g. differentiation on the purpose of re-use, and where combined data from multiple sources carrying different licenses for sections of the data. Current pricing is seen by commercial re-users as a high entry barrier to market. Fee levels vary strongly, e.g. up to a factor of 21 between the highest and lowest price for national observations data in the EU (whereas as mentioned for other MS such observation data is available free of charge).

For some of the long-listed datasets, specifically radar data and NWP model data, the datasets can get quite voluminous. Open standards for scientific data formats (such as HDF5) exist and are in use to help keep the volume of datasets low, such as is the case for radar data. However technological advances are leading to higher volumes of data. In the case of NWP model output, most of the current data is 'deterministic' (i.e. one run of a model), but there is a development towards 'ensemble' NWP model output, which is the same model run many times over with slightly different starting conditions. Such output can easily be a factor 50 larger than the deterministic variety. The result is very high volumes of

data (tens of Terabytes daily), although these can be significantly brought down for data provision by selecting parameters and (vertical) layers that are of meteorological relevance (not every parameter is relevant for use at every layer. E.g. humidity at height might be less relevant than wind speeds at heights relevant for wind energy parks). Similarly current developments see an increase in more fully automated observation stations, and movement towards higher measurement and provision frequencies up to (near) real time. These developments too mean an increase in the data volume concerned. This is creating demand for cloud based access and a shared data space. Such volumes mean that the delivery method is a potential issue in terms of marginal costs, and both bandwidth demands and the need for cloud service accounts introduce new barriers to access and re-use, even if the data itself is free at the point of use.

#### **2.4.2 Analysis of the value of these datasets**

The core public task of national weather services is to 'protect life and property', meaning to reduce the unwanted and unanticipated impacts of weather conditions. In general this means that the more meteorological data is used in decision making the more likely that public task can be fulfilled. Reaching wider groups of stakeholders through a diversity of channels is enabled by open data. It is unlikely that all areas of public interest can all be covered by a national weather service's own activities and outreach. Thus national weather services can extend their reach and engagement using open data as an instrument that allows others to act, provide services, channels and further uses for this data.

The meteorological data market is a global one, as weather and climate are global phenomena. This also allows for comparison between the US and European markets, as the US market has had access to public sector meteorological free to re-use data for much longer. A recent World Bank study (January 2020) comparing the US and UK markets highlights a diverse market ecosystem and overall bigger market when base data is available without barriers. Existing literature and cases further suggest that, for meteorological data, price elasticity is well above one, meaning that when fees are reduced or abolished there is a strong non-linear rise in demand. This points to strong latent demand, and fees charged forming an active threshold to entry for new re-users.

Aspects of the data itself that impact use value are timeliness (some data is time critical for certain uses), and granularity in both time and geography. As mentioned, the volume of the data might be a challenge when looking at how to create value from meteorological HVDs, by way of potentially raising marginal provision costs.

Concerning re-use, re-users see a big variation in current availability of observational data. At the same time complicated licensing structures play a role (e.g. differentiation on the purpose of re-use, and where combined data from multiple sources carrying different licenses for sections of the data. The same variation is visible when it comes to fees being levied at the point of use. Current pricing is seen by commercial re-users as a high entry barrier to market. Fee levels vary strongly, e.g. up to a factor of 21 between the highest and lowest price for national observational data in the EU (and of course in other MS such observation data is available free of charge). Abolishing fees results in loss of revenue for the public data holder. Such fees are sometimes used to fund infrastructure (such as automated observation stations), and to fund increasing need for (super)computing capacity. Typically loss of revenue in these cases requires compensation from the general national budget.

The above long-listed meteorological data has value to a wide variety of sectors. For example, as mentioned by data holders and other stakeholders alike, the energy sector, building and construction, infrastructure management, disaster management and civil protection, media, sports, transport and logistics, tourism, it, research, the insurance industry, flood risk management, environmental protection, climate adaptation efforts, and agriculture (for which climate data is becoming more important, due to changes in growing seasons). When looking at these data from the perspective of the HVD framework

developed, it suggests that they are relevant for all the macro categories of value as they have a strong potential to bring economic, social and environmental benefits, a strong potential for reuse, they allow generation of new and innovative applications and they benefit the public sector as well. The literature offers many examples of how these benefits are realised. For instance, there is strong and long standing evidence that the above mentioned meteorological information help predicting forest fires<sup>1</sup> and pushing for sustainable agriculture.<sup>2</sup> Furthermore, evidence suggests that meteorological data combined with transport data give transport operators and the traveling public improved insight into the anticipated impacts resulting from adverse weather conditions or climate change, enabling more effective mitigation strategies.<sup>3</sup> Finally, it is demonstrated meteorological data can be used for disease control and prevention.<sup>4</sup> As these few use cases suggests, the potential HVD in the meteorological thematic area tick many of the boxes of our HVD framework.

As no country or organisation can go it alone on meteorological data, maximum value is perceived if data and research are as widely shared as possible. This also translates into perceiving data unified across countries as more valuable, harmonization being a value added activity (which is also being outsourced), and having a single point of access (Which e.g. GTS of the WMO aims to be) as being important for ease of use. All three are friction reducing steps. While it is not the aim of the PSI Directive to introduce harmonisation of data or a single point of access as INSPIRE does, the PSI Directive is seen by both data holders and re-users as an opportunity to make the first steps of friction reduction, which is removing the variety in licensing and fees, making access to similar data across the EU much easier and removing barriers to entry. With regard to INSPIRE this would add the HVD open data requirements formulated in this report to the INSPIRE download services for the datasets involved.

The list of sectors where meteorological data is of valuable use is very long, and the use cases vary just as much. However some groupings are possible to make, not on the diversity of sectors involved, but on the type of usage. These different types of use cases are apparent from both the examples of re-use encountered, and mentioned by both data holders and re-user. They may take place within the same sector in parallel, and multiple usage types may be relevant to the same re-user even. These types of use cases spell different needs with regard to the available open data.

A first distinction is that between use cases that depend on timeliness of information, and use cases that depend on the validity of information. Timeliness of information delivery is key everywhere where instant decisions need to be taken. Either to prevent damage (e.g. dealing with heavy rainfall by waste water management facilities), for responsive systems (e.g. building management, internet of things), or to be able to estimate the impact of the current situation (e.g. current expected solar output of your roof panels, events, sports), mitigation (e.g. road management), or to use for guidance in practical decisions (e.g. agriculture). Timeliness of forecasting is key in for instance the wind energy sector, with large financial interests involved in the decisions taken on these forecasts. These type of use cases all need real time data from (automated) observations or radar, and/or dependable frequent availability of up to date forecasts (based on NWP model output). That real time data needs to include all measured variables to cover the various use forms (e.g. in observation data 'classic' elements like temperature and humidity are still important, but other variables such as solar radiation have increased in importance for re-use over time, in this example for solar energy related use cases).

For a different group of use cases validated data, and data over longer time series is key. This is the case for insurance companies for instance, or in the energy sector to determine the likely output of solar or wind energy plans for a given location, as well as for agricultural planning, and climate adaptation efforts.

A second distinction to typify use cases is that between use cases that are (longtime) part of the meteorological service market, and new entrants and novel uses which are not necessarily meteo

companies or even meteo oriented uses. The meteorological service market is providing e.g. forecasting services for the general public or specialised forecasts for specific user groups (e.g. sailing or specific industries), as well as being intermediaries to larger re-user groups. Some MS in our conversations with them report that this market seems rather stable to them, whereas they see more dynamism in other areas, both in terms of the number of companies involved, and the value involved. In this latter group fall data oriented companies that are not meteo companies, and use cases are often based on real time data, that is being used as part or context for non meteo services.

A third distinction is that between use cases increasing efficiency for re-users, and use cases that lead to new value creation.

Denmark in preparation for their current transition to a full open data policy, in 2016 estimated the impact of releasing meteorological data. For three sectors (electricity, district heating networks and agriculture) they quantified the efficiency gains in existing workflows from providing open data, between 6.7 and 18 million Euro annually, with agriculture seeing the highest efficiency gain. It also noted likely impacts from increasing competition, and value created for these three sectors based on additional uses of the data (e.g. To reduce the amount of pesticides used in agriculture), but did not quantify that impact. The study also expects impact in the insurance, media, retail and contracting industry, as well as data services, but did not quantify that impact.<sup>128</sup> These efficiency gains in these three specific sectors only already exceed the expected annual revenue loss of the DMI (805 thousand Euro) by at least a factor 8, and also exceed the sum of the loss of revenue and the cost of DMI’s ongoing open data project (5.4M Euro, in total), on these use cases alone by up to a factor 3.

The Denmark 2017 Rombøll study estimated DMI’s meteo data’s potential for water related services at 14 million Euro annually, which also exceeds the costs of DMI’s open data effort on its own by more than a factor 2. The 2011 POPSIS study into the impact of the then PSI Directive documented private sector employment growth, much higher numbers of re-users (often SMEs) and increasing tax revenue for the cases of the Norwegian, Dutch and Slovenian steps to either free open data or provision at marginal costs. This non-linear rise in re-use, leading to tax revenue over time exceeding the costs of data provision is a common pattern in these cases.

In the following table examples and references for observable impact from the existing wide variety of use cases are provided for different macro categories of value creation. Though currently not evenly spread out over all MS, the overall pattern that stands out across these categories is that for the data sets in scope, the various measurements, historic and model output data, the socio-economic value of opening these data sets is very high, impacting a wide variety of sectors, allowing strong growth of re-use from existing and new companies as well as citizens, with additional tax revenue from those activities easily growing beyond provisions costs or reduction of revenue from data fees by multiples.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
Economic	Competition		Germany’s NWS (DWD) is in the midst of stakeholder research amongst re-users and first results indicate a

<sup>128</sup> Deloitte study for DMI, 2016



Macro characteristics	Value category	Assessment of value for the datasets in scope	Examples/references
			growing and dynamic market, with a growing number of market entrants. Reduction and removal of barriers, like data fees, were shown in the 2011 POPSIS study to significantly increase the number of new commercial re-users for meteorological data.
	Consumer benefits		Finland's FMI in their 2019 stakeholder review sees a variety of use cases emerging for consumer centric services, including rain radar, local weather on mobile, and specifically for different sports (skydivers, surfers) as well as household solar energy planning and forecasting (using solar radiation data). Wider variety of services and apps available, for both broader and niche purposes. A commercial Dutch rain radar app/site is the most used one in the country, e.g. to decide whether to commute by bicycle (and stay dry). When the Finnish FMI opened up their data in 2013, a report estimated benefits the benefit to agriculture, tourism and recreation to at least double from when data was available against a fee. <sup>129</sup>
	Economic output		Over the period 1993-2006 the US market for meteorological services, where NWS data are freely available, grew well over twice as fast (~17%) per year, compared to the EU market (~6.5%) which had extensive data charges and restricted availability, resulting in a market roughly twice the size, while the US and the EU had a comparable overall GDP <sup>130</sup> .  In a 2019 stakeholder survey by the Finnish NWS (FMI), 54 companies reported that open meteo data had enabled new business for them in the previous three years.
	Employment		Growth in employment both for new and existing re-users were shown for the Netherlands (rising by a factor 3 after removal of data fees) and Norway. The German NWS (DWD) in their current stakeholder survey indicates a rising number of start-ups and SMEs are involved with re-using meteo data.

<sup>129</sup> BearingPoint, 2013, for FMI

<sup>130</sup> Pettifer, R., 2008

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
	International competitiveness		The 2011 POPSIS study demonstrated how Norwegian open data resulted in a growing number of SME's from elsewhere in Europe integrating meteo data in their (often media related) services. Similarly the ongoing German stakeholder survey indicates more international companies engaging with their open data than before.
	Product market dynamism		The range of use cases encountered, especially those use cases that are not traditional meteo services, point to a growing diversity of products and services using meteo data. <sup>131</sup> Germany's NWS (DWD) is in the midst of stakeholder research amongst re-users and first results indicate a growing and dynamic market, with a growing number start-ups emerging besides established providers of meteorological services.
	Productivity and commercialisation		Turnover rise in the Netherlands was shown to be a factor 4 for private sector re-use when data fees were removed gradually between 1999 and 2009.
	Public-private coordination		The steady availability of open meteo data has a likely impact on the number and variety of research projects that use meteo data, as indicated by a number of MS.
Climate change	Citizen engagement in addressing climate change		Citizen science communities are emerging (e.g. Germany, Norway, Netherlands <sup>132</sup> ) using meteorological data in comparison with own measurements, in the context of local climate adaptation efforts.
	Energy management and efficiency		Many use cases referred to by NWSs point to the relevance of meteorological data (observations and predictions, climate data) for green energy production. Both large scale for planning facilities (expected output for a location) and operations (e.g. predicting output in the next 24 hours to bring to market), as well as (and increasingly small scale (e.g. individual households planning solar panels, using solar radiation observations). A 2016 report for the Danish

<sup>131</sup> Stakeholders input

<sup>132</sup> E.g. <https://meetjestad.nl>, <https://www.buergerschaffenwissen.de>

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			DMI estimates the efficiency increase from open data for the electricity and distance heating sectors between 800k to 4 million Euro annually. <sup>133</sup>
	Environment management		The agricultural sector is a key (re-)user of meteorological data. A 2016 report for the Danish DMI estimates the efficiency increase from open data for agricultural sector between 3.5 million to 14.1 million Euro annually (including by environmental modelling of the effects of e.g. water and pesticides usage) <sup>134</sup> Multiple MS report use cases concerning environmental protection, nature conservation, and climate change analyses.
Innovation & AI	Citizen innovation		Citizen science communities are emerging (e.g. Germany, Norway, Netherlands <sup>135</sup> ) using meteorological data in comparison with own measurements, in the context of e.g. climate adaptation.
	Entrepreneurialism and private sector innovation		Austrian company UBIMET (founded 2004, over 200 employees), uses AI technology combining all types of available meteo data into high-resolution models for hyper-localised weather information. In a 2019 stakeholder survey by the Finnish NWS (FMI), 54 companies reported that open meteo data had enabled new business for them in the previous three years
Public services and administration	Public sector and public revenue		Additional tax revenue was shown to be higher than the loss of revenue and increase in provision costs for both the NWS of the Netherlands and Norway, with that difference growing over time. (Although that benefit typically does not accrue partly or entirely with the data holder, but with the general government.) <sup>136</sup> Had in 2009 the EU market for meteorological data re-use been as developed as the US market (where such data is available for free), the annual

<sup>133</sup> Deloitte, 2013, for DMI

<sup>134</sup> Deloitte, 2013, for DMI

<sup>135</sup> E.g. <https://meetjestad.nl>, <https://www.buergerschaffenwissen.de>

<sup>136</sup> De Vries et al, 2011

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			treasury revenue would have been at least 300 million euro higher in the EU. <sup>137</sup>
	Public services management		Multiple NWSs report an increase in quality and depth of questions received from re-users / the public, where professional level /technical questions have become more important.
Re-use	Demand for information		NWSs that have opened up their data (e.g. Finland, Netherlands, Norway, Germany) typically see the demand for their data rise non-linearly upon the removal of fees, and then linearly. Current demand (March 2020) in Germany came to 2.4 billion requests on their server, with over 400 TB of data downloaded.
	Trust and confidence in information		The Norwegian NWS, Met.no, which provides all their data as open data since 2007 concluded the direct link to the general public this provided worked as a quality assurance. This consolidated the business case for their open data policy as it underlined the importance of being seen as actively fulfilling their public task reliably in the eye of the public <sup>138</sup>
Social	Disease prediction and prevention		Meteorological data can be used to help chart, mitigate and predict the spread of diseases, e.g. for Dengue fever. While Dengue fever is not an issue in Europe, the current uncertainties around the way the Corona virus spread is influenced by temperature and humidity also point to a role for meteorological data. <sup>139</sup>
	Mobility efficiency and Mobility		When the Finnish FMI opened up their data in 2013, a report estimated benefits from the availability of data for traffic purposes (car, pedestrians, cyclists, and water) of 51

<sup>137</sup> Pettifer, 2009

<sup>138</sup> Pricing of Public Sector Information Study, POPSIS, 2011, Deloitte et al.

<sup>139</sup> Ramadona AL, Lazuardi L, Hii YL, Holmner Å, Kusnanto H, Rocklöv J (2016) Prediction of Dengue Outbreaks Based on Disease Surveillance and Meteorological Data. PLoS ONE 11(3): e0152688. <https://doi.org/10.1371/journal.pone.0152688>

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
planning			million Euro, a factor 22 higher than the loss of revenue involved for FMI. <sup>140</sup>
Public engagement and government transparency understanding			The Norwegian NWS, Met.no, which provides all their data as open data since 2007, actively disseminated all of its information to the general public, creating a more active relationship with the public it serves, and increasing the visibility and transparency of Met.no's work. <sup>141</sup>

## 2.5 Earth observation and environment

In this section, the report provides a detailed overview of the datasets in scope of the analysis for this thematic area, and maps the main benefits and use cases of these data fields.

### 2.5.1 Identification of the datasets in scope

Earth observation as observing the planet's physical, chemical and biological status over time, can be seen as the combination of satellite remote sensing, earth based remote sensing, and in-situ data collection regarding the environment. Environmental data concern both the status of the environment (in the physical, chemical and biological sense), as well as human activities impacting that status (either administrative or regulatory aspects such protective measures or administrative boundaries or allowed levels, as well as interventions in the physical environment such as waste or emissions, or flood prevention activities).

Although this Impact Assessment study is only considering high-value datasets made available by Member States, nevertheless, data used in EU level initiatives (like the Copernicus programme) are regularly provided by Member States. Therefore legislation regulating Earth Observation (hereinafter: EO) and environment data for EU level programmes are still relevant in determining which data sets are considered high value data for the EO and environment theme. The key EU legislations regulating the EO and environment thematic area are:

- The INSPIRE Directive (Directive 2007/2/EC)<sup>142</sup>, which governs several spatial data themes relevant to the environment thematic area of this study.

<sup>140</sup> BearingPoint, 2013, for FMI

<sup>141</sup> Pricing of Public Sector Information Study, POPSIS, 2011, Deloitte et al.

<sup>142</sup> Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1581591451478&uri=CELEX:32007L0002>

- Directive 2003/4/EC<sup>143</sup> on public access to environmental information, aiming to guarantee access to environmental information held by or for public authorities. The Directive is a consequence of the Aarhus Convention.<sup>144</sup>
- The Copernicus Regulation (Regulation (EC) No 377/2014)<sup>145</sup> which establishes Copernicus, the Union Earth observation and monitoring programme and lays down the rules for its implementation;
- The Commission Delegated Regulation concerning GMES (Commission Delegated Regulation (EU) No 1159/2013)<sup>146</sup> which concerns GMES services and data; and
- The Horizon 2020 Regulation (Regulation (EC) No 1291/2013)<sup>147</sup> which establishes Horizon 2020 – the Framework Programme for Research and Innovation (2014-2020).

The table below (Table 16) provides an overview of the relevant provisions from the INSPIRE Directive. Table 17 summarises the main provisions of the Regulations mentioned above, laying down rules for data access and reuse.

Table 16 - Earth observation and Environment legislation

---

### INSPIRE Directive

---

The relevant spatial data themes INSPIRE refers to regarding EO and Environment are:

Annex I: (8) Hydrography and (9) Protected areas

Annex II: (2) Land cover (3) Ortho-imagery and (4) Geology

Annex III: (3) Soil (4) Land use (7) Environmental monitoring facilities (11) Area management/restriction/regulation zones and reporting units (12) Natural risk zones and (19) Species distribution

---

Article 3(2) defines 'spatial data' as "any data with a direct or indirect reference to a specific location or geographical area"

---

Article 7 requires Member States to adopt "implementing rules laying down technical arrangements for the interoperability and...harmonisation of spatial data sets and services". Furthermore, they must "ensure that all newly collected and extensively reconstructed spatial data sets and corresponding spatial data services are available with the implementing rules".

---

In accordance with Article 10(1), Member States must "ensure that any information... needed for compliance with the implementing rules is made available to public authorities or third parties"

---

<sup>143</sup> Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC

<sup>144</sup> Convention on Access to Information, Public Participation in Decision-Making and Access to Justice in Environmental Matters, *Aarhus Denmark*, 25 June 1998.

<sup>145</sup> Regulation (EU) No 377/2014 of the European Parliament and of the Council of 3 April 2014 establishing the Copernicus Programme and repealing Regulation (EU) No 911/2010, <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1581599872382&uri=CELEX:32014R0377>

<sup>146</sup> Commission Delegated Regulation (EU) No 1159/2013 of 12 July 2013 supplementing Regulation (EU) No 911/2010 of the European Parliament and of the Council on the European Earth monitoring programme (GMES) by establishing registration and licensing conditions for GMES users and defining criteria for restricting access to GMES dedicated data and GMES service information, <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1581590073837&uri=CELEX:32013R1159>

<sup>147</sup> Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020) and repealing Decision No 1982/2006/EC, <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1581601294555&uri=CELEX:32013R1291>

---

Article 11(1) requires Member States to establish and operate a network of certain services, which must take into account relevant user requirements and be easy to use, available to the public. Article 13 provides valid reasons for public access limitation. Article 14 requires the discovery service of Article 11(1a) to be free of charge, as well as the viewing service (Art 11(1b)) although charges still may apply for viewing, and re-use for commercial purposes may be prevented. Downloads and access to spatial data services, may be charged for.

---

Article 17 states that Member States must enable public authorities to gain access, exchange and use spatial data sets and services.

---

Table 12 - Earth observation and Environment legislation

<b>The Directive on access to environmental information</b>	<b>The Copernicus Regulation</b>	<b>The Commission Delegated Regulation concerning GMES</b>	<b>The Horizon 2020 Regulation</b>
<p>Article 2 provides the definition of “environmental information”: a) the state of the elements of the environment b) factors, such as substances, energy, noise, radiation or waste c) measures affecting or likely to affect the elements and factors referred to in (a) and (b) (d) reports on the implementation of environmental legislation; (e) economic analyses and assumptions used within the framework of the measures and activities referred to in (c); and (f) the state of human health and safety</p>	<p>Article 4(2) requires Copernicus to deliver “accurate and reliable data and information to Copernicus users...responding to the requirements of Copernicus core users.” Furthermore, Copernicus must provide “sustainable and reliable access to spaceborne [and in situ] data and information”. The achievement of these objectives, according to article 4(3) is measures following the “use of Copernicus data and Copernicus information”. Article 23 (1) therefore encourages “promoting the use and sharing of Copernicus data and Copernicus information”</p>	<p>Article 3 lays down the open dissemination principles: “Users shall have free, full and open access to GMES dedicated data and GMES service information under the conditions laid down in Articles 4 to 10, subject to the restrictions laid down in Articles 11 to 16.”</p>	<p>Annex I of this Regulation indicates that “Activities shall focus on the capabilities, technologies and data infrastructures for Earth observation and monitoring from both remote sensing and in situ measurements that can continuously provide timely and accurate information and permit forecasts and projections. Free, open and unrestricted access to interoperable data and information will be encouraged. Activities shall help define future operational activities of the Copernicus programme and enhance the use of Copernicus data for research activities.</p>
<p>Article 3 requires Member States to “ensure that public authorities are required to make available environmental information held by or for them to any applicant at his request and without his having to state an interest.”</p>	<p>Article 23 (2) states that “Dedicated mission data and Copernicus information shall be made available through Copernicus dissemination platforms, under pre-defined technical conditions, on a full, open and free-of-charge basis, subject to [specified limitations].”</p>	<p>Article 7 lays down conditions regarding use. It states that: “Access to GMES dedicated data and GMES service information shall be given for the purpose of the following use in so far as it is lawful:</p> <ul style="list-style-type: none"> <li>a. Reproduction</li> <li>b. Distribution</li> <li>c. communication to the public</li> <li>d. adaptation, modification</li> </ul>	



- 
- and combination with other data and information
  - e. any combination of points a. to d.
- 

Article 24 enables the Commission to set conditions and limitations of Copernicus data and Copernicus information access and use.

---

The EU legislation discussed so far is rather general in nature. The INSPIRE Directive leaves a certain level of ambiguity when identifying which data constitutes 'spatial data' and the Directive on access to environmental information provides a broad definition of 'environmental data'. Other EU legislation, on the other hand, regulates specific data set themes. The legislation regulating these data sets has been referred to in the INSPIRE priority data set list: a "list of data sets related to environmental reporting, which should be made available by Member States".<sup>148</sup> Key relevant examples of EO and environment related legislation include:

- Directive 86/278/EC on sewage sludge
- Directive 91/271/EC on urban waste-water treatment
- Directive 91/676/EC on nitrates
- Directive 92/43/EC on habitats
- Directive 98/83/EC on drinking water
- Directive 1999/31/EC on landfill of waste
- Directive 2000/60/EC on a water framework
- Directive 2002/49/EC on noise
- Directive 2006/21/EC on extractive waste
- Directive 2006/7/EC on bathing water
- Directive 2007/60/EC on floods
- Directive 2008/50/EC on air quality
- Directive 2008/56/EC on a marine strategy framework
- Directive 2009/147/EC on birds
- Directive 2010/75/EC on industrial emissions
- Directive 2012/18/EC on SEVESO III
- Recommendation 2014/70 on hydraulic fracturing
- Regulation 166/2006/EC on the European pollutant release and transfer register
- Regulation 1143/2014/EC on invasive alien species
- Regulation 2017/852/EC on mercury

The EU level legislation clearly identifies a number of themes containing relevant data sets for EO and environment: water, waste, noise, air, nature and industrial emissions. These themes are also referred to by the Directorate-General for Environment (DG ENV) and the European Environment Agency (EEA) in their overview of INSPIRE priority data set list for e-reporting.<sup>149</sup>

The Directive on access to environmental information also identifies certain themes which contain environmental information (and must therefore be publicly accessible): air and atmosphere; water; soil; land; landscape; natural sites (including wetlands, coastal and marine areas, biological diversity and its components, including genetically modified organisms, and the interaction among these elements); substances; energy; noise; radiation; waste (including radioactive waste, emissions, discharges and other releases into the environment, affecting or likely to affect the elements of the environment); measures (such as policies, legislation, plans, programmes, environmental agreements, and activities affecting or likely to affect the elements and factors of the environment as well as measures or activities designed to protect those elements); reports on the implementation of environmental legislation; cost-benefit and other economic analyses and assumptions used within the framework of the measures and activities; and the state of human health and safety (including the contamination of the food chain, where relevant,

---

<sup>148</sup> <https://inspire.ec.europa.eu/metadata-codelist/PriorityDataset>

<sup>149</sup> <https://inspire.ec.europa.eu/work-programme/priority-list-data-sets-ereporting>

conditions of human life, cultural sites and built structures inasmuch as they are or may be affected by the state of the elements of the environment).<sup>150</sup>

Whereas the INSPIRE Directive and the Directive on access to environmental information set expectations concerning the discoverability and public accessibility of data, the High Value Data list for Article 14 of the PSI Directive will add the requirements for open data (reusable, open license, free of charge, machine readable, and available through an API and/or bulk download). For the purpose of this Impact Assessment it means that the INSPIRE Directive and the Directive on access to environmental information are a useful way to both establish that a data set within their scope is of relevance, as well as that such data is likely to exist within (a majority of) the Member States. (The INSPIRE Geoportal provides a general overview of which MS are providing metadata, viewers or downloads for which datasets, although the actual details are regularly incomplete or outdated.)

Based on the legislated themes, Member State recommendations and stakeholder interviews, a list of themes encompassing data sets within the scope of this Impact Assessment has emerged.

As stated at the beginning of this chapter Earth observation as observing the planet's physical, chemical and biological status over time, can be seen as the combination of satellite remote sensing, earth based remote sensing, and in-situ data collection regarding the environment. Space based remote sensing is mostly done through European and international level efforts, and only a very few MS have their own space based remote sensing capabilities. Additionally where such national capabilities are the result of private participation data may fall outside the scope of the PSI Directive based on third party rights. As the scope of this study concerns data within scope of the PSI Directive which is held at MS level by a majority of MS, the themes that emerged for Earth observation concern in-situ and earth based remote sensing data, concerning the state of the planet in physical, chemical or biological terms, and **space based remote sensing is deemed out of scope.**

These themes are divided into several groups, which map to INSPIRE themes. Environmental data have use value both on their own and as in-situ data for EO services:

Earth Observation:

- Digital elevation models (LIDAR) (INSPIRE Annex II)
- Coastal vulnerability (INSPIRE Annex III)
- Geology / geophysics (e.g. seismic monitoring) (INSPIRE Annex II)
- Hydrography (INSPIRE Annex I)
- Land parcels, use and cover (parcels is INSPIRE Annex I, cover is INSPIRE Annex II , use is INSPIRE Annex III),
- Oceanography (INSPIRE Annex III)
- Ortho-imagery (INSPIRE Annex II)
- It should be noted that other themes like climate data and transport networks are covered in other chapters of this assessment but have relevance here too.

Environment:

- Air quality (e-reporting priority list)
- Biodiversity (INSPIRE Annex III, e-reporting priority list)
- Emissions (industrial) (e-reporting priority list)

---

<sup>150</sup> Article 2(1) of Directive 2003/4/EC of the European Parliament and of the Council of 28 January 2003 on public access to environmental information and repealing Council Directive 90/313/EEC

- Environmental monitoring facilities (INSPIRE Annex III, partly in e-reporting priority list)
- Forestry (INSPIRE Annex II, land use)
- Natural hazards (INSPIRE Annex III)
- Nature preservation/reserves (INSPIRE Annex I, e-reporting priority list)
- Noise (e-reporting priority list)
- Soil (INSPIRE Annex III)
- Waste (e-reporting priority list)
- Water (e-reporting priority list)

This thematic list contains multiple, often many different datasets within each theme. The environmental e-reporting priority list for instance contains over 80 datasets, spread over seven themes. Individual MS suggested up to 300 different datasets to be considered within the thematic area of Earth Observation and Environmental data. What stands out from stakeholder input is that for any given use case it's not a single data set that will be relevant, but rather multiple datasets from across different themes from the list above. Either because these are adjacent to each other and become mostly useful when used together, such as when using air quality measurement data, one also needs the location of monitoring stations involved, as well as models used, areas these models are applied to and air quality management zones involved, to be able to construct a more complete overview. Or because the use case depends on additional data sets from other themes e.g. for context, detection of correlations and cross-reference, for instance combining air quality with land use, and noise in the context of species distribution.

As Earth is a complex system, many different factors mutually influence each other or are mutually dependent. This is reflected in that no single data set or theme within the scope of Earth Observation and Environmental data stands out as most important, or as singularly valuable for re-use on its own. Obviously there are differences in utility of these data sets for specific use cases, but there seems no viable ranking possible. This is also visible from the thematic diversity in the data sets that individual MS suggested to be taken into account as HVD. While there are thematic overlaps, e.g. air quality would be named by most, there are also many differences, so that even with overlap the list of themes and datasets within those themes remains long. Input from re-users reflects this as well, research questions and data usage interests are as varied as the list itself, and provide little contrast through which to shorten the list of data sets to take into consideration.

We therefore see all Earth Observation and Environmental data within scope that is contained in the below listed INSPIRE themes, with data relevant to the Environmental e-reporting priorities forming an additional preference. This places the following themes in scope:

E-reporting priority data list:

These are the datasets used for the preparation of reports by Member States under the environmental acquis as documented in the EEA's reporting obligation database<sup>151</sup> and further detailed in the INSPIRE priority data set list for environmental reporting<sup>152</sup>. These datasets cover topics such as:

- Air Quality,
- Biodiversity,
- Emissions,
- Nature preservation,
- Noise,

<sup>151</sup> <https://rod.eionet.europa.eu/>

<sup>152</sup> <https://inspire.ec.europa.eu/metadata-codelist/PriorityDataset/>

- Waste, and
- Water

INSPIRE Themes:

- Hydrography (Annex I)
- Land parcels (Annex I), limited to geometries, parcel type and parcel code, as per the Geospatial thematic area w.r.t. Cadastral parcels.
- Protected sites (Annex I)
- Elevation (Annex II)
- Geology (Annex II)
- Land cover (Annex II)
- Ortho-imagery (Annex II), excluding satellite sensor derived data
- Bio-geographical regions (Annex III)
- Environmental Monitoring Facilities (Annex III)
- Habitats/Biotopes (Annex III)
- Land use (Annex III),
- Natural Hazards (Annex III),
- Oceanography (Annex III)
- Sea Regions (Annex III),
- Soil (Annex III),
- Species Distribution (Annex III)

Taking this approach builds on a broad existing consensus about what data is relevant, as evidenced by their adoption into INSPIRE, that the data is likely to exist across a majority of MS, and that there will already be commonalities across MS in how this data is being made available today.

For the purpose of analysis we selected three example data sets that represent various relevant aspects of the many data sets involved (and were regularly mentioned by stakeholders as practical examples in discussing various barriers for re-use, and potential) : taken from more Earth Observation oriented and more Environment oriented themes, concerning both measurement and administration, both on and off the e-reporting priority list, and both useful directly and more indirectly within e.g. Copernicus Services. Those three datasets are:

- Air quality measurement data
- Natura 2000 sites
- Land parcels

Their benefits are discussed hereafter.

### **2.5.2 Analysis of the value of these datasets**

Given the extensive scope of this thematic area, use cases are very varied as well. Where traditionally environmental and earth observation data always has been of value to many public sector bodies for e.g. spatial planning and administrative purposes, and for research into environment, biology, geology and urban and land development, the availability of larger volumes of digitised data also allow the use of that same data for more everyday practical purposes of e.g. the agricultural and building sectors. For instance land parcel data and land use data have been useful in the agricultural sector for administrative purposes, such as providing for the documentation needs of EU CAP subsidies. That same data is now also being

used to allow farmers to better select parcels for lease based on previous crops grown there. A Dutch service provider, "Boer en Bunder", provides farmers with a dashboard for the land parcels they work, that brings together detailed crop data, environmental, meteorological and geographic data, and helps them select land parcels to lease for their crops in the coming growing season. Digital elevation data (acquired through LIDAR) in the Netherlands is a key tool for the public sector in water management, but more recently it also has helped detect new archeological sites as it registers small deviations in height and looks below foliage. That same data is used by construction companies to better plan, budget and execute earthworks, as the data allows for precise calculation of earth volumes to be moved, leading to large cost savings in the budgeting phase of projects and faster execution.

Where the above examples show direct use value, there are many ways in which there is indirect use value. That indirect use value comes in three different forms. Generally within one theme, data of a more administrative character, demarcating zones, areas etc., provide crucial context for the use and interpretation of measurement data. Second, use value is derived from combining data across different subjects, providing insight by correlating different data sets within the Earth Observation and Environmental data thematic area, or across to completely different thematic areas, e.g. combining emissions data with health statistics. Thirdly, indirect use value is derived from improving and reinforcing existing valuable services for which data in this thematic area serves as an input. This is true for how in-situ data is used to augment space based remote sensing data, such as in the six Copernicus services (atmosphere, marine, land, climate change, security, emergency). There the broader availability of data that can serve as input, strengthens existing use cases and value creation. For instance the EU wide availability of re-usable land parcel data would immediately improve Copernicus services, as sourcing that data now means confronting different conditions and requirements across MS.

All of the listed categories of datasets are valuable from an environmental and social perspective as the many publications and use cases<sup>153</sup> linked with Copernicus data and services suggest, as does their presence within the scope of the INSPIRE Directive itself. While the benefits for environmental and climate policies are strong<sup>154</sup>, Earth Observation and environmental data also helps in many other policy domains and sectors. For instance, one of the latest studies on the value of EOs Copernicus data explains how these datasets can be used for conservation and valorization of cultural heritage at the Member States level<sup>155</sup>. Furthermore, recently their economic value was also highlighted<sup>156</sup> and especially within the agriculture sector<sup>157</sup>. The European association of earth observation companies (EARSC<sup>158</sup>), representing some 80 members, have identified 22 active market sectors and list over 100 example products and services in which EO and environmental data is used for those sectors. These sectors range from agriculture, forestry and fisheries to energy transition, mining, construction, utilities, insurance, tourism, and emergency services, research, climate adaptation and more.

Researchers, citizens, and both commercial and non-commercial organisations have an existing strong interest in re-usable data for this thematic area, as indicated both by data holders and re-users.

---

<sup>153</sup> <https://www.copernicus.eu/en/use-cases>

<sup>154</sup> There are many reports highlighting the importance of earth observation data for environmental policy making: see for instance: A walk to the park? assessing access to green areas in Europe's cities update using completed Copernicus urban atlas data, 2018 [https://www.copernicus.eu/sites/default/files/2018-10/2018\\_01\\_green\\_urban\\_area\\_0.pdf](https://www.copernicus.eu/sites/default/files/2018-10/2018_01_green_urban_area_0.pdf) or Copernicus ex-ante benefits assessment, 2017, [https://www.copernicus.eu/sites/default/files/2018-10/Copernicus-Ex-Ante-Executive-Summary\\_0\\_6.pdf](https://www.copernicus.eu/sites/default/files/2018-10/Copernicus-Ex-Ante-Executive-Summary_0_6.pdf)

<sup>155</sup> Copernicus services in support to Cultural Heritage, October 2018, [https://www.copernicus.eu/sites/default/files/2019-06/Copernicus\\_services\\_in\\_support\\_to\\_Cultural\\_heritage.pdf](https://www.copernicus.eu/sites/default/files/2019-06/Copernicus_services_in_support_to_Cultural_heritage.pdf)

<sup>156</sup> [https://www.esa.int/About\\_Us/Business\\_with\\_ESA/Global\\_Space\\_Economic\\_Forum/Case\\_study\\_shows\\_economic\\_value\\_of\\_Copernicus\\_Earth\\_observation\\_data](https://www.esa.int/About_Us/Business_with_ESA/Global_Space_Economic_Forum/Case_study_shows_economic_value_of_Copernicus_Earth_observation_data)

<sup>157</sup> <http://earsc.org/Sebs/>

<sup>158</sup> EARSC EO Wiki Space <https://earsc-portal.eu/>

Stakeholders also frequently point out that public sector institutions comprise a large part of likely and existing re-users. Not just within MS where public sector bodies encounter difficulty or fees to source data from other government institutions, but also across MS, where such data may feed into Copernicus Services, or for direct re-use. For instance an environmental agency from a MS indicated to be very interested in API access to seismological data from other MS.

While the scope of data in this thematic area, and the number of use cases and re-use examples all point to re-use value certainly being created, we haven't found much quantitative research to underpin that. Sweden sees a high demand for EO data like land parcel and land use data, and ortho-imagery with a potential of 518 million Euro in value for all geospatial data which includes EO, and sees sizable contributions of that data to value creation in the agricultural sector (in total between 105 million Euro in direct, up to 278 million including indirect impact for all Swedish proposed HVD's), spatial planning (288 to 595 million Euro for all Swedish proposed HVD's), and the public sector itself (105 to 163 million Euro for all Swedish proposed HVD's)<sup>159</sup>. The latter is another sign that open data from this thematic area significantly reduces friction towards use within the public sector itself. A Finnish study<sup>160</sup> found there's an annual economic benefit potential for spatial data services, of which the data in this thematic area were taken as a part, of around 13 billion Euro, of which some 3 billion Euro (22%) has been realised. The estimated value contribution of the Geospatial Platform (fully functional by 2025), which will incorporate data from the Earth Observation and environmental thematic area, will be 150 million Euro annually in direct plus 400 million annually in indirect benefits.

With the insights above and other input from both stakeholder feedback and desk research taken into account the following table lists the areas of benefits for the various identified macro categories of value creation.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
Economic	Competition		SME's are the largest user group of open environmental data in Finland, according to a 2019 user survey <sup>161</sup>
	Consumer benefits		A variety of services for citizens to choose from that inform them about e.g. air quality is visible across Europe. An application e.g. like Airly.eu combines EEA and MS data with data from their own members' sensor network.
	Economic output, employment, market dynamism,		EARSC, the European association of earth observation companies with some 80 members, have identified 22 active market sectors and listing over 100 example

<sup>159</sup> Damvad Analytics, for the Swedish Land Bureau, 2020

<sup>160</sup> Spatineo, 2019, for the Finnish government

<sup>161</sup> SYKE user survey 2017 & 2019

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
	productivity		<p>products and services in which EO and environmental data is used<sup>162</sup>.</p> <p>The 2019 Finnish Spatineo study<sup>163</sup> estimates that over 75% of potential benefits (totalling 13 billion Euro) of spatial data services, which includes the data for this thematic area, is left untapped. The estimated value contribution of the Finnish geospatial platform under development will be 150 million Euro annually in direct benefits, and 400 million indirectly.</p> <p>Sweden sees high demand for EO data, which contributes to a 518 million value potential for all spatial data, with EO data particularly contributing to the agricultural sector, spatial planning and to the public sector itself.<sup>164</sup></p>
Climate change	Citizen engagement in addressing climate change		Environmental monitoring data is used in citizen science efforts in various MS (e.g. air quality, emissions in Netherlands, Germany), more generally environmental monitoring data is used for 'around my home' type of apps.
	Energy management and efficiency		Emissions data plays a role in shaping climate policy w.r.t. energy transition in general, and in decisions w.r.t. powerplants in particular
	Environment management		The agricultural and spatial planning applications of earth observation and environmental data are widely mentioned by MS and re-users w.r.t environment management. Reporting obligations also create their own demand for this data from NGO's and the general public, to make their own status assessments.
Innovation & AI	Citizen innovation		The Swedish environmental agency sees environmental

<sup>162</sup> EARSC EO Wiki Space <https://earsc-portal.eu/>

<sup>163</sup> Spatineo, 2019, for the Finnish government

<sup>164</sup> Damvad Analytics, for the Swedish Land Bureau, 2020



<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			data as key community resource.
	Entrepreneurialism and private sector innovation		New services using a mix of EO/Environmental data emerge, e.g. in the agricultural sector (e.g. 'boerenbunder.nl'), the European Data Portal lists over 40 such new services.
	Public sector innovation		Novel uses w.r.t the execution of public tasks are made possible with EO/environmental data. (E.g. the use of LIDAR for rain water flood management at local level, and detection and protection of archeological sites with the same data, both in the Netherlands)
	Public private coordination		Environmental, climate change and biodiversity related research benefits strongly from public data availability.
Public services and administration	Public sector revenue		The existing economic activity and potential for the data in scope should translate into additional tax revenue, as we've previously seen for other thematic areas and datasets <sup>165</sup> . In parallel, several MS indicated the reduction of charges for data was an efficiency gain for them.
	Public service performance		Open data allows psb's other than the dataholder to better perform their tasks. E.g. Dutch municipalities were required to 'stress test' their rain water flood management, and the only way to do so was to use the open digital elevation model data (lidar). Swedish natural hazard data will similarly be of high use value for spatial planning by municipalities, and the dataholder sees a high demand.
	Public services management		A Polish government feasibility study into open data benefits highlighted the time savings in internal procedures, both for the dataholder and for re-using public entities.
Re-use	Demand for information		All MS indicate demand for research purposes, across all datasets within scope.

<sup>165</sup> De Vries et al, 2011

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			<p>Slovenia sees high demand for digital elevation models, land parcel and ortho-images, from private companies like geodetic companies and landscape designers, as well as from local government. Other MS see similar high interest in these datasets.</p> <p>Reporting obligations also create their own demand for this data from NGO's and the general public, to make their own status assessments.</p> <p>Air quality has a high public interest. Over 90% of the requests the Polish environmental monitoring agency received in 2019 concerned air quality monitoring.</p> <p>Finland has seen a non-linear growth of demand for spatial data, that includes data within this thematic area, since the adoption of an open data policy in 2008. The Finnish Environment Institute SYKE carried out user surveys in 2017 and 2019 amongst 200 re-users, showing high interest in environmental data for decision making (29%), research (17%), commercial use (28%).</p>
	Trust and confidence in information		Ground based in-situ measurements (e.g. of air quality) play an important role in validating and calibrating space based remote sensing data, and as such are crucial in realising the use and value of e.g. Copernicus services.
	Volume and range of information		Many use cases depend on the availability of data from different domains (e.g. combinations of traffic, air quality, species distribution and climate data, water and agriculture, land use and natural hazards etc.)
Social	Public engagement and government transparency understanding		Reporting obligations also create their own demand for this data from NGO's and the general public, to make their own status assessments, in holding public service to account.

Re-use value is deemed to be higher by both data holders and users when data can provide wider coverage (e.g. data is available for most MS), and provided in a timely manner. This is particularly true for

uses other than in policy planning or reporting cycles. In the latter the availability is sufficient if availability keeps the same rhythm as those, but for most other uses being able to access updated data as soon as a change has occurred is much more important.

## 2.6 Statistics

In this section, the report provides a detailed overview of the datasets in scope of the analysis for this thematic area, and maps the main benefits and use cases of these data fields.

### 2.6.1 Identification of the datasets in scope

The statistical production is often governed by the first principle of the Fundamental Principles of Official Statistics (adopted by United Nations Statistics Division) that states “Official statistics provide an indispensable element in the information system of a democratic society, serving the Government, the economy and the public with data about the economic, demographic, social and environmental situation.”<sup>166</sup>

The availability and access to reliable and high-quality data and information about the social, economic and environmental condition of a country are important elements for policy and decision makers, business community and citizens. In general, this data is produced by public bodies as part of their official function and disseminated are by the institutions themselves and/or by a central institution set up at national level - National Statistical Institute (NSI) as the official statistics of the country. When it comes to the governance of the national statistical system, this is often determined by law and is regulated by adherence to professional standards. In general, the system is led or co-ordinated by the NSIs, which become the main gateway for national data to the outside world. The NSIs also ensure that all data provided have the quality required by the statistical standards. As to retain trust in official statistics, the data produced by NSIs are produced in a professionally independent way based on scientific methods, rigorous quality criteria, including relevance. Therefore, the data in scope of the current research is the data produced as official statistics and held by the NSI, including the ones underlying the European statistics.

In addition, at European level the statistical coherence and quality is ensure through the European Statistics Code of Practice. It is a self-regulatory instrument and is based on 16 Principles covering the institutional environment, statistical processes and statistical outputs. Other important documents that oversee the production of official statistics and putting in place the general development guidelines are:

- the *Regulation (EC) No 223/2009 on European statistics*<sup>167</sup>
- the *Quality Assurance Framework of the European Statistical System*<sup>168</sup>,
- the *Commission Recommendation of 23 June 2009 on reference metadata for the European Statistical System*<sup>169</sup>,
- the *General Data Protection Regulation (GDPR)*<sup>170</sup>.

---

<sup>166</sup> <https://unstats.un.org/fpos/>

<sup>167</sup> Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics and repealing Regulation (EC, Euratom) No 1101/2008 of the European Parliament and of the Council on the transmission of data subject to statistical confidentiality to the Statistical Office of the European Communities, Council Regulation (EC) No 322/97 on Community Statistics, and Council Decision 89/382/EEC, Euratom establishing a Committee on the Statistical Programmes of the European Communities, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009R0223>

<sup>168</sup> <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>

<sup>169</sup> <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:168:0050:0055:EN:PDF>

Therefore, the official statistics have high quality standards, full transparency of methods and assumptions, professional independence or submission to independent professional judgement (neutrality), publicly available to everyone at the same moment, coherence and comparability, etc. While the official statistics are not the sole supplier of information, they are seen as a trustworthy source of information based on the way they are produced - professionally independent, based on scientific methods and rigorous quality criteria. Moreover, they are publicly available to anyone at the same time, and offer high quality information. Official statistics are produced to be used and to make an impact on society through a higher degree of openness and transparency, ensuring confidentiality and, at the same time, providing equal access to information to the citizens. Their use contributes to building a society with more empowered people, better policies, more effective and accountable decision-making, greater participation and stronger democratic mechanisms. In the same time, official statistics should reflect users and re-users' needs and be developed accordingly.

Official statistics become high-value datasets based on their usage: who needs them and why, how and where are they used. Several stakeholders, including re-users, agree that all official statistics should be considered high-value datasets.<sup>171</sup> In general, datasets will be of different values for different stakeholders, based on their respective needs. The European Commission guidelines on licences, datasets and charging for re-use of documents<sup>172</sup> propose some of the datasets to be release with higher priority. However, particular situations will bring forward different types of datasets as being of outmost importance. If we take as an example the current COVID-19 pandemic, we identify health data, together with data related to death rates, as most important datasets in understanding the evolution of phenomenon. The above-mentioned guidelines also consider that depending on the circumstances (relevance to strategic goals, market developments, social tendencies, etc.), other categories could be considered "core" or "high-value" data and the responsible public authorities in consultation with the relevant stakeholders should assess in advance which datasets should be released as a priority. Therefore, a selection of few datasets to be included on the high-value list does not imply that the rest of the official statistics are considered "less relevant" but it should be considered to merely be a starting point of the whole process.

A monitoring report of Eurostat digital dissemination (November 2019) showed that within the top 10 European statistics the main table consulted were *population on 1 January, gross domestic product (GDP) per capita in PPS, real GDP growth rate – volume, harmonised unemployment rate by sex, GDP by volumes and at market prices*. In addition, a recent exercise conducted by Eurostat together with the National Statistical Institutes from the Member States revealed that demography, poverty and inequality, national accounts (GDP), labour market, prices, regional statistics, government finances, business and health statistics are considered to be of high-value at both European and Member States levels.<sup>173</sup> The desk research results and stakeholder interviews also showed that from the broad range of official statistics datasets, the demographic indicators, the GDP and labour market datasets are often found within the top 10 most searched for topics across Member States.

The Global Open Data Index (GODI) measures the openness of clearly defined data categories relevant for civil society at large. In its latest open data benchmark, published in 2017, the GODI shows that *national*

---

<sup>170</sup> Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679>

<sup>171</sup> Stakeholder interviews and focus group discussions.

<sup>172</sup> Commission notice — Guidelines on recommended standard licences, datasets and charging for the reuse of documents OJ C 240, 24.7.2014, p. 1–10

<sup>173</sup> The detailed results of the Eurostat proposal are presented in the European Commission document "high-value datasets in the statistics category", Ares(2020)3505834.

*statistics*, next to *government budgets*, is one of the most open dataset types. In terms of datasets, there is no particular preference and the datasets selection varies across Member States. For example, in Netherlands the frequent use of datasets by citizens and other re-users make them high-value datasets. The intense usage confirms the users' interest in the datasets and it also validates the adequacy of the data to their needs. The top five datasets in the case of The Netherlands are Netherlands regionally, population, price indexes, labour and social security, and income and spending, but this can change if the users' needs change.

Due to the broad selection of datasets available in the context of the official statistics, and based different inputs received from stakeholders on use of datasets, we narrowed down the list of option covering social, macroeconomic and business datasets as potential high-value datasets and they are generally described as follows. This list of datasets also considers the proposal of datasets considered as HVDs by the members of the European Statistical System Committee (ESSC), after the consultations and discussion with Eurostat.<sup>174</sup>

**Macroeconomic datasets** refer to a wider range of key statistics for the whole economy, individual sectors and the relationships between them. Consequently, the data will provide an extended picture of the economic environment. In general, the data refers to categories such as gross domestic product (GDP), government finance, sector accounts (financial and non-financial) and labour market. The **national accounts** datasets bring forward information on the health of a country's economy as well as of its specific economic sectors. Within this category, the **gross domestic product** (GDP) data is often used to measure of countries' economic performance. Its components provide insight on the financial performance and on the value added of the economy as a whole, but also at the level of specific activities, components and/or economic sectors. The three methods in calculating the GDP provide information on different economic aspects - gross value added, taxes and subsidies (through production approach), final consumption, breakdown by households and government categories, fixed capital formation, exports and imports (in the expenditure approach) and compensation of employees, mixed income, taxes on production and imports, subsidies (when using the income approach). The gross value added often includes a breakdown by economic sectors. The changes over time in the prices of consumer goods and services acquired by households (price inflation) is an important phenomenon of economic stability. To measure the price stability and convergence across Member States, the Harmonised Index of Consumer Prices (HICP) is used.<sup>175</sup> Thus, HICP is another key aspect of the economic performance of a country. The **government finances** data is an additional component of the macro-economic environment. The information provided via government finances datasets help increase the transparency level of the public revenues and spending, showing where the money went, how have the public funds developed over time, and what type of activities have been funded. The main key attributes refer the expenses and revenues that includes different levels of disaggregation - by general, central and local governments, by social security funds, by government functions (COFOG).

**Social statistics** reflects the countries' population social conditions through the perspective of demography, labour market, poverty and inequality and health aspects. **Demography** is an important key statistic in understanding socio-economic development of countries. Often, the data concentrates on the overall structure and trend of population development, with details on **population**, and *vital statistics* - **births** (including **fertility** rates) and **deaths** (including **mortality** rates) statistics. The information is often used to provide background information for different macro characteristics categories. The datasets

---

<sup>174</sup> European Commission document "high-value datasets in the statistics category", Ares(2020)3505834.

<sup>175</sup> HICP is used as an indicator for assessing price convergence with regard to a possible accession of a country to the monetary union and it is a measure of price stability for the European Central Bank's (ECB) monetary policy strategy. The Governing Council of the ECB has defined price stability in terms of the Harmonised Index of Consumer Prices (HICP) for the euro area.

have a good level of disaggregation - by sex, age and age groups (often five years age group) and place of residence (up to NUTS 2 level information) in most of the European Union Member States. Some countries offer additional breakdown information such as level of education, citizenship, marital status. **Labour market data** provide another set of information about of the socio-economic environment within a country. It helps identifying economic needs and niches of opportunities, and reflects the labour market dynamism via employment and unemployment rates, job vacancies and labour force potential, pointing out to both hotspots for development and structural issues within economies. Most of the datasets provide wide range of breakdowns, from regional level (up to NUTS 3 level in some countries) and economic sector (NACE rev. 2), to sex, age groups, occupations, education level, citizenship, place of residence, duration of unemployment, etc. **Poverty and inequality data** reflect the population social struggle and overall well-being. Material and social deprivation, at-risk-of-poverty and low work intensity households are important aspects of societal development of countries. An aging population needs better healthcare services. Thus, health data and, in particular, **current healthcare expenditure** provide relevant information on costs, general services and providers at countries' level.

**Business statistics** are another important segment of the macro-economic environment. The datasets provide information about structure, dynamic and performance of the economic environment. The information covers the economic sectors of national economies, from the industries sectors to services and trade and to tourism flows across the Member States. It often reflects the business sector's economic performance, through the perspective short-term business statistics, trade and tourism. Business communities are an essential components of the countries' economic development, and short-term business data and trade data provide the relevant information to monitor their competitiveness and performance. In some of the Member States, the tourism flows constitute an important share of economy. The breakdowns include the type of activities (NACE rev. 2) and products, trade partners and regional level (NUTS 2 for specific datasets), and specific breakdowns for the tourism sector, such as duration and destination of trips, means of accommodation and transport, booking modalities.

### 2.6.2 Analysis of the value of these datasets

Providers and re-users of national statistics have diverse perspectives in defining high-value datasets within national official statistics. However, a general consensus shows that both demographic and macro-economic statistics are considered very important. In an information age, the provision of reliable and high-quality data and information by national statistical institutes (NSIs) and other producers of official statistics around the world is increasingly important to our economies and societies.<sup>176</sup> One of the key features of official statistical data is their quality, influencing the value that users experience. The demand for statistics is rapidly growing and the world's realities creates new needs for accurate information about economies and societies. In addition, users' needs are becoming more complex and individualized, needing more detailed information on small population groups and geographic localities, for instance.

The information availability and data interoperability (harmonisation) are factors that consolidate the value brought forward by the official statistics. Official statistics, in general, allow sound international cooperation between countries and supports the design of evidence-based policy at both national and international level. For example, monitoring the progress of the 2030 Agenda on Sustainable Development Goals relies also on socio-economic indicators produced by NSI.<sup>177</sup> When it comes to beneficiaries, policy makers and decision-makers are an important part of them. Data are the lifeblood of decision-making and the raw material for accountability. The GDP, government finances, labour market and demography are

---

<sup>176</sup> United Nations Economic Commission for Europe (UNECE), Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics (New York and Geneva: United Nations, 2018).

<sup>177</sup> <https://unstats.un.org/sdqs/report/2019/The-Sustainable-Development-Goals-Report-2019.pdf>, <https://ec.europa.eu/eurostat/cache/diqpub/sdqs/index.html>

some of the key indicators used to assess the European Semester<sup>178</sup> progress of Member States and to monitor the implementation status of the country-specific recommendation. Without high-quality data providing the right information on the right things at the right time; designing, monitoring and evaluating effective policies becomes almost impossible.<sup>179</sup> "The Oak of Truth"<sup>180</sup> visualises Estonia's progress in achieving the key indicators set out in three strategies - the national sustainable development strategy "Sustainable Estonia 21", the national reform programme "Estonia 2020" and the action plan of the Government of Estonia. It has been created by the Government Office and Statistics Estonia to make the Government's goals easier to understand for the general public, more transparent and goal-oriented.

But they are not the only ones. The scientific community is one of the biggest official statistics consumers. Socio-economic studies rely on qualitative and trustful data and researchers use official statistics together with their own research data. The ability of businesses and people to make well-based decisions relies also on data and their implied quality. In addition, data brokers such as GlobalDatabase.com<sup>181</sup> and Statista.com<sup>182</sup> provide and develop new services by re-using official statistics. And, also the journalists are avid users and re-users of official statistics, when informing readers about different topics relevant to society or just explaining economic terms on simpler terms.

The GDP data is used as an indicator of economic recovery after a crisis, but also as a measure of economic development in general. All the information is available at national level, but recent development showed that countries have also started compiling GDP data at regional level (based on the income approach). Due to its particular structure and coverage, the GDP datasets provide important information on different patterns existing within economies. It is often combined with demographic indicators, business and employment statistics to provide additional insight on the socio-economic context of the countries. The economists and analysts often used the GDP and price stability (HICP) as measures of comparison for economic performance and stability of a country or between countries.<sup>183</sup> Both annual and quarterly data are used in various economic forecasts and data modelling. Often the government finances are associated with Government role with economies. Opening up public finances data is seen as a measure to increase Government transparency. For example, the *OpenCoesione* in Italy focuses on projects that are financed by Italian the department of Cohesion Policies.<sup>184</sup> It shows how (much) money is spent on different subjects in different regions, with visualizations. The data are published for the public to assess whether the projects that are financed meet their needs and if resources are used effectively. By doing so, citizens are able to understand how this specific department within the government is spending their money and thereby transparency is increased. The Accountability Hack is a recurrent semi-annual event in the Netherlands where programmers, policy officers, data scientists and journalists search for solutions to policy issues together using open data.<sup>185</sup> Estonia publishes periodically financial data about the government sector's activities as open data to make government sector's financial activities more transparent and understandable.<sup>186</sup> Easy-to-understand versions of the public finances were also developed by re-users in Finland<sup>187</sup> and France<sup>188</sup> and by the data holder organisation in Ireland<sup>189</sup>.

---

<sup>178</sup> [https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/european-semester\\_en](https://ec.europa.eu/info/business-economy-euro/economic-and-fiscal-policy-coordination/eu-economic-governance-monitoring-prevention-correction/european-semester_en)

<sup>179</sup> A World that Counts: [www.undatarevolution.org/report/](http://www.undatarevolution.org/report/)

<sup>180</sup> <https://tamm.stat.ee/kategooriad/uro-kestva-arengu-kava> (in Estonian)

<sup>181</sup> <https://www.globaldatabase.com/our-data>

<sup>182</sup> <https://www.statista.com/>

<sup>183</sup> <https://www.investopedia.com/ask/answers/what-is-gdp-why-its-important-to-economists-investors/>

<sup>184</sup> <https://opencoesione.gov.it/en/>

<sup>185</sup> <https://openstate.eu/nl/projecten-tools-data/evenementen/accountability-hack-2/> (in Dutch)

<sup>186</sup> <https://riigiraha.fin.ee/QvAJAXZfc/opendoc.htm?document=riigiraha.qvw&lang=en-US&host=local&anonymous=true/>

<sup>187</sup> <https://www.tutkihankintoja.fi/?lang=en>

<sup>188</sup> <https://www.fiscalite-locale.com/plugin.php?e=observatoire>



Demographic statistics are other important key statistics in understanding socio-economic development of countries. Within the demographic dynamic both the population and vital statistics (births and deaths) datasets are used. These datasets offer relevant information on the overall structure and trend of population development and are available for longer periods of time (long time series). In general, the demographic datasets are used to provide background and contextual information for different macro characteristics categories. Often, the economic development is measured also from individual (inhabitant) perspective not only from global (international, national and regional) perspectives. Poverty level of the population is an important indicator in socio-economic monitoring process. Poverty and inequality statistics come complementing the demographic perspective of the population. The income dataset aims to identify the segments of population most affected by poverty and social inequality for better approaches to tackle this issue. Material and social deprivation, at-risk-of-poverty and low work intensity households are important parameters reflecting essential aspects of the well-being of population within countries. Demographic change is also used in various socio-economic forecasts on short-, medium- and long-term. Both population and vital statistics contain information useful in estimating replacement rate of the workforce and age-dependency ratio trends, two important aspects in socio-economic development of countries.

The labour market dataset is the other side of the socio-economic development coin. Employment and unemployment are often used as proxies for identifying economic needs and niches of opportunities. Job vacancies offer an additional layer of information on the countries' labour market dynamism. Unemployment rates might be able to underline development's hotspots, but also points out to the structural issues within economies. With breakdowns by regional level (NUTS 3 and municipalities levels in some countries) and by economic sector (NACE rev. 2), the labour market dataset is a relevant tool for both entrepreneurs and researchers. In the economic analysis, the employment rate is often complemented by the unemployment rate to better explain the economic or crises effects on society. The long-term unemployment is an indicator used to identify possible structural issue of the labour market within countries and economies. The education levels of the workforce and occupations as well as the labour costs offer insights on business developments and possible expansion opportunities.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Examples/references</b>
Economic	Employment	The labour market and demography offer important insight of economic development opportunities through various indicators, such as <i>employment and unemployment rates, job vacancies and training participations, with disaggregation by age, sex, economic activities, occupations, regions, etc.</i> Analysts, economists and journalists often use these datasets when talking about effects of major crisis on society. The International Labour Organisation (ILO) chief mentions 195 million job losses around the world due to COVID-19. <sup>190</sup> An analysis of occupation-level data estimates that the COVID-19 crisis could leave up to 59 million jobs at risk in Europe, a staggering 26 per cent of total employment in the 27 Member States of the European Union, plus the United Kingdom. <sup>191</sup>
	Economic output	Gross domestic product and price statistics are often the measure used by journalists and economists when analysing the countries' economic development, stability and performances. Combined with demographic

<sup>189</sup> <https://whereyourmoneygoes.gov.ie/en/about/>

<sup>190</sup> "COVID-19: impact could cause equivalent of 195 million job losses, says ILO chief," in *Economic Developments, UN News*, 08 April 2020; <https://news.un.org/en/story/2020/04/1061322>

<sup>191</sup> David Chinn, Julia Klier, Sebastian Stern, and Sahil Tesfu, "Safeguarding Europe's Livelihoods: Mitigating the Employment Impact of COVID-19," in McKinsey & Co., 19 April 2020; <https://www.mckinsey.com/industries/public-sector/our-insights/safeguarding-europes-livelihoods-mitigating-the-employment-impact-of-covid-19>



		<p>characteristics and labour market data, it provides important contextual information about the economies and societies development. Trading Economics<sup>192</sup> is a platform providing economic profiles, using data from official sources, for more than 200 countries, including the EU Member States. The indicators used include data on GDP, government finances, labour force and wages, and other factors relevant to the economies (trade, market prices, health, consumers, businesses).</p> <p>Also, the GDP per inhabitant is often used to measure the poverty level of the population. The GDP, labour market and demographic indicators are part of the set of indicators used to monitor the progress of the 2030 Sustainable Development Goals.<sup>193</sup> The gross value added by economic sectors brings forward information on the economic structure and of its key sectors.</p>
	Economy monitoring	<p>Policy makers use GDP, prices and labour market data to monitor economic performance, stability and trends, but also for policy developments.<sup>194</sup> Combining and correlating this information with data on trade provides further insights on the economic development of the countries (through indicators such as <i>real growth rate of regional gross value-added (GVA) at basic prices by NUTS 2 regions - percentage change on previous year, shares of exports/imports in total GDP, GVA by economic sector</i>). Labour market and demographic data bring forward additional information for policy makers and decision-makers. <i>Employment and unemployment rates, labour costs per capita</i> are valuable information for both policy makers and entrepreneurs. Often labour costs are an important factor for business development across the world.</p>
	Product market dynamism	
	International competitiveness	
Innovation & AI	Entrepreneurialism and private sector innovation	<p>The population, macro-economic and labour market datasets provide contextual information of the economic and industrial structure of a country through various indicators available, such as <i>population by age groups, regions (NUTS 2 or NUTS 3 level), education attainment, GDP aggregates by industry, GVA by industry sectors, employment and unemployment rates by education attainment and regions (with various level of disaggregation), job vacancies by economic sectors</i>.</p> <p>Often this complementary information helps entrepreneurs to better understand the socio-economic context of a region or a country and offer insights on business development opportunities. The public government datasets, housing datasets, geographic datasets, finance and economics datasets and employment and education data are also data that can be used by AI and machine-learning algorithms.<sup>195</sup></p>

<sup>192</sup> <https://tradingeconomics.com/>

<sup>193</sup> For example, social and macroeconomic statistics are used in the monitoring Goal 1: *End poverty in all its forms everywhere* (People at risk of poverty or social exclusion, People at risk of income poverty after social transfers, Severely materially deprived people), Goal 3: *Ensure healthy lives and promote well-being for all at all ages* (Standardised preventable and treatable mortality, Mortality rate due to tuberculosis, HIV and hepatitis, Mortality rates for new-borns and children under 5 years of age), Goal 8: *Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all* (Employment and unemployment rates, Long-term unemployment rate, Young people neither in employment nor in education and training, Real GDP per capita, Share of government spending in social protection and employment programmes from total government budget, Investment share of GDP by institutional sectors), Goal 10: *Reduce inequality within and among countries* (Income distribution (quintile share ratio), Income share of the bottom 40% of the population), Goal 17: *Strengthen the means of implementation and revitalize the global partnership for sustainable development* (Total government revenues, including the share finance by taxes, Government debt); see also Eurostat, Sustainable development in the European Union: Monitoring report on progress towards the SDGs in an EU context: 2020 edition (Luxembourg: European Commission, 2020) and <https://sustainabledevelopment.un.org/sdgs>

<sup>194</sup> Stakeholder interviews.

<sup>195</sup> Michael Chui, Martin Harryson, James Manyika, Roger Roberts, Rita Chung, Ashley van Heteren and Pieter Nel, *Notes from the AI Frontier: Applying AI for Social Good* (New York: McKinsey Global Institute, 2018).

Public services and public administration	Public sector revenue	Government finances datasets provide the information related to transparency of government expenditures and revenues. The datasets include detailed information that allows the identification of general spending patterns and of sources of revenues for the public administration. The data are often used to increase government transparency in public spending and collection of revenues. “Where your money goes” is an Irish platform and easy-to-use tool for examining gross government expenditure over a period of ten years. <sup>196</sup> The Finish “Explore public budgets” platform provides citizens with information on the purchases made by the state and Finnish municipalities as well as how public funds are being spent. For companies, the service provides information on current market conditions. <sup>197</sup>
	Public services management	
Environmental	Citizen engagement in addressing climate change	The population is often used to provide contextual information of countries and regions, using different indicators such as: <i>median age of the population by NUTS regions, percentage of young/old age population by NUTS regions, population density</i> . This complementary information about the demographic structure and characteristics helps to better understand the consumption patterns of energy and the level of engagement of the population in addressing issues related to the climate change.
	Energy management and efficiency	
Social	Disease prediction and prevention	Providing general or contextual information of the socio-economic environment by combining population dataset with other indicators from GDP and health datasets. Population density and age demographic distribution are useful information in developing plans for prevention and containment of epidemics. The COVID-19 pandemic effect on economy and society holds a lot of media attention. Statistics such as demography and labour market are datasets often used by journalists to illustrate the virus’ impact on society. The population density and age distribution (especially in small communities) are mentioned as part of the key factors that determine how vulnerable places are to the virus by Mr Florida, in its article “The Geography of the Corona Virus.” <sup>198</sup>
	Crime and justice	The labour market and unemployment datasets provide additional contextual information of the socio-economic environment: <i>employment and unemployment by education attainment, age and regions, job opportunities, employment and unemployment rates at regional level</i> . In addition, the GDP adds another level contextual information about the poverty level (e.g. <i>GDP per capita by regions</i> ) at national or at regional levels. Poverty and inequality, and unemployment are often considered as part of the factors influencing criminality level within communities. The information is useful for policy makers in designing strategies to reduce the poverty and, indirectly, to prevent rising levels of illegal activities. Population density also help planning transport networks based on needs and particular characteristics of regions. Population structure (e.g. age distribution, education level, activity status) help better design mobility services and make them more efficient.
	Mobility access	
	Mobility efficiency	
	Mobility planning	

<sup>196</sup> <https://whereyourmoneygoes.gov.ie/en/about/>

<sup>197</sup> <https://www.tutkihankintoja.fi/?lang=en>

<sup>198</sup> Richard Florida, “The Geography of Corona Virus,” published in *Bloomberg’s CityLab*, 03 April 2020; <https://www.citylab.com/equity/2020/04/coronavirus-spread-map-city-urban-density-suburbs-rural-data/609394/>

## 2.7 Mobility

In this section, the report provides a detailed overview of the datasets in scope of the analysis for this thematic area, and maps the main benefits and use cases of these data fields.

### 2.7.1 Identification of the datasets in scope

As anticipated in the premises of the study, the scoping of the “mobility” thematic area was a challenging exercise in terms of its semantic definition and in terms of the identification of the datasets to be considered.

First, the PSI Directive does not provide for a definition of “mobility” and more generally speaking, there appears to be no commonly agreed definition of the term. While “transport” can be pinned down to the movement of people and/or goods by means of networks leveraging the four modes of transport – road, rail, water and air<sup>199</sup>, “mobility” is considered as wider/looser in scope. Therefore, depending on one’s understanding of the concept, the study team was faced with various interpretations of the thematic area, its overall purpose, and consequently the preliminary ideas or ‘wish lists’ of possible datasets in scope, both from the data holder and the data reuser perspectives. In order to cater to the second and most challenging aspect of the thematic area i.e. identifying a “long list” of possible datasets in scope of the study – without creating overlaps with other existing policy initiatives, the development of a *formal* definition of “mobility” was kept on hold, as explained hereafter.

Similarly to other thematic areas, a review of relevant (and mostly transport mode-specific) EU legislation related to “mobility” made it possible to derive an initial pool of data and/or information that are to be collected and/or made available by Member States. In particular, the following pieces of EU law should be highlighted:

- The **ITS Directive** (2010/40/EU) which aims to create a favourable environment for the deployment of Intelligent Transport Systems, through the adoption of technical, functional and organisational specifications in relation to (the procedures for provision of) certain **road transport and other multimodal data**.
- The **Rail Interoperability Directive** (EU/2016/797) which aims to ensure the interoperability of the rail system within the EU, through the development of technical specifications for interoperability (TSIs), including for **data exchange** among competent authorities and parties, notably on rail transport **operations and traffic management**.
- The **VTMIS Directive** (2002/59/EC) which establishes a European **vessel traffic monitoring** and information system (VTMIS) for the tracking of movements of **ships** transiting along the **coast lines** of Member States by competent authorities.
- The **RIS Directive** (2005/44/EC) and implementing regulations which establish harmonised rules for the **provision of river information services (RIS) and data** on all inland waterways of the Member States of class IV and above which are linked by a waterway of class IV or above to a waterway of class IV or above of another Member State, including the ports on such waterways, in order to ensure the safety and sustainability of **inland waterways** in the EU.
- The **ANS Regulation** (EC/550/2004) and **EATMN Regulation** (EC/552/2004) which respectively ensure the interoperability of **air navigation services** and of the European **air traffic management (ATM)** network.

---

<sup>199</sup> See: [https://eur-lex.europa.eu/summary/chapter/transport.html?root\\_default=SUM\\_1\\_CODED=32](https://eur-lex.europa.eu/summary/chapter/transport.html?root_default=SUM_1_CODED=32)

- The **INSPIRE Directive** (2007/2/EC) which defines **transport networks** as one of the 34 spatial data themes needed for environmental applications and to be made available by public sector organisations as well as private organisations acting on behalf of public services.

Following a first assessment and an alignment with the Commission’s Directorate-General for Mobility and Transport (DG MOVE), it was concluded that information falling under the scope of the ITS Directive, with the exception of the datasets also covered by the INSPIRE Directive to which the ITS Directive refers to, would not be considered as potential high-value datasets for the purpose of this study, in order not to interfere with ongoing initiatives related to the implementation of the ITS Directive. The impact of this exclusion on the current study is two-fold.

First, it drastically reduces the length of the initial list of possible high-value datasets to be considered in the field of “mobility”:

- The ITS datasets are recognised by all stakeholders as having a high-value added for re-use – which is of course the reason why they have been retained under the ITS Directive in the first place.
- The ITS Directive covers an extensive range of datasets, as presented in the table below. While originally the Directive was rather focussed on road transport/traffic and safety thereof, it should be noted that the most recent implementing legislation also encompasses other transport modes and wider “mobility” aspects. In particular, the Delegated Act on EU-wide Multimodal Travel Information Services (EU/2017/1926) also covers the requirements set to Rail Infrastructure Managers under the TSIs for rail operation and traffic management, as well as vessel/ship and voyage/traffic datasets related to the VTMISS and RIS Directives, thereby excluding these datasets, too.

Table 13 – Data themes and datasets covered by the ITS Directive and Delegated Acts

<b>Delegated Act on Road safety-related minimum universal traffic information free of charge to users</b> <i>(Commission Delegated Regulation (EU) No 886/2013)</i>	
Status	In force (applicable since 2013)
Datasets in scope	<b>Location, category</b> (including temporary slippery road; animal, people, obstacles, debris on the road; unprotected accident area; short-term road works; reduced visibility; wrong-way driver; unmanaged blockage of a road; exceptional weather conditions) and <b>driving behaviour advice</b> concerning <b>road safety-related events</b> or conditions.
<b>Delegated Act on Information services for safe and secure parking places for trucks and commercial vehicles</b> <i>(Commission Delegated Regulation (EU) No 885/2013)</i>	
Status	In force (applicable since 2013)
Datasets in scope	<b>Static data</b> related to the <b>parking areas</b> (name, address, entry point longitude/latitude, total number of places, price and currency, etc.), <b>information on safety and equipment</b> of the parking area, contact information of the parking operator, and <b>dynamic data on availability</b> of spaces (full, closed, or number of places)
<b>Delegated Act on EU-wide real-time traffic information services</b> <i>(Commission Delegated Regulation (EU) No 962/2015)</i>	
Status	In force (applicable since 2017)

Datasets in scope **Static road data** (including road network links and their physical attributes, road classification, traffic signs, speed limits, traffic circulation plans, freight delivery regulations, locations of tolling stations, identification of tolled roads with user charges and payment methods, location of parking places and services areas, location of charging points for electric vehicles and conditions for their use, location of compressed natural gas, liquefied natural gas, liquefied petroleum gas stations, location of public transport stops and interchange points, location of delivery areas), **dynamic road status data** (including road/lane/bridge closures, bans on heavy goods vehicles, roadworks, accidents and incidents, dynamic speed limits, poor road conditions, etc.) and **traffic data** (volume, speed, location and length of traffic queues, travel times, waiting time at border crossings to non-EU Member States).

### Delegated Act on EU-wide Multimodal Travel Information Services

(Commission Delegated Regulation (EU) 2017/1926)

Status In force (applicable since 2017)

Datasets in scope This Delegated Act applies to all transport networks and modes in the EU, such as schedule based (air, rail including high speed rail, conventional rail and light rail, long-distance coach, maritime including ferry, metro, tram, bus, trolley-bus, cableways), transport on demand (shuttle bus, shuttle ferry, taxi, ride-hailing, car-share, car-pool, car-hire, bike-share, bike-hire, dial-a-ride) and personal based (car, motorcycle, bicycle, walking).

**Static travel and traffic data** (including location – address, topographic places, points of interest; trip plan computation for scheduled modes – connection links, network topology, transport operators, timetables, hours of operation, vehicles and accessibility, etc.; trip plan computation for personal road transport modes – road network, cycle network, pedestrian network<sup>200</sup> and accessibility facilities; location for demand responsive modes – park & ride stops, bike sharing stations, car-sharing stations, publicly accessible refuelling/charging stations, secure bike parking, detailed common standard and special fare query for all scheduled modes, etc.) and **dynamic travel and traffic data** (including passing times, trip plans and auxiliary information, disruptions, delays, cancellations, estimated departure and arrival times, cycling network closures, car/bike sharing availability, car parking spaces available, etc.)

Second, the scope of the ITS Directive as presented in the table above is still evolving, meaning that certain data/information not explicitly covered 'today' may be subject to inclusion 'tomorrow'. Indeed, the updated work plan on the implementation of the ITS Directive<sup>201</sup> foresees:

- For the Delegated Act on EU-wide real-time traffic information services, a possible extension of the geographical scope of the current specifications to relevant data types at urban level, in view of covering the whole road transport network (timeline 2019-2020), as well as looking into the accessibility of additional data types such as recharging and refuelling points, UVAR data, vehicle data for road operation purposes etc. (priority area I of the ITS Directive).
- For the Delegated Act on Multimodal Travel Information Services, to look into the accessibility of static and dynamic information to support interoperable payment and ticketing, and specifically pricing information on the whole road network (timeline 2018-2020).

Beyond the restrictions posed by the ITS Directive, in the area of air transport, it became apparent that ATM data used in the context of air navigation services also falls out of the scope of this study. This is primarily because Article 13 of the EATMN Directive provides for a restricted access regime for ATM data,

<sup>200</sup> These data points overlap in the Delegated Act and the ITS.

<sup>201</sup> See: [https://ec.europa.eu/transport/sites/transport/files/legislation/c20188264\\_en.pdf](https://ec.europa.eu/transport/sites/transport/files/legislation/c20188264_en.pdf)

which can only be used for operational purposes, by certain bodies: “[...] *relevant operational data shall be exchanged in real-time between all navigation service providers, airspace users and airports, to facilitate their operational needs*”, that “*data shall be used only for operational purposes*”. In addition, the Directive requires that “*certified services providers, airspace users and airports [...] establish standard conditions of access to their relevant operational data.*” Since ATM data can only be shared or reused between navigation service providers, airspace users and airports for operational purposes – it is essentially impossible to make this data fully available for commercial reuse. Lastly, ATM data have close ties with security and defence, which the PSI Directive recognises as out of scope.

To come back to the “*undefined*” status of the thematic area introduced earlier on in this section, the flip side of the coin in the current situation is the *possibility* to go about the exercise from a more “creative” perspective when it comes to the identification of the (remaining) datasets in scope. In other words, thanks to the “loose” interpretation of the thematic area i.e. “*anything possibly related to the movement (before, during and after) of people, goods and things, including related means of transport/vehicles as well as networks/infrastructure/geospatial attributes*”, the study team was able to find certain specific datasets which could still be analysed for the purposes of the study, despite the restrictions posed by other legislation explained above. Additionally, this is a key advantage for any future revision of the secondary legislation relating to the HVDs to be considered under the PSI Directive.

Among the list of “mobility” datasets already in possession of Member States as per the previously outlined EU legislation, datasets which can be considered as potential high-value datasets for the purposes of this study, but not necessary in the scope of the Open Data Directive, are those pertaining to:

- **Transport networks** data under the INSPIRE Directive. It should be noted that while these datasets have also been identified as relevant under the Geospatial thematic area, these will be assessed as part of the mobility thematic area (only);
- **Inland waterway and river infrastructure**-related data covered by the RIS Directive. It should be mentioned that ongoing efforts by Member States related to the provision of inland waterway infrastructure data in the context of the TENtec Portal<sup>202</sup> will be taken into account, and synergies will be sought as relevant. It should be noted that inland waterways represent only a small segment of the ‘mobility’ domain and concerns only a few Member States (those which have navigable inland waterways, see section 3.6.2.1). While the various datasets considered hereafter can be considered of high value for the sector specifically (as presented in section 3.6.1.2), their reusability outside the inland waterway community, and therefore, their impact on the overall EU data economy can be expected to be limited compared to other HVDs considered in the study.

Table 14 – Data themes and datasets in scope of the analysis for PSI mobility thematic area

<b>Transport networks</b> <i>as per INSPIRE Directive</i>	<b>Inland waterway infrastructure data</b> <i>as per RIS Directive</i>
The spatial data theme transport networks includes datasets related to:	Member States in the scope of the Directive are required to provide RIS users:
<ul style="list-style-type: none"> <li>• Road, rail, air and water transport networks and related infrastructure (topographic features)</li> </ul>	<ul style="list-style-type: none"> <li>• At least in an accessible electronic format: all relevant data concerning navigation and voyage planning on inland waterways such as</li> </ul>

<sup>202</sup> See : <https://ec.europa.eu/transport/infrastructure/tentec/tentec-portal/map/maps.html>

- Links between different networks (multimodal nodes)
- Referencing of transport flows enabling navigation services

Strong links are to be noted with other spatial themes including hydrography and addresses.

waterway axis with km indication, restrictions for vessels or convoys in terms of length, width, draught and air draught, operation times of restricting structures, locks and bridges, location of ports and transshipment sites, reference data for water level gauges, at least in an accessible electronic format; this data is made available in the form of the RIS Index through the European Reference Data Management System (ERDMS)<sup>203</sup>, operated by the Commission

- Electronic navigational charts, for all inland waterways of class Va and above (when appropriate against a reasonable cost-related charge);
- Notices to skippers, including water level (or maximum allowable draught) and ice reports of their inland waterways, as standardized encoded and downloadable messages. The standardised message shall contain at least the information necessary for safe navigation.

### *Inland waterway and river infrastructure data as per RIS Directive*

The RIS Directive regulates data provision by and exchange with a wide array of parties (both public and private), however, as discussed above, the current study considers only those datasets which relate to waterway infrastructure. As such, any vessel, traffic<sup>204</sup> or freight related data are out of scope of this study.

In particular, should be considered as “waterway infrastructure data” under this study, all waterway and **fairway** related datasets, including both **static and dynamic/urgent data**. In a view to ensure coherence with the RIS Directive and existing data provision efforts by Member States, the provisions of the RIS Directive should be respected and synergies should be sought to the extent possible with the TENtec portal<sup>205</sup>, which stores infrastructure-related data concerning all modes of transport. The TENtec Open Method of Coordination (OMC) Glossary<sup>206</sup> identifies the following parameter categories relevant to inland waterways’ infrastructure:

- 02 - Inland Waterways (hereinafter 02 ILW/..)
- 03 - IWW Locks (hereinafter 03 LO/..)
- 04 - IWW lock chambers (hereinafter 04 LC/..)
- 05 - IWW Bridges (hereinafter 05 BR/..)
- 06 - Ports (hereinafter 06 PO/..)
- 07 - Port Terminal (hereinafter 07 PT/..)

<sup>203</sup> [https://webgate.ec.europa.eu/RIS/EUERDMS\\_WEB](https://webgate.ec.europa.eu/RIS/EUERDMS_WEB)

<sup>204</sup> Traffic refers to pre-journey data (e.g. raw timetables and fares data, seat maps, contract of carriage conditions, planned disruptions, delays or cancellations), on-journey data (e.g. delays and cancellations), and post-journey (consolidated delays and cancellations of rail services, actual occupancy rates).

<sup>205</sup> See TENtec Portal (2020), <https://ec.europa.eu/transport/infrastructure/tentec/tentec-portal/site/en/tentec.html>

<sup>206</sup> See TENtec OMC Glossary (2017), [https://ec.europa.eu/transport/sites/transport/files/tentec\\_omc\\_glossary.pdf](https://ec.europa.eu/transport/sites/transport/files/tentec_omc_glossary.pdf)



- 08 - Mooring Places (hereinafter 08 MP/..)

The table below highlights the “minimum” data points included in the inland waterway category, following the implementing legislation on Fairway Information Services (FIS). It also identifies any links with existing parameters in the TENtec Glossary presented above.

Table 15 - High-Value Datasets in the mobility thematic area, inland waterways

Static	Dynamic / Urgent
<ul style="list-style-type: none"> <li>• Fairway characteristics (02 ILW/01-21 &amp; 33)</li> <li>• Long-time obstructions in the fairway and reliability (02 ILW/22-27)</li> <li>• Rates of waterway infrastructure charges (02 ILW/34)</li> <li>• Other physical limitations on waterways (03 LO/1; 04 LC/1-5 &amp; 8-9; 05 BR/01-06)</li> <li>• Regular lock and bridge operating times (03 LO/2 &amp; 4-9; 04 LC/6-7; 05 BR/6)</li> <li>• Location and characteristics of ports and transshipment sites (PO/3-65; 07 PT/1-2)</li> <li>• List of navigation aids and traffic signs</li> <li>• Navigation rules and recommendations</li> </ul>	<ul style="list-style-type: none"> <li>• Water depths contours in the navigation channel (02 ILW/14)</li> <li>• Temporary obstructions in the fairway (02 ILW/22-25)</li> <li>• Present and future water levels at gauges (02 ILW/15-16)</li> <li>• State of the rivers, canals, locks and bridges (02 ILW; 03 LO; 04 LC; 05 BR)</li> <li>• Restrictions caused by flood and ice (02 ILW/24-25; 03 LO/6-7)</li> <li>• Meteorological data (incl. wind direction) Short term changes of lock and bridge operating times (03 LO/2; 05 BR/6)</li> <li>• Short term changes of aids to navigation</li> </ul>

Building on the static and dynamic/urgent data highlighted in the table above, should also be considered in the scope of the current study – and as potential high-value datasets, **digital representations of the inland waterway network** as prescribed by the technical specifications for inland electronic navigational charts (ECDIS)<sup>207</sup>. In particular, electronic inland ECDIS made available by Member States should contain the following information/features:

- Waterway axis with kilometres indication (02 ILW/01-07)
- Links to the external xml-files with operation times of restricting structures (03 LO/2; 05 BR/6)
- Location of ports and transshipment sites (06 PO/58-59; 07 PT/1-2)
- Reference data for water level gauges relevant to navigation (02 ILW/15-16)
- Bank of waterway at mean water level
- Shoreline construction (02 ILW/9-12; 04 LC/1-4; 05 BR/3-6)
- Contours of locks and dams (02 ILW/15-16)
- Boundaries of the fairway/navigation channel (02 ILW)
- Isolated dangers in the fairway/navigation channel under and above water (02 ILW/22-25)
- Official aids-to-navigation (e.g. buoys, beacons, lights, notice marks)

The benefits and use cases of these datasets and charts will be detailed in section 0 below.

<sup>207</sup> See: <https://eur-lex.europa.eu/legal-content/GA/ALL/?uri=CELEX%3A32013R0909>



### *Transport networks data as per INSPIRE Directive*

As indicated above, the Transport Networks data theme is defined within the INSPIRE Directive as: "Road, rail, air and water transport networks and related infrastructure. Includes links between different networks. Also includes the trans-European transport network"<sup>208</sup>. In the same vein, a transport network can be defined as a spatial network that describes a structure that allows movement of vehicles and/or flows of other commodities. The network is the entirety of transport lines or connections with nodes, links, hubs, or corridors through which flows pass.<sup>209</sup> Thanks to mathematical graph theory, a topological and mathematical representation of the nature and structure of these networks can be produced schematically and digitally, leading to the development of transport network data (models) and schemes.<sup>210</sup> In particular, should be considered as "transport network data" under the current study, all datasets/-models published by Member States under the Transport Networks data theme on the INSPIRE Geoportal<sup>211</sup>. In line with the different modes of transport, the INSPIRE data specifications<sup>212</sup> distinguish between six sub-themes:

- Common transport network elements
- Road transport networks
- Rail transport networks
- Water transport networks
- Air transport networks
- Cableways networks

For each sub-theme's network schemes, the table on the next page provides a list of specific features or "minimum" data points to be included. It should be noted, as explained in the INSPIRE specifications for the Transport Networks data theme that any "non-geographic data", such as "asset condition report" (any information on the condition/status of transport assets e.g. means of transport, "traffic flow records" (any information on traffic flows recorded on transport networks), "images of assets" (any pictures of transport assets), "statistics", "timetables", "noise data" (any information on noise produced by transport/on transport networks) and so on, is out of scope of the specifications and, therefore, of this impact assessment study. Further, some of these elements are covered by the ITS Directive, which *de facto* makes them out of scope for the current study.

---

<sup>208</sup> OJ L 228, 9.9.1996, p. 1. Decision as last amended by Council Regulation (EC) No 1791/2006 (OJ L 363, 20.12.2006, p. 1

<sup>209</sup> Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1-3), 1-101. <https://doi.org/10.1016/j.physrep.2010.11.002>, pp. 1-3.

<sup>210</sup> Rodrigue, J. P., Comtois, C., & Slack, B. (2013). *The Geography of Transport Systems*, pp. 328-330; *The Geography of Transport Systems. What is Transport Geography?* Via: [https://transportgeography.org/?page\\_id=40](https://transportgeography.org/?page_id=40)

<sup>211</sup> See: <https://inspire-geoportal.ec.europa.eu/overview.html?view=themeOverview&theme=tn>

<sup>212</sup> See: <https://inspire.ec.europa.eu/Themes/115/2892>

Table 16 – List of INSPIRE transport network features

Common	Road	Rail	Water		Air		Cableways
<ul style="list-style-type: none"> <li>• Access Restriction</li> <li>• Condition Of Facility</li> <li>• Maintenance Authority</li> <li>• Marker Post</li> <li>• Owner Authority</li> <li>• Restriction For Vehicles</li> <li>• Traffic Flow Direction</li> <li>• Transport Area</li> <li>• Transport Link</li> <li>• Transport Link Sequence</li> <li>• Transport Link Set</li> <li>• Transport Network</li> <li>• Transport Node</li> <li>• Transport Object</li> <li>• Transport Point</li> <li>• Transport Property</li> <li>• Vertical Position</li> </ul>	<ul style="list-style-type: none"> <li>• E-Road</li> <li>• Form Of Way</li> <li>• Functional Road Class</li> <li>• Number Of Lanes</li> <li>• Road</li> <li>• Road Area</li> <li>• Road Link</li> <li>• Road Link Sequence</li> <li>• Road Name</li> <li>• Road Node</li> <li>• Road Service Area</li> <li>• Road Service Type</li> <li>• Road Surface Category</li> <li>• Road Width</li> <li>• Speed Limit</li> <li>• Vehicle Traffic Area</li> </ul>	<ul style="list-style-type: none"> <li>• Design Speed</li> <li>• Nominal Track Gauge</li> <li>• Number Of Tracks</li> <li>• Railway Area</li> <li>• Railway</li> <li>• Electrification</li> <li>• Railway Line</li> <li>• Railway Link</li> <li>• Railway Link Sequence</li> <li>• Railway Node</li> <li>• Railway Station Area</li> <li>• Railway Station Code</li> <li>• Railway Station Node</li> <li>• Railway Type</li> <li>• Railway Use</li> <li>• Railway Yard Area</li> <li>• Railway Yard Node</li> </ul>	<ul style="list-style-type: none"> <li>• Beacon</li> <li>• Buoy</li> <li>• CEMT Class</li> <li>• Condition Of Water Facility</li> <li>• Fairway Area</li> <li>• Ferry Crossing</li> <li>• Ferry Use</li> <li>• Inland Waterway</li> <li>• Marine Waterway</li> <li>• Port Area</li> <li>• Port Node</li> <li>• Restriction For Water Vehicles</li> <li>• Traffic Separation Scheme</li> <li>• Traffic Separation Scheme Area</li> <li>• Traffic Separation Scheme Crossing</li> <li>• Traffic Separation Scheme Lane</li> </ul>	<ul style="list-style-type: none"> <li>• Traffic Separation Scheme Roundabout</li> <li>• Traffic Separation Scheme Separator</li> <li>• Water Link Sequence</li> <li>• Water Node</li> <li>• Water Traffic Flow Direction</li> <li>• Waterway</li> <li>• Waterway Link</li> <li>• Waterway Node</li> </ul>	<ul style="list-style-type: none"> <li>• Aerodrome Area</li> <li>• Aerodrome Category</li> <li>• Aerodrome Node</li> <li>• Aerodrome Type</li> <li>• Air Link</li> <li>• Air Link Sequence</li> <li>• Air Node</li> <li>• Air Route</li> <li>• Air Route Link</li> <li>• Airspace Area</li> <li>• Apron Area</li> <li>• Condition Of Air Facility</li> <li>• Designated Point</li> <li>• Element Length</li> <li>• Element Width</li> <li>• Field Elevation</li> <li>• Instrument Approach Procedure</li> <li>• Lower Altitude Limit</li> <li>• Navaid</li> <li>• Procedure Link</li> </ul>	<ul style="list-style-type: none"> <li>• Runway Area</li> <li>• Runway Centreline Point</li> <li>• Standard Instrument Arrival</li> <li>• Standard Instrument Departure</li> <li>• Surface Composition</li> <li>• Taxiway Area</li> <li>• Touch Down Lift Off Area</li> <li>• Upper Altitude Limit</li> <li>• Use Restriction</li> </ul>	<ul style="list-style-type: none"> <li>• Cableway Link</li> <li>• Cableway Link Sequence</li> <li>• Cableway Link Set</li> <li>• Cableway Node</li> </ul>

The benefits and use cases for transport networks schemes will be detailed in section 0 below.

### *Other datasets to consider for possible future revisions*

As previously mentioned, the study team was confronted with diverging interpretations by stakeholders of the overall thematic area, including the scope of application the ITS Directive. Indeed, some stakeholders' experience is that certain datasets which, based on the legal mapping, should be covered by the ITS Directive, are in practice not necessarily or completely so. This section provides a summary of these findings.

While the provision of **rail (traffic) data** appears to be fully covered by the ITS Directive on the basis of the legal mapping, it appears that in practice, there is room for certain Member States to restrict the provision, especially of **dynamic data**. Rail (traffic) data reusers note that the following datasets of high value could still be covered by the PSI Directive:

- Pre-journey data, with the right to reuse to do price prediction: raw timetables and fares data, seat maps, contract of carriage conditions, planned disruptions, delays or cancellations;
- On-journey data, with the right to reuse to do travel and delay prediction: real time platforms, real time delays and cancellations;
- Post-journey, with the right to reuse to do delay prediction and repay: Day+1 accurate and consolidated delays and cancellations of rail services, actual occupancy rates.

According to stakeholders, in certain Member States, these datasets are fully available only to the national (public) transport operators, whereas these could widely be used by third party service providers in order to make pricing, delay, voyage predictions for consumers. Nevertheless, as indicated by DG MOVE, these points will be looked at by the upcoming revision of the Delegated Regulation 2017/1926. Therefore, it would not be necessary for the PSI Directive to take it into account in order to avoid redundancies.

In addition to rail data, stakeholders consider that certain **road transport data**, in particular **technical vehicle data**, could also still fall outside the practical scope of the ITS Directive. In the Netherlands, the RDW has been publishing and updating daily the following datasets as open data since 2012<sup>213</sup>:

- Vehicle base registry including approximately 100 attributes per vehicle, for 14.5 million vehicles;
- European type approval information on vehicles;
- Recall-information on vehicles and vehicle-parts;
- Registry of Dutch companies performing specific legal tasks such as PTI's or modifications on cars;
- Information on the results of individual PTI's and other vehicle inspections.

Currently, the RDW counts 150 million API-calls per month for the aforementioned data. A call can vary from the data of a given vehicle, to a full download of the data of all vehicles registered. The platform provides specific information for re-users to facilitate the re-use of data, such as extended API documentation for each dataset.<sup>214</sup> Stakeholders believe that these datasets could be widely reused across the EU, if only all Member States would provide such information as open data. From the Dutch experience, main reusers and/or use cases are the following:

- Municipalities for their citizen communication on access to environmental zones;
- Automatic number-plate recognition (ANPR) to detect repeating visits at certain points of interest;
- Automotive industry after sales services to search and find replacement components based on the technical vehicle information;
- Different tools for calculating vehicle value and/or maintenance costs.

---

<sup>213</sup> See: <https://opendata.rdw.nl/>

<sup>214</sup> See: <https://dev.socrata.com/foundry/opendata.rdw.nl/m9d7-ebf2>

Finally, as regards **waterborne** transport, **ship/vessel automatic identification system (AIS) tracking data** is also considered as high value for reuse by certain stakeholders. Such data is already collected by Member States following the RIS Directive, in order to be able to track movements of ship. However, these datasets were considered out of scope for the current study, due to sensitivity/security considerations.

According to stakeholders, AIS tracking data could nonetheless be of high value for reuse, notably in the context of transport and logistics services provision. Indeed, the availability of AIS tracking data as open data would facilitate the tracking and tracing services for waterborne transport. By knowing the exact position of vessels and therefore goods transported thereon, it would be possible for shipping operators to communicate the most accurate time of departures/arrivals at ports. In turn, this would allow to further plan the next pick-ups throughout the logistics value chain, thereby enhancing the effectiveness and efficiency of multi-modal transport.

On the other hand, as indicated by DG MOVE, AIS can already be received freely now. However, AIS data is considered sensitive and is subject for GDPR, as it allows for the identification of individuals (e.g. a ship owner that lives on-board his/her vessel). Therefore, an in-depth assessment on the provision of AIS openly should be first conducted.

Overall, the study team recommends to closely follow-up the implementation of the ITS Directive as regards the aforementioned datasets, and where relevant and only if necessary, possibly consider these as potential high-value datasets to be analysed as part of the extensions of the Implementing Act.

### 2.7.2 Analysis of the value of these datasets

As presented in the next sub-sections, the main re-users of inland waterway infrastructure data and transport networks data are some sort of “mobility” (service) providers be it for people and/or goods. Their main benefits are related to the economic, social and environmental spheres, which overall are key enablers for an EU-wide shift towards greener and safer multimodal mobility. These datasets can also be expected to have strong reuse potential in the domain of connected autonomous vehicles and their further development, both on land and waterborne transport modes.

#### *Inland waterways and river infrastructure data*<sup>215</sup>

The main benefit of the creation and provisions of inland waterway and river infrastructure data as regulated by Directive 2005/44/EC is that they enable the provision of **River Information Services (RIS)** free of charge. Overall, RIS aim at contributing to a safe and efficient transportation system, and more notably, to the full utilisation of inland waterways, thereby justifying the objective to render the available technology interoperable with other modes of transport.<sup>216</sup> There are three levels of information offered by RIS, namely **Fairway Information (FIS)**, Tactical Traffic Information (TTIS), and Strategic Traffic Information (STIS).<sup>217</sup> Naturally, among these information services, inland waterway and river

---

<sup>215</sup> It should be noted that a key reference used throughout this section, and the indicator table presented hereafter, is the evaluation report of the RIS Directive (2020), which naturally identified a number of benefits and impacts related to inland waterway and river infrastructure data: European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*.

<sup>216</sup> European Commission. (2007). *Commission Regulation No 414/2007 of 13 March 2007 concerning the technical guidelines for the planning, implementation and operational use of river information services (RIS) referred to in Article 5 of Directive 2005/44/EC of the European Parliament and of the Council on harmonised river information services (RIS) on inland waterways in the Community*. Via <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007R0414&from=EN>, p. 7

<sup>217</sup> European Commission. (2007). *Commission Regulation No 414/2007 of 13 March 2007 concerning the technical guidelines for the planning, implementation and operational use of river information services (RIS) referred to in Article 5 of Directive 2005/44/EC of the European Parliament and of the Council on harmonised river information services (RIS) on inland waterways in the Community*.

infrastructure data contribute particularly to the provision of FIS. It shall be noted that the identified high-value datasets in this analysis are already regulated by Directive 2005/44/EC on harmonised river information services (RIS) on waterways in the Union (RIS Directive), which is currently being evaluated according to the Better Regulation Principles of the European Commission. This evaluation feeds into a possible revision of the legislative framework for River Information Services in the future.

Generally speaking, RIS and FIS are employed by the inland waterway ecosystem to improve planning of traffic, route, voyage, and logistics, and to enhance safety on the fairways. In short, these services allow commercial users of inland waterways to use the fairway infrastructure more efficiently, while authorities are able to manage traffic and safety in ways previously not possible. In the following sections it will become clear that the reuse of inland waterway and river infrastructure data in the context of RIS and FIS represents valuable **economic, social and environmental benefits** to the sector as such, as well as to the overall EU transport sector.

Indeed, Europe has more than **37 000 kilometres of canals and rivers connecting hundreds of important industrial cities and areas**. Two main rivers, the Rhine and the Danube are the core of the network. The Netherlands, Belgium, Luxembourg, France, Germany, Austria, Slovakia, Hungary, Poland, Croatia, Serbia, Romania, Bulgaria, Moldova and the Ukraine form the core network of about 10 000 kilometres. Yet also a wealth of canals and tributaries link several smaller economic centres and cities. Of the 21 Member States that have inland waterways, 13 are interconnected through waterway networks, highlighting the importance of seamless cross-border data-exchange.<sup>218</sup> Of the 37 000 kilometres of inland waterways, 20 000 are accessible to 1000 tonne vessels. Moreover, on average, 550 million tonnes are shipped by waterways every year. In 2017, the main types of goods transported were metal ores, coke, and refined petroleum products and products of agriculture. These accounted for more than half of all goods transport on EU inland waterways. Transport performance of containers rose by almost 5% in that same year. This could indicate a transition towards an **energy-efficient and low-carbon economy**, as well as **increased integration of inland waterway transport in the supply chain**.<sup>219</sup>

In fact, the **strengths** of inland waterway transport lie mainly in its ability to **carry vast volumes of goods per vessel, its low shipping costs, and its ecological friendliness**. In addition, it has a **high level of safety, low infrastructure costs, and still free capacity**. The key limitations for this transport mode are its reliance on variable fairway conditions, especially on free-flowing river stretches, and the resulting variable load factor for the vessels. For example, prolonged drought due to climate change can severely impact the sector.<sup>220</sup> Despite the vast network, **inland waterways still have an unfulfilled capacity potential** that is not fully exploited. Hence, its future development requires the introduction of modern concepts, technologies, and solutions. If not, inland waterway transport could lose the aforementioned competitive advantages compared to other transport modes. Overall, inland waterways represents between 6% and 7% of the total EU freight transport.<sup>221</sup> It is not unconceivable that the further use of RIS and FIS – which exist in a harmonised way since 2005, could increase or at least consolidate that share. This is why **inland waterway and river infrastructure data** are **considered key enablers in this development** and thus being regulated by Directive 2005/44/EC, as further detailed hereafter.

---

<sup>218</sup> European Commission - DG MOVE. *Inland waterways*. Via: [https://ec.europa.eu/transport/modes/inland\\_en](https://ec.europa.eu/transport/modes/inland_en)

<sup>219</sup> Central Commission for the Navigation of the Rhine. (2019). *Inland Navigation in Europe: Market Observation. Annual Report 2019*. Via: [https://inland-navigation-market.org/wp-content/uploads/2019/11/ccnr\\_2019\\_Q2\\_en-min2.pdf.pdf](https://inland-navigation-market.org/wp-content/uploads/2019/11/ccnr_2019_Q2_en-min2.pdf.pdf), p. 28.

<sup>220</sup> IMPREX (EU). (2017). *Vulnerability of Inland Waterway Transport and Waterway Management on Hydrometeorological Extremes*. Via: <https://www.imprex.eu/system/files/generated/files/resource/d9-1-imprex-v2-0.pdf>, p. 27.

<sup>221</sup> European Commission - DG MOVE. (2019). *EU transport in figures: Statistical pocketbook 2019*, p. 76.

The use of inland waterway data renders the ILW sector more efficient and competitive since commercial data re-users can engage in more accurate voyage planning and execution, resulting in **better use of fairway infrastructure and improved resource planning**. Furthermore, precise RIS and FIS data and its communication among stakeholders enhances the **safety and predictability** of the overall sector significantly. While inland waterways are not present in all EU Member States, a well-performing sector benefits Europe as a whole. For example, waterways are used by operators from all Member States, and the shift to efficient multimodal transport benefits the EU27 in its entirety. Moreover, accurate voyage planning leveraged by RIS enables a better connection to other modes of transport, thereby contributing to objectives of **multimodality and socio-economic sustainability** of the overall EU transport sector.<sup>222</sup> This was acknowledged in the 2011 Transport White Paper as well, which recommends the further development of multimodal and inland waterway transport to prevent congestion and lift other environmental barriers to the **sustainable socio-economic growth of the EU**.<sup>223</sup> Otherwise, transport could lose its fundamental role for and of the EU economy and endanger around 11 million jobs.<sup>224</sup>

In line with the categorisation of inland waterway infrastructure datasets presented in section 0 above, FIS provides **dynamic information** (data that changes on a short timescale, e.g. water levels) as well as **static information** (data that changes now and then, e.g. regular operating times of locks and bridges) regarding the use and status of the inland waterway infrastructure. Additionally, **urgent information** on short-term changes to the fairway is also communicated through FIS. For example, when the water levels at sea and in the rivers, canals, and lakes change, it has immediate consequences for the users of waterways. Additionally, fairway traffic can be impeded by short-term changes in infrastructure operating times. That is why FIS related data measures and communicates water elevation, the status of barrages and regimes impacting the state of the fairway, malfunctions of aids to navigation, physical limitations on waterways, and information on vertical clearance, among others. FIS thereby supports tactical and strategic navigation decisions.<sup>225</sup>

In practice, in the framework of the RIS Directive, the provision of dynamic information is standardised in the **Notices to Skippers (NtS)**, while static information is encompassed in the **Inland ECDIS systems**. These systems and their specific benefits are discussed in the table below.

---

<sup>222</sup> European Commission. (2018). *Commission Staff Working Document SWD(2018)427 on Digital Inland Navigation*. Via: <https://ec.europa.eu/transport/sites/transport/files/legislation/swd20180427-digital-inland-navigation.pdf>, pp. 12-13.

<sup>223</sup> European Commission. (2011). *White Paper COM/2011/0144: Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system*. Via: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52011DC0144&from=EN>

<sup>224</sup> European Commission – DG MOVE (2019), *Transport in the European Union: Current Trends and Issues*, via: <https://ec.europa.eu/transport/sites/transport/files/2019-transport-in-the-eu-current-trends-and-issues.pdf>, p. 3.

<sup>225</sup> OJ. (2007). *Commission Regulation (EC) No 416/2007 of 22 March 2007 concerning the technical specifications for Notices to Skippers as referred to in Article 5 of Directive 2005/44/EC of the European Parliament and of the Council on harmonised river information services (RIS) on inland waterways in the Community*. Via <http://data.europa.eu/eli/reg/2007/416/oj>

Table 17 – Notices to Skippers and Inland ECDIS Systems, and their benefits

Description	Benefits <sup>226,227</sup>
<p><b>Notices to Skippers</b> As previously noted, the Notice to Skippers communicates <b>dynamic information</b> related to inland waterway infrastructure. Regulation No 416/2007 (amended by Implementing Regulation (EU) 2018/2032 standardises the NtS to provide a uniform structure of data-sets to facilitate the integration of notices in voyage-planning systems. Consequently, the Notice attributes to facilitating data-exchange between different countries. As such, the notice is compatible with the data-structure of Inland ECDIS to facilitate integration of Notices to Skippers into Inland ECDIS (see below). The overall objective is an obligation Member State authorities to provide information in a uniform manner relating to the <b>safety of inland navigation and information needed for voyage planning</b>.<sup>228</sup> The data format can be used both for publishing notices on the Internet (pull services) or for distribution by e-mail (push services).<sup>229</sup></p> <p>The NtS is mainly used to communicate water level and depth information (including ice information) in a standardised format. Moreover, NtS also provides information on the long-term and short-term status of the inland waterway infrastructure (i.e. bridges and locks), failures of aids to navigation, temporary blockages of waterway sections or other types of infrastructure, works, and weather messages. This data is crucial for inland navigation as it</p>	<ul style="list-style-type: none"> <li>• <b>Better voyage planning</b> resulting from the availability of real-time information on traffic, weather, water levels, etc., which can lead to more efficiency and cost-savings from better resource planning on the side of RIS users/users of the waterway.</li> <li>• <b>Improved safety and lower accident rate</b> consequential from better planning through traffic information, as perceived by skippers and authorities alike;</li> <li>• <b>Improved communication</b>, facilitated by the automatic translations of encoded notices to skippers, which is conditional to the messages being provided in line with the requirements of the Directive.</li> <li>• The standardised format leads to harmonised data-exchange, cross-country communication and interoperability with other systems (e.g. Inland ECDIS) – <b>ameliorated communication improves safety</b>;</li> <li>• After the initial costs associated with setting-up and implementing NtS, <b>authorities save on administrative costs</b> thanks to standardised, encoded and downloadable format. The benefits are mainly economical;</li> <li>• Skippers avoid wasting time (and paper) using paper NtS;</li> <li>• In the long term, <b>cost savings and harmonisation of notices to skippers</b> are expected to lead to an interoperable RIS system, and in turn <b>improve the competitiveness</b> (e.g. skippers can get to their destination faster) <b>and stimulate innovation</b> of the inland navigation sector as a whole. Ideally, this would lead to a modal shift and reduce the negative impact of the sector on</li> </ul>

<sup>226</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*. Via: <https://op.europa.eu/en/publication-detail/-/publication/1f0e2c53-7ebe-11ea-aea8-01aa75ed71a1> pp. 29-30.

<sup>227</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, pp. 20-22, 25-27.

<sup>228</sup> OJ. (2007). *Commission Regulation (EC) No 416/2007 of 22 March 2007 concerning the technical specifications for Notices to Skippers as referred to in Article 5 of Directive 2005/44/EC of the European Parliament and of the Council on harmonised river information services (RIS) on inland waterways in the Community*. Via <http://data.europa.eu/eli/reg/2007/416/oj>

<sup>229</sup> Comité européen pour l'élaboration de standards dans le domaine de la navigation intérieure (CESNI). *About RIS*. Via: <https://ris.cesni.eu/30-en.html>

Description	Benefits <sup>226,227</sup>
directly impacts <b>voyage planning, accessibility as well as safety</b> . <sup>230</sup>	the environment.
<p><b>Inland ECDIS Systems</b> Inland ECDIS systems (regulated by Implementing Regulation (EU) No 909/2013, amended by Implementing Regulation (EU) 2018/1973) seek to contribute to the safety, reliability, and efficiency of inland navigation by reducing the workload of the skipper steering the vessel and to enhance their situational awareness.<sup>231</sup> The RIS Directive stipulates that Member State authorities must ensure “electronic navigational charts suitable for navigational purposes are available to RIS users”.<sup>232</sup> Inland ECDIS communicate mainly <b>static information</b> (e.g. regular operating times of locks and bridges) regarding the use and status of the inland waterway infrastructure, and thereby support tactical and strategic navigation decisions. Furthermore, the technical requirements for electronic navigational charts specify that all kinds of geographical objects necessary for safe navigation (e.g. boundaries of the fairway, shoreline constructions, beacons, and so forth); fairway water depth information; and additional information from parties other than the competent authorities are to be included, provided it does not affect the minimum information required for safe navigation.<sup>233</sup></p>	<ul style="list-style-type: none"> <li>• Fairway information made available by electronic charts <b>improves resource planning of waterway users</b> as it enables users to efficiently plan voyages and navigate the waterways faster;</li> <li>• The availability of electronic charts <b>enhances the information on the fairway, which improves the safety</b> on waterways as RIS users have a clear view on use and status of the inland waterway infrastructure. <b>Improved navigation decisions reducing the number of accidents are crucial</b> in this point;</li> <li>• <b>Improves interaction with other traffic management systems</b> of other transport modes, enabling the shift towards multimodal transport;</li> <li>• The overall use of the infrastructure is optimised;</li> <li>• Inland ECDIS contributes to the <b>harmonised exchange of information between RIS providers</b> and cross-border exchange of information.</li> </ul>

<sup>230</sup> Comité européen pour l'élaboration de standards dans le domaine de la navigation intérieure (CESNI). *About RIS*. Via: <https://ris.cesni.eu/30-en.html>

<sup>231</sup> Comité européen pour l'élaboration de standards dans le domaine de la navigation intérieure (CESNI). *About RIS*. Via: <https://ris.cesni.eu/30-en.html>

<sup>232</sup> OJ. (2005). *Directive 2005/44/EC Of the European Parliament and of the Council of 7 September 2005 on harmonised river information services (RIS) on inland waterways in the Community*. Via <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:255:0152:0159:EN:PDE>, Article 4 (3b).

<sup>233</sup> OJ. (2005). *Directive 2005/44/EC Of the European Parliament and of the Council of 7 September 2005 on harmonised river information services (RIS) on inland waterways in the Community*. Via <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2005:255:0152:0159:EN:PDE>, Annex II (2) “Inland ECDIS”, Directive 2005/44/EC



Last but not least, in terms of final beneficiaries of the services provided thanks to the reuse of inland waterway and river infrastructure data, the combination of FIS data and sub-services are used by a variety of stakeholders, including shipmasters, lock/bridge operators, waterway authorities, terminal operators, calamity centres, fleet managers and cargo shippers. Every stakeholder leverages the data in a way that is most fitting to their agenda. First of all, mainly for statistical, asset-management, and planning purposes, **inland waterway authorities** use fairway information. Fairway information data enables them to more effectively manage the traffic, and thus, the safety on the fairway. Similarly, public search and rescue services use FIS data to respond quicker and more efficiently in case of calamities. Furthermore, **lock, bridge, and berth operators** must be aware of the condition of the fairway and possible constraints to manage traffic most efficiently. They use fairway information data for lock, bridge, and berth operations to predict high traffic and, if possible and appropriate, to make provisions to ensure effective planning. Consequently, both authorities and infrastructure operators' **main benefit stemming from fairway information is social, as FIS enhances safety.**<sup>234</sup>

On the other hand, commercial actors leverage fairway information services to improve their operations and save on resources, such as fuel costs. Next to experiencing increased safety on the fairway, **the benefits are mainly economical of nature** for this group. For instance, data on water depth influences how much ships can be charged. As such, the provision of water depth data can lead to optimal use of the fairway, as it helps commercial parties to load their vessel in such a way as to ensure an optimal draft for the entire journey. In the case of seamless and timely FIS data-sharing they can modify their itinerary and plan their resources accordingly. More precisely, **skippers** use the information on the availability of fairways as conditions for navigational safety and route planning. Traffic information on locks and bridges may, for example, be used to deviate from the travel planning. Put differently, traffic information can give additional input on journey durations of vessels and, thus, result in an improved estimated time of arrival for the managed cargo. In order to dispatch their barges efficiently, **fleet managers** use accurate fairway information about the inland waterway infrastructure. Moreover, **cargo receivers and senders** gain more information about the transport possibilities and the position and/or routes of their goods to enhance the management of the transport chain. Furthermore, **terminal operators** benefit from reliable voyage information, leveraged by fairway information services, to plan and improve transshipment processes. Lastly, **service providers from other transport modes** may use fairway information services to integrate inland waterways transport into their chain of operations. It may also be used by road transport enterprises to gain insights on waiting times at movable objects, such as open bridges.<sup>235</sup>

A third and last category of benefits leveraged by the use of FIS data is **environmental** in character. As FIS data enhances the efficiency of inland waterway transport through improved use of resources, emissions decrease as a result thereof. Furthermore, heightened safety also carries environmental benefits. This can be attributed to the fact that fewer accidents occur, thereby averting oil and other dangerous cargo discharges in the natural surroundings. Additionally, precise information on accidents allows authorities to act faster and to minimise the environmental impact of spills.

To conclude, the table below highlights the key value characteristics for inland waterway and river infrastructure data.

---

<sup>234</sup> European Commission (2019), *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS): Final Report – Technical Annexes*, via: <https://op.europa.eu/s/n67h>, pp. 73-74.

<sup>235</sup> CoRISMa, *Functional and technical requirements study for a European FIS and TIS Register and Portal – Milestone 6*, p. 8.

Table 18 – Assessment of value for inland waterway and river infrastructure data

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references:</b>
Economic (benefit area)	Competition		Overall, there is little evidence that the use of inland waterway data enhanced competitiveness of the sector. However, the sector benefitted from professionalisation stemming from data for navigation and resource planning. Examples include providers using data for investments in equipment, fleet maintenance, repairs, accident prevention, etc., which professionalises the sector. <sup>236</sup>
	Economic output		There is no hard evidence to support that the use of RIS and FIS data increases economic output. As an indicative measure, the implementation of the RIS Directive led to a 1% growth in tank barge market in terms of tonne-km. However, this effect is not statistically significant, as growth in the inland navigation sector is more likely caused by external effects, such as economic growth, industrial production and other factors. <sup>237</sup>
	Employment		There is no quantitative proof that the use of inland waterway infrastructure data increased employment in the sector. However, the promotion of the use of RIS and FIS has led to the creation of new educational programmes, training personnel in ECDIS and related systems. <sup>238</sup> It can be assumed that the introduction of high-tech solutions has increased the quality of work in the sector, as well as higher safety standards resulting from the use of inland waterway transport data.
	International competitiveness		Similar to “competitiveness” characteristic above, there is no evidence that the use of inland waterways data has improved international competitiveness.
	Productivity and commercialisation		The use of inland waterway data has enabled data-driven decision-making through the time and money saved by waterway users resulting from better voyage planning, efficient ship loading based on current fairway conditions, and interoperability with the full supply chain and other modes of transport in Europe. The optimised use of the fairway, leveraged by FIS data, has the potential to improve productivity of the sector. <sup>239</sup>
	Public-private coordination		The quality of coordination between public and private organisations is crucial in the case of inland waterway data since the information relating to safety is provided and certified by the waterway authorities in a harmonised way. This, in turn, is reused by users of the fairway. <sup>240</sup>

<sup>236</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, pp. 17-18.

<sup>237</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 52.

<sup>238</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 49.

<sup>239</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 50.

<sup>240</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 79.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references:</b>
Environmental	Energy management and efficiency		Measuring energy (fuel) use of the IWT sector is difficult, since vessels operating on the fairway differ in size and no standardised measurements are non-existent. A study conducted in the Netherlands concluded that waterway users using RIS systems reportedly used 6.7% less fuel, compared to vessels that did not use RIS. <sup>241</sup> Furthermore, it is estimated that, between 2009 and 2017, the average annual reduction in the energy efficiency of the final energy consumption of inland navigation for the 13 relevant Member States is 1.86%. However, the data does not take into consideration external factors such as the more rigorous technical requirements and fleet renewal, or high fuel prices pushing skippers to ecologic sailing. However, on the basis that 40 percent of RIS users use their equipment to minimise fuel consumption, it can be inferred that the RIS and FIS data has a good return on investment in this regard. <sup>242</sup>
	Environment management		Environmental protection through more efficient calamity abatement is difficult to quantify, as it is nearly impossible to compare accidents in MS over time. Nevertheless, inland waterway data enhances safety on the fairway resulting in fewer accidents, and fewer oil spills, which decreases the negative environmental impact. Relatedly, if information about a ship is known through data-sharing, calamity abatement will be faster and more efficient, which can also reduce the impact an accident has on the environment. <sup>243</sup> As established above, better awareness and proper usage by authorities and end-users of RIS and FIS data, it is fair to assume that fuel consumption, and thus pollution and climate change, are impacted. Eurostat data indicates a fall in overall emissions of carbon dioxide, nitrous oxide and methane for water transport in the 13 Member States since 2006. <sup>244</sup> As mentioned above, however, correlation does not imply causation, since more stringent regulations regarding cleaner engines have been implemented.
Innovation & AI	Entrepreneurialism and private sector innovation		The standardised notices have led to innovation and new developments such as (semi)-autonomous sailing. New technological needs have also been spurred by the use of inland waterway data. <sup>245</sup>
	Public sector innovation		The standardised and systematic data collection and sharing by RIS authorities has pushed them to adopt new systems and to introduce programmes to educate personnel. It can be assumed that, overall, RIS authorities are higher skilled thanks to inland waterway data-sharing initiatives. Public authorities also innovate by

<sup>241</sup> Ecorys (2011), *Monitor VoortVarend Besparen, Eindmeting*. Via: [https://www.cbrb.nl/nieuws/documenten/doc\\_download/268-rapport-eindmonitor-voortvarend-besparen](https://www.cbrb.nl/nieuws/documenten/doc_download/268-rapport-eindmonitor-voortvarend-besparen), p. 35.

<sup>242</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. xv.

<sup>243</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 56.

<sup>244</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, Technical Annex 6: Ex-post Social Cost-Benefit Analysis Section 6.2.2.6.

<sup>245</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 80.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references:</b>
			developing necessary regulatory frameworks and through specific digitalisation initiatives. <sup>246</sup>
Public services and public administration	Public sector revenue		It can be assumed that more efficient use of fuel and the overall waterway infrastructure, public sector revenue has decreased. In the sector, fuel taxes, water pollution charges, port charges, fairway dues, and dues for locks and bridges are collected by authorities. <sup>247</sup> Since the use of RIS renders the sector more efficient, it can be expected that public sector revenue decreases. On the other hand, more efficient use of the fairway can lead to cost-savings on the side of RIS authorities, which increases savings.
	Public services management		The majority of costs to make RIS data available are carried by authorities, while most benefits are reaped by RIS users. The data is shared free of charge. While there are no quantifiable data available, the investments made by authorities related to RIS implementation are estimated to be at least EUR 200 million. <sup>248</sup> Furthermore, RIS authorities incur annual costs related to the updating of electronic navigational charts, which also implies that staff needs to be trained to create the charts, which costs time and money. However, authorities stipulate that they do not consider the costs high compared to the benefits they bring to the overall sector. Lastly, authorities estimate the preparation and implementation of the RIS systems to have involved anywhere between 0.2 and 8 full-time public staff (4.5 on average). <sup>249</sup>
	Public services performance		The sharing of IWT data has enhanced safety on the fairway, and along with it the efficiency of calamity abatement processes. As information is kept up to date and widely shared, rescue agencies immediately know what to do and how to react. In the case of dangerous cargo being transported, this can help avoid excessive pollution, but can also help save the lives of crew on board. <sup>250</sup>
Re-use	Demand for information		The demand for information on inland waterways data is naturally high, as all users of the fairway must have systems in place on their vessels to receive and share data. Member States report that the coverage of the shared information is nearly 100%. <sup>251</sup>
	Trust and confidence in		There are variations in the consistency and precision of some information between

<sup>246</sup> European Commission. (2017). *Digital Inland Waterway Area: Towards a Digital Inland Waterway Area and Digital Multimodal Nodes*. Via: <https://ec.europa.eu/transport/sites/transport/files/studies/2017-10-dina.pdf>, p. 5.

<sup>247</sup> European Commission. (2019). *Transport taxes and charges in Europe An overview study of economic internalisation measures applied in Europe*. Via: <https://ec.europa.eu/transport/sites/transport/files/studies/transport-taxes-and-charges-in-europe-isbn-978-92-79-99561-3.pdf>, p. 98.

<sup>248</sup> European Commission. (2014). *Evaluation of RIS Implementation for the period 2006-2011*. Via: <https://ec.europa.eu/transport/sites/transport/files/modes/inland/studies/doc/2014-07-evaluation-of-ris-implementation-main-report.pdf>, p. 155.

<sup>249</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, pp. 57-58.

<sup>250</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 2.

<sup>251</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 15.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references:</b>
	information		countries, as some data are not up-to-date and the RIS Directive does not stipulate the pace of revision and provision. The data is also self-reported, and may not be completely reliable. However, confidence on the shared data is present on behalf of users and the availability thereof is seen as useful and beneficial. <sup>252</sup>
Social	Mobility efficiency		Although limited, there is evidence of a reduction in congestion, as the increase in harmonisation and data exchange leads to time and cost savings for users and authorities alike. The provision of data used for navigation and voyage planning, and optimised use of the infrastructure allows for better resource planning, which saves costs. <sup>253</sup>
	Mobility systems planning		There is evidence of a reduction in congestion, though limited, as the increase in harmonisation and data exchange leads to time and cost savings for both users and authorities. However, there is no indication of a modal shift towards inland navigation occurring as a result of the use of RIS data. <sup>254</sup> As discussed above, other factors probably play a more important role in shifting towards multimodality.

<sup>252</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 16.

<sup>253</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, pp. 14, 56.

<sup>254</sup> European Commission. (2020). *Study supporting the evaluation of Directive 2005/44/EC on Harmonised River Information Services (RIS)*, p. 51.

### Transport networks data

The main benefits related to transport networks data (-models) are related to the optimal and safe usage of transport networks across modes, thereby contributing to the three principal objectives of the EU's Trans-European Transport Networks (TEN-T) policy-making: **fostering growth, creating jobs, and mitigating climate change.**<sup>255</sup>

Indeed, a well-functioning transport infrastructure network is essential to maintaining the European Union's competitiveness and wealth. For one, it is estimated that 3.5 million people travel across an internal EU border every day.<sup>256</sup> The sum of 27 Member States currently has around five million km of paved roads, of which 50 500 km are E-roads. On those roads, in 2018, 13 500 000 thousand tonnes or 1 765 123 million tonne-kilometres of freight were transported in the EU27.<sup>257</sup> The European rail network amounts to more than 215 000 km on which 9.6 billion passengers travelled and around 416 billion tonne-kilometres of freight were transported in 2017.<sup>258</sup> The sum of navigable canals and rivers adds up to 37 000 km. On these inland waterways, almost 550 000 thousand tonnes of cargo was carried in 2018 (EU27).<sup>259</sup> When it comes to air, in 2019, over 1 billion passengers were recorded and 13 700 000 tonnes of freight and mail were transported.<sup>260</sup>

The TEN-T policy's overall goal is to develop primary roads, bridges, inland waterways, airports, seaports, inland ports and traffic control networks in an organised fashion, by offering interconnected and multimodal high-speed long-distance routes while bearing sustainability as a key driver as well.<sup>261</sup> As such, sustainable transportation can enhance economic growth and improve accessibility while respecting the environment and improving the resilience of cities, urban-rural linkages, and productivity of rural areas.<sup>262</sup> In fact, it is estimated that the finalised TEN-T project, resulting in an optimised transport network, could reduce CO2 emissions from freight transport by almost 12.5 Mt CO2 by 2030 (-1.4%). Economically, the total EU GDP could grow by 1.6 per cent by 2030, generating approximately 797 000 jobs.<sup>263</sup>

In this context, digital representations of transportation networks can be used for a variety of purposes, ranging from **infrastructure governance to freight planning and fleet management**, etc. In the next section, applications for different transport modes will be discussed. Nevertheless, broadly speaking, five main use cases of these network data models can be distinguished, as presented below. It will become clear that transport network data increases its value for reuse when combined with traffic/vehicle data, which in turn, are covered by the ITS Directive.

- **Topology:** As discussed above, the purpose of a network data model is to provide an accurate representation of a network as a set or arrangement of links and nodes. For topology, transport networks data is leveraged in order to **represent reality as accurately as possible**, by encoding

<sup>255</sup> European Commission. (2018). *The impact of TEN-T completion on growth, jobs and the environment*. Via: <https://ec.europa.eu/transport/sites/transport/files/studies/ten-t-growth-and-jobs-synthesis.pdf>, p. 9.

<sup>256</sup> European Parliament. (2018). *Briefing: A Europe without internal borders? Free movement of persons*. Via: [https://www.europarl.europa.eu/ReqData/etudes/BRIE/2018/621815/EPRS\\_BRI\(2018\)621815\\_EN.pdf](https://www.europarl.europa.eu/ReqData/etudes/BRIE/2018/621815/EPRS_BRI(2018)621815_EN.pdf)

<sup>257</sup> Eurostat. (2019). *Transport data – Road Transport* [Dataset]. Via: <https://ec.europa.eu/eurostat/web/transport/data/database>

<sup>258</sup> European Commission - DG MOVE. (2019). *EU transport in figures: Statistical pocketbook 2019*. Via: <https://op.europa.eu/en/publication-detail/-/publication/f0f3e1b7-ee2b-11e9-a32c-01aa75ed71a1>, p. 78

<sup>259</sup> Eurostat. (2019). *Transport data – Inland Waterways Transport* [Dataset]. Via: <https://ec.europa.eu/eurostat/web/transport/data/database>

<sup>260</sup> Eurostat. (2019). *Transport data – Air Transport* [Dataset]. Via: <https://ec.europa.eu/eurostat/web/transport/data/database>

<sup>261</sup> European Commission. (2018). *Support Study for an impact assessment on measures for the streamlining of TEN-T*. Via: <https://ec.europa.eu/transport/sites/transport/files/studies/2018-09-19-support-study-ia-measures-streamlining-ten-t.pdf>, p. 8

<sup>262</sup> European Commission - DG MOVE. (2019). *EU transport in figures: Statistical pocketbook 2019*, p. 66.

<sup>263</sup> European Commission. (2018). *The impact of TEN-T completion on growth, jobs and the environment*, pp. 19, 28-29.

each node with the connectivity it permits, and providing accurate information on, for example, direction, route ends, etc. Additionally, transport networks data allow to increase the positional accuracy, as they make it possible to locate each node at the exact coordinate with matches the intersection it represents in real life.<sup>264</sup>

- **Cartography:** Certain elements, such as road type can be encoded or assigned in a transport data model. This, in turn, allows for each segment or object to be displayed on a map with its relative importance, such as highways, main street, street. Similarly, cartographic representations of networks leverage transport network data to be enriched with descriptive labels signifying important element and directional signs, or additional layers of information, such as landmarks. Moreover, nodal attributes can display traffic lights or other related information. These contribute to the cartographic utility and, subsequently, to the purpose of navigation and orientation of users. Cartographic representations of transport network data models are commonly used as **road or touristic maps** for the general public.<sup>265</sup>
- **Geocoding:** Assuming that a linear referencing system is embedded into a transportation network model, geocoding can be possible to derive a precise location. The most commonly used linear reference system is the address system, in which each link has a corresponding street name and address<sup>266</sup>. Once this information is included, it is possible to use transport network data for geocoding. In other words, transport network data enables the conversion of street addresses into geographical coordinates for accuracy in latitude and longitude. Hence, the network can be used to establish a location or address with relative accuracy.<sup>267</sup> Translating customers' addresses to location coordinates is done by a variety of businesses that need to **navigate exact locations, ranging from delivery services, ride hailing, to supply-chain companies**. Even insurers leverage transport network data to **identify zones that are prone to natural hazards** for the purpose of determining claims. Finally, marketers use geocoding to explore areas allowing them to define new market opportunities.<sup>268</sup>
- **Routing and assignment:** transport network data models can be leveraged to identify optimal paths and routes for commuting from A to B, in line with given capacity constraints. To do so, a topology is needed in which the relationship of each link with other intersecting segments must explicitly be specified. Many re-users, ranging from logistics companies, parcel delivery businesses, to passenger transportation enterprises leverage this data to calculate, and eventually minimise, their marginal costs incurred due to **bottlenecks in traffic infrastructure**. Lastly, traffic regulators and mobility service providers can develop alternative **multimodal transportation systems** through analysis of transport network data and determine optimal traffic flows. These kinds of innovative mobility schemes can be improved by supplementing transport network data with real-time traffic data.<sup>269</sup>
- **Infrastructure planning and construction:** As the demand for infrastructure increases drastically, infrastructure planning and construction can be improved through the use of transport network data

---

<sup>264</sup> The Geography of Transport Systems. *Topology of a Network Data Model*. Via: [https://transportgeography.org/?page\\_id=7594](https://transportgeography.org/?page_id=7594)

<sup>265</sup> The Geography of Transport Systems. *Cartography of a Network Data Model*. Via: [https://transportgeography.org/?page\\_id=7600](https://transportgeography.org/?page_id=7600)

<sup>266</sup> It should be noted that the INSPIRE Data Specifications for Transport Networks underlined this key relationship with the data theme Addresses.

<sup>267</sup> The Geography of Transport Systems. *Geocoding in a Network Data Model*. Via: [https://transportgeography.org/?page\\_id=7604](https://transportgeography.org/?page_id=7604)

<sup>268</sup> Pitney Bowes. *Geocoding and Big Data?* Via: <https://www.pitneybowes.com/us/location-intelligence/case-studies/when-is-geocoding-a-big-data-problem.html>

<sup>269</sup> Liu, J., Mirchandani, P., & Zhou, X. S. (2020). *Integrated vehicle assignment and routing for system-optimal shared mobility planning with endogenous road congestion*, pp. 4-5.



models. In combination with traffic data, it allows policymakers to identify where congestion occurs across modes and locations, arising from exceeding capacity and where random, but predictable events take place (i.e. accidents, natural hazards, and so forth). Upon identification of points that pose problems, practical solutions can be developed, such as **increasing/decreasing the capacity of sections or improving safety**. This not only concerns road transport networks but other modes and transport hubs as well. Additionally, reliable information on critical points in transport networks allows efficient planning of construction as well. For one, works can be phased or take place at night based on estimations of flows of traffic (linked to the network specificities, but also actual traffic data) to minimise the impact of construction on these sections.<sup>270</sup>

These various use cases for transport networks data feed into series of more specific applications and usages for the various modes of transport, as presented hereafter.

**Efficient road transport** – parcel deliveries, cargo enterprises, mobility service providers, and so forth – relies on accurate information of the networks’ characteristics and their conditions for optimal use, voyage planning, last-mile delivery of goods (i.e. poses greatest risk in terms of unexpected charges), and so forth. For example, regulations and circulation plans inform freight managers and authorities to avoid inapt use of sensitive parts of transport networks. Additionally, transport network data about dangerous sections, road works, and other infrastructural bottlenecks allow efficient routing and helps to cope with the negative impacts of emissions and noise pollution. As such, reusing transport network data and developing **accurate digital maps that describe the road network’s geometry, topology, and infrastructural-related map attributes**, such as traffic regulations, circulation plans, restrictions for heavy freight vehicles, and so forth, are crucial for freight operations. Integration of these maps with freight management systems represent a multitude of benefits for operators – not only economically, but environmentally as well. Furthermore, this information can also be used by network managers to optimise the use of the infrastructure by **strengthening multimodal links**, which attributes to ecological objectives.<sup>271</sup> With the same objective in mind, transport network data can be used by energy companies to strategically plan (electric) refuelling stations along routes to enable the shift towards renewable energy use. Evidently, navigational systems providers use transport network data as a basis for their maps, upon which some augment their services with other data sources. Finally, in times of crisis (e.g. current COVID-19 crisis), transport authorities can leverage data on their transport network to ensure that deliveries of medical equipment, food and other vital goods are sustained.<sup>272</sup>

Some EU funded projects are relevant in this area, e.g. the EULF Transportation Pilot carried out in Norway and Sweden. The aim is to improve the flow of accurate road safety data – including some elements included in transport networks – between public authorities and private map developers. This, in turn, increases predictability of the infrastructure for road users and enhances overall safety as well.<sup>273</sup> Another EU project leveraging transport networks data is the EuroRoadS that demonstrates on-trip information and warning services on current speed limits on the road.<sup>274</sup>

---

<sup>270</sup> The Geography of Transport Systems. *Improving Transport Infrastructure*. Via: [https://transportgeography.org/?page\\_id=6436](https://transportgeography.org/?page_id=6436)

<sup>271</sup> European Commission - DG MOVE. *Availability and access to road data*. Via: [https://ec.europa.eu/transport/themes/its/road/action\\_plan/availability\\_and\\_access\\_to\\_road\\_data\\_en](https://ec.europa.eu/transport/themes/its/road/action_plan/availability_and_access_to_road_data_en)

<sup>272</sup> Conference of European Directors of Road (CEDR). *Information on road transport during the COVID19 emergency*. Via: <https://www.cedr.eu/17594/information-on-road-transport-during-the-covid19-emergency/>

<sup>273</sup> European Commission. (2015). *EULF Transportation Pilot: Road data exchange in Norway and Sweden*. Via: <https://joinup.ec.europa.eu/sites/default/files/inline-files/EULF%20Factsheet%20Transportation%20Pilot%20v1%20final.pdf>

<sup>274</sup> European Commission - Joint Research Centre. *EuroRoadS*. Via: <https://inspire.ec.europa.eu/glossary/EuroRoadS>



In **railway transport**, network data is used to improve rail operations and performance of the rail system. For one, known obstructions in rail infrastructure can be identified through the use of transport network data, allowing network managers to upgrade the network. Leveraged by network data, they can **increase capacity and safety** by applying digital traffic management systems or AI to sections prone to congestion, disruption or calamities. Naturally, a railway network which is used more **efficiently also has environmental benefits**.<sup>275</sup> Another example is that of Infrabel, the public undertaking responsible for railway infrastructure in Belgium, who publishes transport network data to give users an overview of the geographical position of track crossings. The benefits are rather on the social side, as the purpose is to allow reuse of those datasets, for non-commercial purposes notably by citizens in order to prevent any potential dangers.<sup>276</sup> Similarly, across Europe numerous public railway companies collect and publish data about their networks to **improve services, achieve greater transparency and communicate better with citizens**.<sup>277</sup> Project administrators and infrastructure operators use reliable maintenance data to track the safety of the railway network. Commercial operators may use network data to improve marshalling and distribution processes, automate payment and invoicing mechanisms, improve link management with other modes of transport, and so on.<sup>278</sup>

When it comes to **air transport**, the main benefit of transport network data lies in **predicting future of airline routes**. For both private operators as public authorities, predicting passenger as well as freight flow capacity of existing routes can be considered as crucial. Based on this information, commercial operators can optimise their processes, which can result in better and efficient resource planning, such as fuel use. This can also entail environmental benefits resulting from increased efficiency. Additionally, identifying vulnerable points in the network can increase the overall security of air transport networks. Lastly, predicting future traffic volume and expected growth or decline of transport hubs also has an **impact on the design of future aircrafts**. When analysis of a transport network model predicts point-to-point transportation between smaller hubs as the preferred option, airline carriers will prefer small to medium-size aircrafts to optimise their cost structure. On the other hand, airline carriers preferring to operate through large hubs that can efficiently handle extensive volumes of take-off and landing and logistics will opt for high-capacity jumbo-jets.<sup>279</sup>

Benefits of transport network data for **inland waterways** is mainly related to more efficient use of the waterway infrastructure. As discussed in depth in section 0, freight and cargo operators benefit greatly from accurate voyage planning, as it allows them to plan their resources efficiently. For example, information on the waterway infrastructure enables users to calculate how much cargo they can move from one point to the other. Transport network data supplemented with real-time fairway data sharing can enable the automation of waterway infrastructure. Moreover, automated lock systems can be developed; a multitude of sensors and aquatic drones can be used to monitor infrastructure and operations; the availability and operational efficiency of cargo handling machines (i.e. cranes, trucks,...) and multimodal connections can be predicted through the use of transport network data.<sup>280</sup> This has **economic**

---

<sup>275</sup> Network Rail Limited. *Targeting digital systems to improve capacity and performance on the railway*. Via: <https://www.networkrail.co.uk/running-the-railway/railway-upgrade-plan/digital-railway/benefits-for-britain/>; see also German Federal Ministry of Transport and Digital Infrastructure, via: <https://www.bmvi.de/SharedDocs/DE/Artikel/DG/mfund-projekte/zekiss.html>.

<sup>276</sup> Infrabel. *Liste et position géographique des passages à niveau*. Via: [https://opendata.infrabel.be/explore/dataset/geo/pn/information/?disjunctive.fld\\_actief\\_passief&disjunctive.fld\\_postcode\\_en\\_gemeente&location=7,5.50.53185,4.4.31655&basemap=jawg.streets](https://opendata.infrabel.be/explore/dataset/geo/pn/information/?disjunctive.fld_actief_passief&disjunctive.fld_postcode_en_gemeente&location=7,5.50.53185,4.4.31655&basemap=jawg.streets)

<sup>277</sup> See: SNCF (France): <https://ressources.data.sncf.com/explore/?sort=modified>; SBB (Swiss) : <https://data.sbb.ch/pages/home20/>; Deutsche Bahn (Germany): <https://data.deutschebahn.com/>; Nederlandse Spoorwegen: <https://apiportal.ns.nl/>.

<sup>278</sup> European Commission – DG MOVE. Rail. Via: [https://ec.europa.eu/transport/themes/its/rail\\_en](https://ec.europa.eu/transport/themes/its/rail_en)

<sup>279</sup> Guo, W., Toader, B., Feier, R., Mosquera, G., Ying, F., Oh, S.-W., ... Krupp, A. (2019). Global air transport complex network: multi-scale analysis. *SN Applied Sciences*, 1(7), pp. 10-11.

<sup>280</sup> Inland Navigation Europe. *EU Digital Policy*. Via: <http://www.inlandnavigation.eu/what-we-do/eu-digital-policy/>

**advantages and results in better environmental performance of the sector.** Lastly, **increased safety** is also a benefit stemming from the use of transport network data.

Finally, transport network data on “the world’s safest means of transport”<sup>281</sup>, or **cableways**, include height of valley station, height of mountain station, route distance, travel time, driving speed line, opening hours, and so forth.<sup>282</sup> This data can feed into **apps for holidaymakers** in mountainous environments to calculate travel time from point A to another. Additionally, waiting times at lifts and openings thereof could also be integrated into such apps. Data collected can also be used to develop interactive world roadway maps that allow real-time tracking of cableways.<sup>283</sup> Furthermore, cableway transport maps can be used to **identify connectivity gaps to develop remote mountainous zones and to connect them to less-remote areas**. Also in urban areas transport network data on cableway can be useful; it can be used for city planners to **fill in gaps and complement public transport systems** in urban areas. It can be especially useful when bridging difficult topographical terrain. Examples include Medellin (CO), London (UK), Luxembourg City (LU), Caracas (VE), New York (US), among others.<sup>284</sup> Finally, cablecars represent a relatively cheap, noise-free and environmentally friendly method of transportation.<sup>285</sup>

---

<sup>281</sup> International Organization for Transportation by Rope (OITAF). *What is OITAF?* Via: [http://www.oitaf.org/index\\_e.htm](http://www.oitaf.org/index_e.htm)

<sup>282</sup> Lift-World. *Lifts in the world - Lift-Database*. Via: [https://www.lift-world.info/en/lifts/place/alpe\\_dhuez/index.htm](https://www.lift-world.info/en/lifts/place/alpe_dhuez/index.htm)

<sup>283</sup> Doppelmayr GmbH. *Interactive Ropeway Map*. Via: <https://www.doppelmayr.com/doppelmayr-interactive/references/>

<sup>284</sup> BBC. *The rise of the urban cable car*. Via: <https://www.bbc.com/future/gallery/20190103-the-rise-of-the-urban-cable-car>

<sup>285</sup> World Bank. *Innovation in the air: using cable cars for urban transport*. Via: <https://blogs.worldbank.org/transport/innovation-air-using-cable-cars-urban-transport>

Table 19 – Assessment of value for transport networks data

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
Economic	Competition		There is evidence from 'real-life uses cases' presented above, that the reuse of transport network data improves firms' performance through the creation of new value chains, enabled by the development of new services.
	Consumer benefits		Since societies shift to a knowledge economy based in cities, transport network data in urban areas enables consumers to make better informed decisions by providing them reliable information on transit infrastructure. <sup>286</sup> Products and services using open data enable more transparency, potentially rendering them more attractive to consumers.
	Economic output		Transport network data helps decision makers to identify weaknesses in the network they manage in a secure and timely manner. While the precise economic performance cannot be measured, effective transport infrastructure provides economic and social benefits by increasing consumer efficiency and competitiveness, maintaining sustainable regional economic growth, generating jobs, fostering labour mobility and linking societies. <sup>287</sup> Congestion losses in Europe equate to around 1 percent of GDP; alternate routes based on transport network data improve mobility on different roads and eliminate traffic jam distances. Mitigating these losses will boost economic performance. <sup>288</sup> Finally, transport network data can be used in a range of applications, such as on-demand services, ride-hailing, etc. Subsequently, businesses using data from the transport network will increase their economic performance substantially. All of the above fuels economic output.
	Employment		As described above, an effective transport network leveraged by the data thereon increases market efficiency and competitiveness, while maintaining sustainable regional economic growth, generating jobs, fostering labour mobility and linking societies. <sup>289</sup> Furthermore, businesses that build new supply chains by leveraging data from the transport network generate new, highly qualified jobs.
	Product market dynamism		Similar to "economic output" section above; increased availability of transport network data will improve the inputs for the products and services. The more transport network data is used and reused, the more efficient the network will become. This means that availability of transport network data optimises prices, improves the services and along with it

<sup>286</sup> Deloitte. (2019). *Toward a mobility operating system: Establishing a lingua franca for urban transportation*. Via: <https://www.deloitte.com/us/en/insights/focus/future-of-mobility/urban-transport-mobility-platforms.html>

<sup>287</sup> Organisation for Economic Co-operation and Development (OECD). *Transport - Infrastructure investment*. Via: <https://data.oecd.org/transport/infrastructure-investment.htm>

<sup>288</sup> European Data Portal. (2017). *Re-using Open Data: A Study on transforming Open Data into economic & societal value*. Via: [https://www.europeandataportal.eu/sites/default/files/re-using\\_open\\_data.pdf](https://www.europeandataportal.eu/sites/default/files/re-using_open_data.pdf) p. 21.

<sup>289</sup> Organisation for Economic Co-operation and Development (OECD). *Transport - Infrastructure investment*. Via: <https://data.oecd.org/transport/infrastructure-investment.htm>

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
			attractiveness and diversity of the value chain. Indeed, these services can increase the positive user experience and the dynamism of the product market. <sup>290</sup>
	Productivity and commercialisation		Data-driven decision making stemming from the identification of bottlenecks in the transportation network to enable optimised route planning, detection of networks with poor infrastructure, and calculation of travel delay, naturally drives productivity. <sup>291</sup>
	Public-private coordination		There is an important relationship between the public and private sector with regard to transport network data. The public sector is responsible for the network and for the collection of data thereon. As such, the private sector relies heavily on the public sector to produce high-quality and high-quantity data. On the other hand, the private sector uses the data on transport networks to create products and value. These applications can increase the efficiency of that network, also benefitting the public sector. This win-win situation creates a potentially long-lasting interdependency. <sup>292</sup>
Environmental	Citizen engagement in addressing climate change		There are several examples of citizen-led sustainability projects that use transport network data to combat congestion and air pollution. In fact, the data provides a detailed description of the evolution of air quality, and allows to better understand and address it more effectively. To citizens and decision-makers alike looking to improve the air quality, this is of tremendous importance. <sup>293</sup>
	Energy management and efficiency		The use of data from transport networks helps various types of mobility systems to properly combine different networks into a more effective operation. This provides incentives for a more sustainable transport infrastructure as it uses the existing resources of the different transport networks more effectively and reduces the burden on public space. Data on transport networks are also used in logistics and freight operations to reduce their environmental impact by reducing the kilometres driven and congestion. <sup>294</sup>
	Environment management		Similarly as above, transport network data can be used to reduce congestion and render the network more efficient, air quality will improve and CO2 emissions will lower. New services might also contribute to modal shifts away from cars to more sustainable options such as bicycle or e-scooter sharing.
Innovation & AI	Citizen innovation		Freely available transport network data can improve citizen self-serving by allowing citizens to directly research the status and general information about transport data networks they use. There are examples of citizen-oriented projects aimed at improving local transport

<sup>290</sup> European Environment Agency (EEA). (2019). *Transport and environment report 2019: The first and last mile — the key to sustainable urban transport*. Via: <https://www.eea.europa.eu/publications/the-first-and-last-mile>, p. 68.

<sup>291</sup> McKinsey Global Institute. (2016). *The age of analytics: Competing in a data-driven world*. Via: <https://cutt.ly/TyDjIXI>, pp. 57-59.

<sup>292</sup> European Commission. (2019). *How public-private partnerships can transform urban mobility* | *Intelligent Cities Challenge*. Via: <https://www.intelligentcitieschallenge.eu/how-public-private-partnerships-can-transform-urban-mobility>

<sup>293</sup> Transport&Environment. *How citizen science is helping combat air pollution in Brussels*. Via: <https://www.transportenvironment.org/news/how-citizen-science-helping-combat-air-pollution-brussels>

<sup>294</sup> European Environment Agency (EEA). (2019). *Transport and environment report 2019: The first and last mile — the key to sustainable urban transport*, pp. 47, 56.

Macro characteristics	Value category	Assessment of value for the datasets in scope	Examples/references
			networks. As discussed above, citizens can be empowered to use data in order to take action and innovate to tackle air pollution or congestion, for example. <sup>295</sup>
	Entrepreneurialism and private sector innovation		The sector of re-users of transport data network has been growing steadily. Entrepreneurialism and innovation are central to this evolution, as these companies mostly see the light as small high-tech start-ups that grow exponentially. For one, the ride-hailing market value could very well reach \$220 billion by 2025. <sup>296</sup>
	Public sector innovation		The public sector benefits significantly from transport network data. The public sector uses the data for maintaining and improving the network, as well as creating new routes. City planning can also be done in innovative ways through the use of this data. Available data can be aggregated to conceive an innovative transportation strategy. Moreover, since 2016, the yearly INSPIRE Conference has hosted an INSPIRE Hackathon to share knowledge and demonstrate to a larger audience the benefits of transport network data. <sup>297</sup>
Public services and public administration	Public services management		As mentioned above, data on the transport network enables the public authorities to manage the transport network and the related services they provide effectively. This ranges from adequate planning of construction and maintenance to improving urban public transport. Generally speaking, the proper utilisation of public infrastructure requires reliable, efficient and user-friendly transport network data. <sup>298</sup>
	Public services performance		Data on transport networks can have a positive impact on public health. As transport affects the health of populations directly and indirectly through emissions, but also road traffic accidents. The predictive nature of data analytics on transport network enables public authorities to mitigate the negative effects related to transport. <sup>299</sup>
Re-use	Demand for information		Preliminary research conducted by the European Data Portal on the use of publicly available data has indicated that about 18.5% of interviewed businesses use transport network data or in combination with other dataset on regions and cities for example. <sup>300</sup> Considering the ever increasing market share of transport network data companies as mentioned above, it can be assumed that the demand is high for such data.
	Trust and confidence in information		Considering that transport network data is official information from public administrations, reliability of companies using this data increases. It also creates trust and reliability on behalf of end-users (i.e. customers of these companies). <sup>301</sup>

<sup>295</sup> POLIS Network. *Five European cities pioneer new citizen-oriented project aimed at improving local mobility with data*. Via: <https://www.polisnetwork.eu/news/five-european-cities-pioneer-new-citizen-oriented-project-aimed-at-improving-local-mobility-with-data/>

<sup>296</sup> 360iResearch. *Global Ride Sharing Market*. Via: [https://www.reportlinker.com/p05871396/Global-Ride-Sharing-Market-Premium-Insight-Competitive-News-Feed-Analysis-Company-Usability-Profiles-Market-Sizing-Forecasts-to.html?utm\\_source=GNW](https://www.reportlinker.com/p05871396/Global-Ride-Sharing-Market-Premium-Insight-Competitive-News-Feed-Analysis-Company-Usability-Profiles-Market-Sizing-Forecasts-to.html?utm_source=GNW)

<sup>297</sup> European Data Portal. *Dubrovnik INSPIRE Hackathon 2020*. Via: <https://eudatasharing.eu/events/dubrovnik-inspire-hackathon-2020>

<sup>298</sup> European Environment Agency (EEA). *INSPIRE Directive*. Via: <https://www.eea.europa.eu/about-us/what/seis-initiatives/inspire-directive>

<sup>299</sup> World Health Organization. *Data and statistics*. Via: <http://www.euro.who.int/en/health-topics/environment-and-health/Transport-and-health/data-and-statistics>

<sup>300</sup> European Data Portal. (2017). *Re-using Open Data: A Study on transforming Open Data into economic & societal value*, p. 28.

<sup>301</sup> European Data Portal. (2017). *Re-using Open Data: A Study on transforming Open Data into economic & societal value*, p. 37.

<b>Macro characteristics</b>	<b>Value category</b>	<b>Assessment of value for the datasets in scope</b>	<b>Examples/references</b>
	Volume and range of information		While little figures exist on the download volumes of transport network data, stakeholders have indicated that they would like to see the volume of available data increase. Possibly, it can be assumed that the download volume is high. Transport network data is regularly used in combination with other public sector information, such as information on cities, populations, and so forth. <sup>302</sup>
Social	Disease prediction and prevention		Since transport is one of the main sources of air pollution, it has negative effects on mortality and is linked to respiratory and cardiovascular diseases. In Europe, each year about 100 000 citizens die prematurely due to air pollution. <sup>303</sup> The socioeconomic cost of road traffic injuries is estimated to be about 2% of European countries' GDP. Transport network data can be an important enabler of new ways of reducing negative externalities of transport in terms of reducing emissions and making transport infrastructure safer, among others.
	Mobility access		As improved mobility access is at the core of the benefits related to transport network data, it is straightforward that it has the potential to increase the number of public transport users. Additionally, applications employ transport data network to develop healthier and more environmental alternatives to individual car use. <sup>304</sup>
	Mobility efficiency		Transport network data can wave the whole transportation landscape into a connected network, which saves time and resources due to more efficient use of the infrastructure and reduced congestion. A plethora of applications are already in place with the specific objective of increasing the efficiency of mobility, enabled by transport network data. <sup>305</sup>
	Mobility planning		As discussed above, the more efficient use of transport networks through transport network data is one of the most important use cases. The data used improves voyage planning of individuals or freight operations alike. As trips are executed more efficiently, the ease and satisfaction of users increases as well. Stakeholder consultations have also pointed out that transport data is leveraged for designing new and upgrading existing infrastructure to meet citizens' needs and enhancing safety.
	Mobility systems planning		Transport network data is crucial for authorities when it comes to identifying gaps or bottlenecks in their systems, as well as coming up with solutions, such as better integration of parallel systems.

<sup>302</sup> European Data Portal. (2017). *Re-using Open Data: A Study on transforming Open Data into economic & societal value*, p. 26, 73.

<sup>303</sup> World Health Organization. *Air pollution and climate change*. Via: <http://www.euro.who.int/en/health-topics/environment-and-health/Transport-and-health/data-and-statistics/air-pollution-and-climate-change2>; World Health Organization. *Economic cost of transport-related health effects*. Via: <http://www.euro.who.int/en/health-topics/environment-and-health/Transport-and-health/data-and-statistics/economic-cost-of-transport-related-health-effects2>

<sup>304</sup> European Commission. (2017). *Study on urban mobility – Assessing and improving the accessibility of urban areas: Final report and policy proposals*. Via: [http://docs.confibus.org/CE\\_MovilidadUrbana\\_MI-04-16-271-EN-N.pdf](http://docs.confibus.org/CE_MovilidadUrbana_MI-04-16-271-EN-N.pdf) , pp. 6-7.

<sup>305</sup> Opendatasoft. *The Importance of Sharing Mobility Data*. Via: <https://www.opendatasoft.com/blog/2018/09/24/mobility-data-sharing>

## 2.8 Horizontal considerations from the Open Public Consultation

The European Commission conducted an open written public consultation, with the support of an online survey, from 20 February 2020 until 3 June 2020. The goal of the consultation was to obtain information from all stakeholders with an interest in the EU data economy, including governments, companies and business organisations, public sector content holders and users, experts and academics as well as citizens, about the European Data Strategy. The online general questionnaire was available in all EU languages on the public consultation website of the European Commission and contributions could be made in any of the 24 official EU languages.

The questionnaire itself is structured in two sections: the first part (Section 1 of the consultation) has the objective to collect views on the data strategy in general, while the second part (Section 2) covers more specific topics, which are divided into sub-sections. This second section aims to collect information on three specific aspects announced in the data strategy, namely:

- How data governance mechanisms and structures can best maximise the social and economic benefits of data usage in the EU;
- The EU-wide list of high-value datasets that the Commission is to draw up under the recently adopted Open Data Directive; and
- The role of self-regulation to implement rules on data processing.

Given that the public consultation has covered the overall European Data Strategy, which has wider scope than the current assignment, the analysis presented in this report only focuses on answers received for the questions in Section 2.2 - *Specific questions on future actions: identification of high-value datasets* to remain in the scope of the study. The section contains a mixture of open and closed questions and covers the following aspects:

- Most important factors to be considered when selecting datasets for the future list of high-value;
- Type of arrangements relevant to improve the re-usability of specific high-value datasets;
- Type of activities to support from EU Funding to enhance the availability and re-use of high-value across Europe.

Finally, it should be noted that as per the timing of the Open Public Consultation, the feedback from respondents was available to the study team when the overall scoping and selection of HVDs was already well advanced. However, no significant contradiction of the existing findings was identified. On the contrary, the feedback collected through the online survey made it possible to validate:

- The overall need for a policy intervention such as the PSI HVD regime enhancing the availability and reuse of PSI;
- The assumptions made as part of the methodological framework for the identification of HVDs presented in section 2.1 above regarding the most relevant factors for the selection of HVDs;
- The potential HVDs (or parts thereof) presented in sections above and assessed as part of Chapter 3;
- The salience of additional publication arrangements, similar to those assessed and recommended as part of Chapter 3 to improve the availability and reuse of PSI.

The summary of the 761 replies to the online general questionnaire is presented hereafter and the graphs resulting from the replies to closed questions are displayed in Annex A.



### **2.8.1 Relevance of high-value datasets and their selection factors**

In general, the establishment of a list of high-value datasets is positively perceived among the respondents. The majority of the respondents (82%) strongly agreed or somewhat agreed with the statement: *"The establishment of a list of high-value datasets, to be made available free of charge, without restrictions and via APIs, is a good way to ensure that public sector data has a positive impact on the EU's economy and society."* Furthermore, more than half of the respondents strongly agreed that the above mentioned elements have positive impact on the EU's economy, and only approx. 7% of the respondents disagreed (strongly disagree or somewhat disagree) with this statement.

Respondents were asked about which factors would increase the re-use of datasets and how relevant are they. In general, all of the following four factors are perceived positively: more than half of the respondents think that these factors are very relevant, or relevant, to take into account when selecting datasets for the future list of high-value:

- The re-use of the dataset would increase if it was provided free of charge;
- The dataset belongs to a thematic area in which there are few EU-level requirements for opening up data;
- The re-use of the dataset would increase if its availability under uniform conditions was ensured across the entire EU;
- The re-use of the dataset would increase if it was available via an application programming interface (API).

The most relevant factor considered to select datasets was to *ensure uniform conditions for available data*. According to 84% of the respondents, this factor is either very relevant or relevant. 78.84% of the respondents found the factor of making datasets available through APIs very relevant or relevant, and 75% found the availability of datasets free of charge very relevant or relevant. The least appealing factor was *"the dataset belongs to a thematic area in which there are few EU-level requirements for opening up data"*, which was indicated as very relevant or relevant by 52%.

When it comes to additional relevant factors to select the future list of high-value, most of the respondents found data quality, data protection and creation of taxonomies as the most important factors. Most of the answers addressing the questions of data quality from different angles: up-to-date, consistent datasets with high quality, and proper granularity, are significant factors according to respondents. Liability and reliability, as well as known provenance (i.e. transparency about the data; how they are collected, what do they cover, limitations etc.) are also factors which need to be taken into consideration. It was also expressed by several respondents that datasets should be reliable, clearly documented and dataset structures should be based on generally accepted taxonomies with as much semantic information as possible. Another important factor, which was considered by several respondents is data protection, there is a need to fully respect intellectual property rights as well as GDPR and data anonymisation rules.

Beyond additional relevant factors, several respondent expressed their views about governance of datasets. Selecting high-value should be based on use cases with potential societal benefits and the issue of who can benefit from the opening of databases needs to be considered. Another important element can be capacity building on proper use of such datasets.

### **2.8.2 Relevance of additional arrangements to improve data re-usability**

Survey respondents were also consulted about the relevance of additional arrangements indicated below to improve the re-usability of high-value. Generally, all of the following four arrangements listed below are considered as relevant elements. More than half of the respondents think these arrangements are very



relevant or relevant to take into account to improve the re-usability of data when specifying the characteristics of such datasets:

- Standardised formats of data and metadata
- Licensing and other terms applicable to re-use
- Possibility of user feedback
- Specific technical arrangements for dissemination

The most relevant arrangement is to provide standardised formats of data and metadata to improve re-usability of datasets. 84% of the respondents found this element very relevant or relevant. 80% of the respondents found licensing and other terms applicable to re-use very relevant or relevant, and 67% found the possibility of user feedback very relevant or relevant arrangement. The least relevant arrangement was "*Specific technical arrangements for dissemination*", 53 % of respondents found it relevant or very relevant.

Respondents pointed out that additional arrangements, such as open standards, available data in different formats (EST API, GraphQL, file download etc.) would also improve re-usability. Respondents also emphasised the importance of licencing: according to several opinions it is key to avoid that global players monopolise the added value of these datasets – and standardised data license agreements can facilitate new collaborative approaches for sharing & reusing data resources. Clearly specified formats such as metadata descriptions, catalogues etc. would be also advisable to use according to several survey respondents. FAIR principles (data should be Findable, Accessible, Interoperable and Reusable) are also advised to take into consideration when creating standardised formats of data and metadata. Furthermore, user feedback may help identify anomalies and be a driver for standardisation. User friendly API and bulk download solutions, capacity building and proper guidance is also important to improve data reusability according to respondents.

Furthermore, several respondents also stressed that the previously mentioned data protection (e.g. options for data anonymisation and pseudonymisation, GDPR rules etc.) and data quality aspects (e.g. granularity, completeness, accuracy, and timeliness) are also very relevant factors when it comes to the improvement of data re-usability.

Regarding data dissemination solutions and their technical arrangements, most respondents share the views that there is a need for user-friendly tools to download datasets with standardised (or common) data exchange protocols and well-documented APIs.

The arrangements should include multiple access points and ensure 24/7 access for users via thematic portals or dedicated websites at the European level (e.g. Linked Open Data according to W3C standards). The dissemination solution could even re-use of INSPIRE principles and standards and ensure Machine to Machine access and the availability of high performance data analytics platform.

### **2.8.3 Relevant activities enhancing data re-use requiring EU funding**

Survey contributors were also asked about which the relevant activities (which may need EU funding to enhance the availability and re-use of high-value datasets) are. In general, all of the following three activities contribute to the availability of HVDs; more than 70% of the respondents think these activities are very relevant or relevant when enhancing the availability and re-use of HVDs:

- Improving the quality (e.g. machine-readability) and interoperability of the data/metadata;
- Ensuring sustainable data provision via application programming interfaces (APIs);
- Engaging with re-users (promoting the data, co-defining use cases).

The most relevant activity considered to enhance data re-use and data availability was to *improve the quality (e.g. machine-readability) and interoperability of the data/metadata*. 89% of the respondents found this activity very relevant or relevant. 81% of the respondents found the activity of *ensuring sustainable data provision via application programming interfaces (APIs)* very relevant or relevant, and 70% found the *engagement with re-users (promoting the data, co-defining use cases)* very relevant or relevant.

As additional activities, respondents highlighted that it is advisable to create pilot projects prototypes, pilots with stakeholders for testing, validation and self-assessment at the preliminary stage of the opening of databases. In terms of engagement, it is needed to set conditions and obligations for secondary data users and data aggregators and to engage with potential re-users and discovering their needs. This cooperation fostering amongst users and providers allows that availability and demand are aligned.

Furthermore, the involvement of a wide range of stakeholders is essential: supporting the participation of experts, in particular from SMEs and financing digital sites with individuals to facilitate access to data for people in digital poverty. It is also advised to design and deploy impact measuring system for data re-use and to facilitate measuring the business benefits.

Centralisation of data at European level is also key: creation of an Agency in charge of maintaining a global catalogue and being a well-known reference to contact. This would ensure cybersecurity, guaranteeing that data are not accidentally or deliberately manipulated.

#### 2.8.4 List of concrete datasets provided by respondents

The followings specific datasets were listed by respondents that should be listed in each of the thematic categories of high-value datasets:

Thematic area	List of datasets provided by respondents
<b>Geospatial</b>	<ul style="list-style-type: none"> <li>• Building information</li> <li>• EU land/coastlines</li> <li>• Address register</li> <li>• National Digital Cadastral Map (DKM)</li> <li>• Digital Landscape Model (DLM)</li> <li>• Agricultural data (soil consumption, soil fertility, land use, CAP incentives)</li> <li>• Blue economy activities</li> <li>• All the information referred to in the INSPIRE directive in Annex III</li> <li>• Civic numbers,</li> <li>• Cadastral maps</li> <li>• Marine datasets</li> <li>• ESRI Open Data Hub</li> <li>• Natural Earth Data</li> <li>• USGS Earth Explorer</li> <li>• GEOROC</li> <li>• Critical infrastructure (distribution of hospitals, data centres, electricity grid, water supply, airports, internet)</li> <li>• Parking areas, availability of free parking spaces</li> </ul>
<b>Earth observation and environment</b>	<ul style="list-style-type: none"> <li>• Data on biodiversity</li> <li>• Hydrographic data</li> <li>• Digital Elevation Model with high spatial resolution (e.g. 5 m)</li> <li>• Aerial orthophotographs</li> <li>• Satellite imagery</li> <li>• River discharge and water analysis</li> <li>• Environmental permit information</li> <li>• Real-time emissions usage of pesticides</li> <li>• Early warning incidents</li> </ul>

---

	<ul style="list-style-type: none"> <li>• Forests data,</li> <li>• Hydrology data,</li> <li>• Engineering geology data,</li> <li>• Geothermy data,</li> <li>• Air quality,</li> <li>• Fishing districts</li> <li>• Copernicus Marine Environmental Monitoring Service (CMEMS),</li> <li>• EMODnet,</li> <li>• SeaDataNet</li> <li>• Natural disasters risk data</li> <li>• Renewable energy prediction data</li> </ul>
<b>Meteorological</b>	<ul style="list-style-type: none"> <li>• Weather/pollution stations</li> <li>• All ESA data</li> <li>• Rainfall extremes of short duration.</li> <li>• Climate model data;</li> <li>• Historical climate data</li> <li>• Atmospheric pressure</li> <li>• ECMWF Weather Forecast</li> <li>• Data necessary to determine road conditions</li> <li>• Marine observations and predictions</li> </ul>
<b>Statistics</b>	<ul style="list-style-type: none"> <li>• Economic data (local, national, EU)</li> <li>• Health related data</li> <li>• Demographic data</li> <li>• Epidemiological data</li> <li>• Data on crime and law enforcement</li> </ul>
<b>Companies and company ownership</b>	<ul style="list-style-type: none"> <li>• Corporate registers</li> <li>• All merchant and market registers</li> <li>• Market position - Datasets by companies</li> <li>• insolvencies, company hierarchies, company financials</li> <li>• List of EU companies with sectors</li> <li>• Address data of companies</li> <li>• Annual financial statements</li> <li>• Employment</li> <li>• Non-financial information which is essential for strengthening the data base of sustainable finance</li> <li>• Life Cycle Assessment indicators</li> </ul>
<b>Mobility</b>	<ul style="list-style-type: none"> <li>• Traffic lights status and timings</li> <li>• Public transport data</li> <li>• accurate timetables</li> <li>• inspections</li> <li>• Traffic information (traffic jams, etc.)</li> <li>• Real-time anonymised data of people flow in a city for research purposes</li> <li>• Restrictions</li> <li>• Parking information</li> <li>• Aggregated urban mobility data</li> <li>• Accidents</li> <li>• UVAR</li> <li>• INSPIRE themes</li> </ul>

---

### 2.8.5 Additional feedback provided by respondents

Many survey respondents also ensured that they completely support the Commission's view that public authorities should make a broader range of data, which have significant potential for re-use and can benefit the general interest, available. However, several factors need to be taken into consideration when doing so.

The efforts to open up specific high-value datasets provided they are free of charge, in a machine-readable format and via an API was highly welcomed by several respondents, but it was expressed by most of them that clear legal rules, and sound data policies are essential, containing information about the origin of data (e.g. how they were collected, and the owner of the data). These policies should ensure that any non-sensitive government-generated data asset is made freely available, and these also need to be clear in order for private companies to use that data. To foster easy re-use, the use of the European Open Data portal is envisaged by several respondents. It is therefore essential that national data portals are kept up-to-date.

It is also important to analyse the impacts of such an activity: opening up the data economy by the use of high-value is perceived as an accelerator, which will foster data use for innovative businesses and for the public good. However, governments should carefully consider the effect that procurement policies can have on the availability of data. When creating HVDs, it is believed that the datasets falling under the scope of the Directive should also conform to good practices in the field.

Data protection is also a very crucial element, which needs to be carefully analysed when creating HVDs. Several respondents stressed that a more concrete relation to GDPR and privacy related rights is necessary. Furthermore, the role of anonymisation and pseudonymisation becomes increasingly more relevant. Anonymised data is key to foster more data sharing and usage, as well as, proper anonymisation is important to build and maintain the trust of individuals in the data economy.

Respondents also expressed that the aim of creating such datasets would be even more interesting when combined with ecosystem of private and public sectors. Furthermore, some opinions underlined that making high-value available is not enough, and as it aims to be "leading by example", the Commission should go a step further by committing investment into actual data creation.

# 3 Micro-level Analysis of High-value Datasets and Recommended Policy Options

## 3.1 Company and company ownership

This section presents the micro-level assessment for the thematic area of company and company ownership. It illustrates the current state of play of the provision of these datasets. Furthermore, it provides the recommended measures for publication together with the costs and benefits of including these datasets as high-value datasets under the PSI Directive. Lastly, it details the three policy options proposed for this thematic area.

### 3.1.1 As-is situation: how Member States provide these datasets today

The responsible data holders and the modes of collection and provision of company datasets vary greatly across European countries and it is difficult to establish categorisations, which are universally applicable. Nonetheless, for basic company information (personal and non-personal) and company documents and accounts, four categories of approaches are the most frequent:

- 1) *Category I*: national authorities (ministries of justice or ministries of business, central agencies and also statistical offices) collecting, holding and providing these datasets (i.e. Denmark, Bulgaria, Ireland, Greece, Finland, France, Malta);
- 2) *Category II*: courts collecting and holding these datasets and making them available through interconnected IT systems managed by central authorities (i.e. Germany, Croatia, Hungary);
- 3) *Category III*: external partners (mainly chambers of commerce) delegated with the task of collecting, holding and providing these datasets based on national legislation (i.e. Italy<sup>306</sup>, the Netherlands, ...)
- 4) *Category IV*: a mix of national authorities and “external partners”, in which there is a split of tasks between, for instance, national authorities and chambers of commerce (i.e. Austria or Luxembourg, with the latter setting up an economic interest grouping for instance<sup>307</sup>)

There are of course exceptions: for instance, in France, company basic information and company documents and accounts are not provided by the same authority (with INSEE being in charge of the former and INPI/Infogreffe sharing the latter). Furthermore, as mentioned in Section 3.1.1 – Datasets in scope, countries count on many different registers depending on the type of companies at hand (i.e. companies having legal entity, artisans, and specific sectorial companies).

---

<sup>306</sup> See in particular Law 580, 29 dicembre 1993, n. 580 Riordinamento delle camere di commercio, industria, artigianato e agricoltura, [https://www.cameracommercio.ci.it/moduli/Legge%20580-1993\\_TESTO\\_aggiornato\\_al%20D%20Lgs219%202016.PDF](https://www.cameracommercio.ci.it/moduli/Legge%20580-1993_TESTO_aggiornato_al%20D%20Lgs219%202016.PDF). The delegation actually dates back to prior to this legislative text but Law 580/1993 establishes the financial independence of the Chambers of Commerce which are not financed by the state anymore but can charge for their services. A second legislative initiative which was implemented in 2014 halved the amount of the fees that Chambers of Commerce can charge for their services: see Decreto legislativo 14 marzo 2013, n. 33, <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2013-03-14:33!vig=>

<sup>307</sup> <https://www.lbr.lu/mjracs-lbr/jsp/IndexActionNotSecured.action?time=1589372460668&loop=2>

Depending on the countries (and less so on the type of approach chosen), company insolvency information are also provided by the same bodies in charge of basic company information or not: i.e. in France INSEE only provides basic companies information, while in Denmark insolvency information are provided by the same central authority in charge of company basic information. In general, countries in which courts play a key role in the collection and provision of company basic information also see them playing a role in the provision of insolvency data.

Concerning beneficial ownership information, the situation is relatively less clear as the implementation of the 5<sup>th</sup> Anti-Money Laundering Directive<sup>308</sup> is still ongoing. According to DG FISMA, a majority of countries is still setting up their beneficial ownership register and therefore the modes of provision are not entirely crystallised (i.e. in France, Czech Republic ...). This is also reflected in the more limited information gathered by the countries' experts. Where beneficial ownership registers already exists (as it is the case in Ireland, Slovenia or Luxembourg) they are very often provided by the same organisations in charge of basic company information.

Finally, it is worth reminding here that countries have different systems and processes for not only providing but also collecting and validating information. In some Member States, a human validation is always required before a piece of information provided by a company is made available in the datasets for users (i.e. Italy, Greece). In other countries, no validation process exists and the information provided by company A appears, in almost real time, in the database for reuse (i.e. Denmark). This question is particularly important for the beneficial ownership information as these are self-declared by companies in many countries. For these information, the question of quality and veracity of data is crucial as the value of this datasets depends on their reliability. When the entries to the beneficial ownership registers are self-declared by companies and stakeholders advocate for "verification" mechanisms to be put in place. Although this aspect falls beyond the scope of the PSI Directive, it must be taken into account by the analysis as including low quality data in the HVD list might not be optimal.

The table below provides a high level overview of all information gathered on today's modes of provision of company information across Member States<sup>309</sup>. The table makes the distinction between basic company information, company documents accounts, company ownership and company insolvency status as much as possible. Furthermore, the table clearly indicates the data gaps and the type and scale of information missing. In particular, concerning the topics of data linking, shared vocabularies and taxonomies, the data collection did not allow to gather very extensive information. These topics seemed to be less relevant for data holders and almost never came across from the interviews and discussion with Member States experts. For these reasons, these characteristics of data provision are often marked as Not Applicable in the table below, although they are shortly discussed in Section 3.1.3.1 – Recommended measures for provision. Finally, the table indicates with an asterisk close to the name of the country (\*) those Member States for which information were only gathered through desk research and interviews could not be scheduled.

---

<sup>308</sup> Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L0843>

<sup>309</sup> It should be noted that company information is provided in different registers across Member States, having different characteristics, policies, and conditions.

Openness-data specification					Documentation					Completeness				
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timeliness	Granularity	Key attributes
Austria	National license	No (around 10 000 per year for full access)	XML	Yes	API available. Bulk download not available for the Firmenbuch but yes from Chamber of Commerce	Available	N/A	Available	N/A	N/A	N/A	Updated every morning	Individual company and individual person	Key identifiers available
Belgium <sup>310</sup>	Two different licenses for commercial and non-commercial reuse of data <sup>311</sup> . Terms of use and log in for individual searches.	Two open datasets can be used free of charge: one is open to the general public for searches, and the other one with basic information for reuse in applications (to download the entire file though you need to pay).	CSV, XML	Yes	Bulk download and web service allowing reuse available (no API)	Available	N/A	Available	N/A	N/A	Not required	Almost daily update (for paying services), less frequent for open datasets (weekly or monthly)	Individual company	Key identifiers available
Bulgaria	No terms of use or license for the data on the agency's website but registration required to consult entire database <sup>312</sup> / open	Consultation of the individual company data and entire register is for free.	PDF for the, XML for the open dataset	Not for the data on the agency's website (but open dataset available on top)	Bulk download yes (for the open dataset) / API no	No	N/A	N/A	N/A	N/A	N/A	Real time update for the agency's data / Update every three month for the open dataset	Individual company	Key company identifier available (no person identifier)

<sup>310</sup> The data reported in the table refers to the datasets of the Crossroad Bank for Enterprises (BCE/KBO). This register includes information on companies (including their status, i.e. if they are undergoing a bankruptcy procedure) but does not include the beneficial ownership information nor the insolvency document.

<sup>311</sup> It should be noted that despite being two separate license agreements (one for commercial, and one for non-commercial purposes), there is no distinction anymore.

<sup>312</sup> The central register agency is in charge of the commercial register and the non-profit entities register. These registers are electronically provided but the possibility of reusing the data is limited as there is no bulk download possible (only individual searches or searches through the entire register). An open dataset is provided on the Open Data Platform of Bulgaria and contains basic information on companies and legal entities. This can be bulk downloaded.

Openness-data specification			Documentation						Completeness					
	license for the open dataset (provided on the open data platform)													
Croatia (*)	Information not available	Yes	Information not available	Information not available	No	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available
Cyprus (*)	Information not available	Basic info are for free but complete information are charged for	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available
Czech Republic	Open license	Yes	CSV.GZ, XML.GZ, XML, CSV	Yes	Only individual search and bulk download possible.	Available	N/A	Available	N/A	N/A	N/A	The business register is updated regularly but without prior notification to the reusers. The insolvency register is updated annually.	Individual company and individual person	Key identification number provided (ICO)
Denmark <sup>313</sup>	No license for the download of the data but registration required for the API of the CBR, and some terms and conditions for marketing reuse	Yes	JSON (CBR) XBRL, XHTML, and PDF (data from financial reports)	Yes	Yes, CBR data require a login but not conditions attached. The API technology makes possible to bulk download (but no user support is offered for this feature). The Årsrapporter does not require a login.	Not available	N/A	Available	N/A	N/A	N/A	Real time/ eventually consistent (every time a user makes an update)	Individual company and individual person	Key identification number provided (CBR)

<sup>313</sup> The information contained in the table refer to the CVR and the Årsrapporter (årsregnskaber) – data from financial reports. The CBR datasets include the beneficial ownership information.



Openness-data specification					Documentation					Completeness				
Estonia	Terms of use (contracts) with main reusers  Open License (CC3.0) for the open data sets	Single searches are for free and there are some basic open datasets, bulk download is charged for. API is also charged for and available for few users only.  Beneficial ownership information are charged for (except for organisations which have the obligations to check for the beneficial ownership).	CVS/XML	Yes	Yes, API available but for biggest reusers only on contractual basis. Bulk download available.	Available	N/A	Available	N/A	N/A	N/A	Real time for all datasets except the open data (once a week).	Individual company and individual person	Unique identifier (company number) available  For beneficial owners and persons the Estonian eID system is used
Finland <sup>314</sup>	No license but regular customers can sign an agreement with the data holder. Terms of use for bulk download.	Basic info for free but more advanced charged for.	CVS/XML	Yes	API and bulk download both available.	Available	N/A	Available	N/A	N/A	N/A	Real time	Individual company and individual person	Key identification number provided
France <sup>315</sup>	Open national license (license Etalab - compliant with creative commons) +	INSEE's data are for free. INPI-RNCS data also for free.	CSV/XML/PDF	Yes for both INPI and INSEE	INSEE and INPI have both API and bulk downloads.	Available for INSEE and INPI	N/A	Available for INSEE and INPI	N/A	N/A	N/A	Monthly for INSEE / Real time for INPI	Individual company and individual person	Key identification number provided by INSEE (numéro SIRENE)

<sup>314</sup> No insolvency register available in Finland. Insolvency information are held by courts.

<sup>315</sup> In France there are several relevant register: 1) the INSEE database SIRENE, 2) the Registre national du commerce et des sociétés provided by the INPI and 3) several smaller registers (the register of Craft businesses, ie "le Répertoire des Métiers" [data holder: Chambres des Métiers et de l'Artisanat, the register for commercial agents, ie "le registre des agents commerciaux, the register of "agricultural assets" ie "le registre des actifs agricoles, the Agriculture Register, ie "le Registre de l'Agriculture, the register of inland waterways businesses, ie "le registre de la batellerie artisanale". According to a French law published in May 2019, all these dedicated registers, whose objective is the information of third parties, will be all "integrated" in one register: "the General Register of Enterprises", which should work from 2023 and should not be confused with the INSEE register. For this analysis, we focus on the INSEE, Infogreffe and the INPI's registers which are already available.

Openness-data specification					Documentation						Completeness			
	terms of reuse													
Germany <sup>316</sup>	Terms of use: the information cannot be reused but just consulted.	Registration and searching for free. Paying for retrieval of documents.	PDF	No	No API nor bulk download	Not available	N/A	N/A	N/A	N/A	N/A	Updated when changes occur	Individual company and individual person	N/A
Greece	Terms of use (and registration)	Search and general statistics are for free but to get official documents you need to pay	XPRL* (only starting now to collect data in this format but they idea is to make everything available in XPRL)	Yes	Yes but only for the general statistics and not for company accounts or other company information	Available for statistics	N/A	Available for statistics	N/A	N/A	N/A	Mostly real time (some data needs to be validated by the administration)	Individual company and individual person	Key identifiers provided
Hungary (*)	Information not available	Basic info for free. More advanced and web services charged for (monthly fee).	XML	Yes	No API but web service for bulk download available for companies to get data automatically.	Available	N/A	Available for web service	N/A	N/A	N/A	Updated weekly	Individual company and individual person	Key identifiers provided
Ireland	Licenses for bulk download. Terms and conditions for the API.	Beneficial ownership information are charged for. Some company's information for free and some charged for.	XML, JSON	Yes for company data/beneficial ownership only search function	API available (improved version under construction)	Available	N/A	Yes	N/A	N/A	N/A	Real time	Individual company and individual person	Key identifiers provided
Italy	License imposing only traceability of data.	Only basic info are for free	XML	Yes	Both available	Available	N/A	Yes	N/A	N/A	Imposed by the license.	Real time	Individual company and individual person	Key identifiers provided
Latvia (*)	Information not available	Basic information for free and in open data	.csv, .txt or .xlsx	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Updated daily	Information not available	Information not available

<sup>316</sup> Germany's position is that data contained in the commercial and insolvency registers does not fall in the scope of the Directive. The data are held by courts which, according to the German interviewees, are not covered by the Directive. Furthermore, the data is not stored centrally but provided through a decentralized IT infrastructure managed by one Lander on the behalf of all others. All the courts are connected to this infrastructure.

Openness-data specification					Documentation						Completeness			
Lithuania	Contractual basis for regular users/terms of use for single users	Free statistical information and basic company information (up to 100 searches per day). All the rest is charged for.	xlsx, csv	Yes	Not yet. However the new open data portal will have an API and might include some company datasets.	Not available	N/A	Not applicable yet	N/A	N/A	N/A	Every 3 months	Individual company and individual person	Key identifiers provided
Luxembourg	No license (reuse defined by law – only searches by company name or number)	Documents are provided for free but l'extrait de registre de commerce is paid for <sup>317</sup> .	PDF	No	No bulk download, no API	Information not available	N/A	Not applicable yet	N/A	N/A	N/A	Real time	Individual company and individual person	Key identifiers provided
Malta	N/A	Basic info for free. More advanced services charged for	PDF	No	API available (but only for public authorities), bulk download not available	Not available	N/A	Yes	N/A	N/A	N/A	Real time	Good	Unique identifier (company number) available
Netherlands	Yes	Basic info for free. More advanced services charged for	PDF, XML	Yes	API and bulk download available	Not available	N/A	Not available	N/A	N/A	N/A	Real time	Individual company and individual person	Key identifiers provided
Poland	No license	Data are generally free of charge. However in some cases, when a user needs specially prepared, comparative dataset, it is possible to obtain such paid service on an individual request <sup>318</sup>	Xlsx, xml	Yes	API and search function available / bulk download available through the API	Available	N/A	Available	N/A	N/A	N/A	Real time (when updates are available)	Individual company and individual person	Key identifiers provided

<sup>317</sup> [https://www.lbr.lu/mjrsc/jsp/webapp/static/mjrsc/en/mjrsc/pdf/tarifs.pdf?FROM\\_MENU=true&time=1589374427482&pageTitle=menu.item.geninfoprices&currentMenuLabel=menu.item.geninfoprices](https://www.lbr.lu/mjrsc/jsp/webapp/static/mjrsc/en/mjrsc/pdf/tarifs.pdf?FROM_MENU=true&time=1589374427482&pageTitle=menu.item.geninfoprices&currentMenuLabel=menu.item.geninfoprices)

<sup>318</sup> In 2019, 143 specific requests were logged through the REGON system and this led to a revenue of around 135 000 PLN (around 30 000 euro).

Openness-data specification				Documentation						Completeness				
Portugal (*)	Information not available	Access is free, extracts are paid for	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available
Romania (*)	Login requested to search the database	Some info for free and some not	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available
Slovakia (*)	No license	Access is free, retrieval is paid for	Information not available	Yes	No API, no bulk download	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available	Information not available
Slovenia	National terms of use for full datasets / CC4.0 for open datasets	There are some limited data available in open data but access to full datasets of company data is paid for. Impossible to access fully beneficial ownership and insolvency <sup>319</sup>	XML, CSV, PDF	For company data, yes	API and bulk download for company data (charged for). Bulk download for open datasets <sup>320</sup> .	Yes	N/A	Yes	N/A	N/A	Yes. Obligation to provide the source of the data (AJPES) and date of retrieval.	Open data updated quarterly. Paying datasets updated monthly, weekly or daily depending on subscription.	Individual company and individual person	Unique identifier available for companies (and publicly available). Unique identifier for beneficial owners only available to authorised bodies.
Spain <sup>321</sup> (*)	Yes	No	PDF, TIFF	No	API available, bulk download not available	No	Yes	No	N/A	N/A	Yes. Obligation to mention the source.	Daily	Individual company and individual person	Unique identifier available for companies (company number). Unique identifier for individual persons (ID) <sup>322</sup>
Sweden (*)	Terms of use are imposed to those registered	Access to basic information is free, more detailed	Information not available	Yes	No	N/A	N/A	N/A	N/A	N/A	N/A	Information not available	Individual company and individual person	

<sup>319</sup> Only individual searches possible, reuse not possible.

<sup>320</sup> Only individual searches possible, reuse not possible.

<sup>321</sup> There are two types of business registers in Spain: the central register (*Registro Mercantil Central*), and the local ones (*Registro Mercantil Provincial*). The table above displays the information for the latter, as they are the business registers collecting the information, and providing the information to the reusers.

<sup>322</sup> Only for those individuals part of the management, or those representing the sole shareholder (and the sole personality of the company has been declared).

---

**Openness-data specification****Documentation****Completeness**

---

as regular  
clients. information  
and extracts  
are paid for.

---

A few considerations can be deducted from the table above and should be taken into account when identifying options for the future:

- Almost all countries rely on licenses and terms of use to regulate their relation with reusers.** Only very few have moved to open licenses (whether international such as Creative Commons 0 or national such as the Etalab license) and no terms of use. Frequently, a log in is required to access the data or use the APIs and terms of use are imposed through this registration system. The reason for having licenses/terms of use is always linked to the importance of ensuring that reusers use the data in a legal way (i.e. respecting GDPR) and to limit liability for the data providers in case of wrong practices. There is a strong concern amongst the holders on the risks of using open licenses and no terms of use due to the loss of control on how their data could be used for. In some countries (i.e. Slovenia), the Data Protection Authorities (DPAs) have even sanctioned certain types of reuse which are considered as totally legal in other countries, such as the possibility to build networks of connection (for instance linking all companies sharing the same directors)<sup>323</sup> and procedure are ongoing to further restrict personal data provision in this area. In this context, terms of reuse and licenses are considered the only way to protect the data providers from having to suffer severe legal consequences and sanctions from DPAs. Furthermore, terms of use are often considered critical to maintain the IT system working: for instance, data holders in multiple countries put a cap on the number of API calls per day to avoid overload and crash of the system.
- Charging for data or services is still a very common practice.** Only a very small minority of countries is already providing all types of company information (or a majority of them) entirely for free. The three best examples in this respect would be France and Denmark (within the European Union) and the United Kingdom (outside the European Union). For those data holders charging for data, the business models, charging approaches and prices vary considerably but, generally speaking, there are two formulas which are the most widespread: 1) charging for retrieval of authenticated documents/documents/special documents (i.e. Luxembourg, Lithuania) and 2) charging for access to entire datasets (often including personal information which are not provided for free, i.e. Sweden, Italy, Slovenia, Belgium). The revenue generated by countries also varies, as shown by the examples included in the table below. It should be noted that the availability of some company data fields would not necessarily imply the loss of all these revenues, but the proportional share of it.

Table 20 – Yearly revenue linked to company information charging per country

Country	Revenue (approximation/per year)
Belgium	450 000 euro
Finland	2.7 million euro
(France	6 million euro for INSEE and 2.4 million for INPI) <sup>324</sup>
Germany	20 million

<sup>323</sup> Stakeholders interviews.

<sup>324</sup> Before the transition to an open data model in 2017, a compensation of 11 million euro was attributed to INSEE since but no compensation was foreseen for INPI.

Estonia	2.6 million euro
Ireland	7 million euro
Italy	58-60 million euro
Poland	Around 30 000 euro (for specific services only, the data being provided for free)
Slovenia	1 million euro
The Netherlands	50 million euro

- **There are differences in the format used for data provision but a relatively important number of countries uses xml and csv.** Some countries, i.e. Greece, are now working on changing the format of their entire datasets to make them more machine friendly and allow automatised business processes and reuse.
- **In terms of machine readability, countries have made very good progress and this is a reality for a strong majority of them already** (coherently with the widespread use of the xml format). When machine readable data is not provided this is sometimes due to a conscious choice to limit reuse (i.e. in Germany) and it is not because of a lack of investments or limited digitalisation efforts.
- **Only around ¼ of the countries have set up APIs for the provision of company information today and an even smaller number has both API and bulk download.** To some extent, this is linked to the fact that many data holders are sticking to the individual search approach as the most relevant approach for searching company information. Individual searches, which fulfilled the need of the past use/business cases (i.e. an individual/organisation looking for one company's information), are not adequate anymore and many data holders are already transitioned/are transitioning to APIs in particular. Again, in some cases not having an API is a conscious choice to avoid reuse of data (i.e. Germany). Amongst the countries who have already set up an API, bulk download is not always provided on top. Vice versa, bulk download is possible in a number of countries which do not have an API: yet, the combination of the two is quite rare. When API and web services are available (i.e. France, Finland, Ireland, Slovenia), these are always or almost accompanied by the provision of metadata and supporting documentation.
- **Timeliness of data varies significantly depending on the categories of datasets at hand and on the country.** There are various philosophies in terms of frequency of update of the data: some countries work in real time (meaning, updating the datasets every time there is a change) while others plan daily, weekly or even monthly and quarterly updates. This also depends on the underpinning IT system and on the validation process which is applied for companies when they update their business information in the registers: in Denmark for instance, the IT infrastructure connects back ends and front ends. This means that, when company A updates for instance the name of one of its director, this information becomes immediately available in the datasets provided to reusers. In other countries however (such as Italy and Greece for instance), a validation from an official needs to take place before the information is approved by the system. Furthermore, in some countries (especially the Nordic countries) companies have to provide some documents and information once a year and this means that the biggest update of the datasets happens around that time. Finally, some countries provide more or less timely data to reusers depending on their type of subscription (i.e. Slovenia offers

subscription packages with monthly, weekly and daily updates at different prices and the same happens in Belgium).

- **Key attributes such as companies identifiers are almost always made available.** Each country has its own system of companies and individual identifiers for disambiguation. Individual identifiers are very important for beneficial ownership information but sometimes also represent very sensitive information to share with reusers: in Estonia for instance, the persons’ identifiers are their Estonian ID number. This requires the data holders to be extra careful on sharing this information and allowing its reuse. In some countries such as Austria, Slovenia, The Netherlands, beneficial owner identifiers are not shared to reusers due to their sensitivity, and are only available for dully authorised public authorities (e.g. law enforcement).

### 3.1.2 To be situation: extending the PSI HVD rules to these datasets

This section presents the recommended measures for publication, as well as the expected costs and benefits of including these datasets under the scope of the PSI Directive as HVD.

#### 3.1.2.1 Recommended measures for publication

As discussed in the previous section, measures for publication of company datasets in scope of this analysis differ to a very large extent, and despite this, users are pretty much aligned on what they need and how data should be published for them to exploit it. The discussion with reusers pointed at very clear measures for publication for the datasets in scope and, in fact, these measures have been largely adopted already by the countries which are considered as best practices in terms of company data provision (i.e. France, Ireland, Denmark and Finland in Europe, and the United Kingdom outside Europe<sup>325</sup>).

The table below summarises the recommended measures for publication for the four categories of datasets which have been considered (basic company information, company documents and accounts, company ownership, company insolvency status).

Table 21 – Recommended measures for publication of company datasets

Dimensions		Basic information	Company documents and accounts	Company ownership	Company insolvency status
Openness-data specification	License (terms of use)	CC.BY 4.0 No terms of use No database right			
	Format	XML - Json			
	Machine-readability	Mandatory			
	Availability of API, bulk download	Both API and bulk download			
Documentation	Metadata (dataset)	Complete (*.csv document available) <sup>326</sup>			

<sup>325</sup> The analysis of OpenCorporates on the accessibility of company datasets across European Countries confirms that France, Denmark, Finland and Ireland are in the top five of the ranking, together with Bulgaria, Cyprus (which could not be reached out by the study team) and Latvia (which could not be reached out by the study team), see: <https://opencorporates.files.wordpress.com/2020/06/eu-company-data-state-of-the-union.pdf>

<sup>326</sup> Based on the ESS consultation, metadata should include: i) definition of variable; ii) validity time stamp of variable (reference time of the variable); iii) source of updating; iv) data when last updating; v) code lists; vi) standards; etc.



Dimensions		Basic information	Company documents and accounts	Company ownership	Company insolvency status
	<i>content description</i> )				
	<i>Data linking</i>	No specific recommendation			
	<i>Documentation (incl. structure and semantics)</i>	Complete and web-available			
	<i>Shared vocabularies</i>	Recommended but not mandatory			
	<i>Taxonomies</i>	Recommended			
Completeness	<i>Traceability</i>	Not necessary			
	<i>Update frequency and timeliness</i>	When available (real time) <sup>327</sup>			When available (real time) (minimum daily for insolvency data) <sup>328</sup>
	<i>Granularity</i>	Individual company level (plus identifier)	Individual owner (plus identifier)	Individual company level (plus identifier)	
	<i>Key attributes</i>	Company code for disambiguation	Beneficial owner code for disambiguation	Company code for disambiguation	

As the table suggests, the same recommended modes of provision apply to the four categories of datasets in scope, except for the aspects concerning the completeness of information for which some nuances must be made. The justifications for each of these recommended measures are the following:

- Concerning **licenses and terms of use**, reusers and open data frontrunners in this domain agree that open licenses must be the norm and terms of use should not be imposed as they defeat the purpose. Any restriction in terms of license or terms of use in fact hampers the economic and social value of the data as it is detrimental to the reuse of the information. Countries like Denmark or France already apply the open license, or even no license policy, although they might still impose some terms of use in certain cases (i.e. reuse of Directors' names for marketing purposes in Denmark is forbidden). While CC0 license is preferred by reusers, open national licenses could still be accepted although a multiplication of non-harmonised national open license could possibly hamper the cross border reuse of these datasets. From this perspective, CC0 should be used whenever possible. It is also important to highlight here that data holders should not have the possibility to claim datasets rights over the entire datasets, as this would not be coherent with an open data approach.
- Format** wise, there is a strong consensus amongst stakeholders on the relevance of the XML formats whenever applicable as it is both human and machine readable and it is already in use in many countries, and not only in the open data frontrunner countries. The JSON format, which is particularly

<sup>327</sup> This recommended measure is based on the findings from the data consultation activities conducted under this study. Please note that this diverges considerably from the outcome of the ESS-wide consultation communicated by Eurostat, where quarterly updating was the more prevalent position.

<sup>328</sup> Idem.

appreciated by reusers, is also already leveraged in Ireland but does not seem to be in use in any other country at the moment. Other widespread formats (i.e. CSV) could be accepted but possibly coming on top of the XML and JSON formats.

- While provision of these datasets through APIs would be mandatory under the PSI Directive for HVDs, **both APIs and bulk download should be made possible** in order to foster the reuse of these datasets. This is because these two provision modes suit different use cases: APIs allow the provision of real time services (i.e. for companies such as Altares which offer KYC services) while bulk download allows to carry out analysis (i.e. NGOs analysing business transparency in a country) and can be used to feed and train Artificial Intelligent systems. Providing these datasets in API only would therefore limit the number of use cases which could build on these datasets.
- Although not always provided today, **metadata (complete and in csv format) and complete documentation (web available)** are considered as absolutely necessary for the reusability of these datasets. As many countries only allow individual search of information and do not provide API or consent bulk download of the data, reusers are now used to live without them<sup>329</sup>. Nonetheless, from a HVD perspective and when API and bulk download will be the norm, the absence of metadata and full documentation could have profound consequences on the extent to which the data can be reused. Hence, the provision of accurate metadata and complete and web available documentation should become mandatory (although it would require some investments in all countries, as discussed in the next section). As indicated by Eurostat, "it is imperative that high-quality metadata are made available for each national companies and company ownership HVD, including variable definitions (with the data source clearly specified for each variables, validity time stamp of the variable, code lists applied, standards etc.)".<sup>330</sup>
- Concerning **data linking**, no specific recommendations can be made based on the data available. This topic does not seem to be a top priority for reusers at the moment.
- **Shared vocabularies and taxonomies** in the domain of company data are highly valued by reusers. Nonetheless, it was often mentioned that a lot of work is still required from the stakeholders' community to agree and progress on such standards and that it would be too much complicated for data holders to deal with this question right now. Considering the costs and efforts that would go into finding an agreement on this matter, it is suggested to recommend their use when possible without fully imposing them on data holders.
- View on **traceability** differ: for reusers imposing traceability of data would mean adding a lot of burden to their activities, especially today when datasets are combined from countless sources and in the context of AI applications. According to some data holders however (i.e. in Italy or Slovenia), traceability is necessary to ensure that the end consumer of the data recognises the validity of the information. As practices varies in this domain and only a minority of data holders oblige reusers to trace back the data to their database, it is recommended not to impose any form of traceability and leave it up to the data holder to ask for this feature, when needed.
- From a reuser's perspective and to facilitate the reusability of the data, the question of **update, frequency and timeliness** of data is pivotal and it was made very clear that non-timely data hold no social and economic value whatsoever. Therefore the timeliness of the information is a pre-condition to qualify these datasets as high value. In fact, "we are entering a world where there will be an explosion in the number, speed and complexity of companies, brought about by companies being incorporated by computer programs... this will bring a phenomenon of what 'firefly companies' – firms that exist for mere hours, minutes, even seconds, and corporate networks that change every day, driven by programmatic company formation"<sup>331</sup>. For this very reason, data should made available in real time,

<sup>329</sup> A country which provides already today good metadata and documentation is Finland.

<sup>330</sup> Eurostat, High-value (HVDs) on Companies and company ownership, Feedback based on an ESS consultation.

<sup>331</sup> <https://blog.opencorporates.com/2020/03/05/oecd-anti-corruption-integrity-forum-can-aml-survive-in-a-fireflies-world/>

which, for this type of datasets, means that changes in the database should appear to reusers as soon as validated by the data holders. There are some differences across the categories of datasets in terms of when such changes could occur though: for instance, basic company information can change at any time. There are countries in which you can open and close a company within the same day (i.e. Estonia). However, in many countries companies' accounts are only provided once a year (i.e. in Denmark it would be during the month of May) and therefore this is the period in which most of the changes to the database would occur. Changes to company ownership and insolvency datasets could also happen daily. For reusers, it is adamant that the datasets should be then provided in "real time" and that change logs should be made available for them to be able to only consider changes since their last retrieval of the data. Today, only a few countries operate all these datasets in a real time or near to real time mode though.

- Concerning the **granularity of the datasets**, this is not a particularly difficult topic in the area of company data as the granularity required is at the individual company or individual person level for the different datasets and this is how the data is provided today.
- Finally, in terms of **key attributes**, it is very important to underline the role that identifiers play for the disambiguation of companies and persons<sup>332</sup> in this domain. For these datasets to be valuable, reusers must be able to disambiguate between homonyms. As a majority of countries provide unique identifiers and these can be used for disambiguation, it is suggested to oblige data holders to make these available as key attributes. Besides this, as pointed out by Eurostat, the Global Legal Entity Identifier could be considered.<sup>333</sup>

### 3.1.2.2 Expected costs

Taking into account the scope of the datasets to be made available as HVD and building on the recommended modes of provision suggested above, it is possible to discuss the expected costs that Member States would have to face when adapting to the PSI HVD rules for company and company ownership. In this respect, it is first necessary to describe what are the categories of costs that data holders bear today for the provision of company data in order to have a baseline and some insights on the magnitude of present costs. The table below summarises the main cost drivers for the provision of company data and indicates the information collected for each of these categories.

Table 22 – Cost drivers for company datasets

Cost category	Description	Insights from the data collection
Infrastructural costs	Costs related to infrastructural investments such as portals, APIs, Servers (cloud), etc.	<p>Infrastructural costs vary very significantly across countries and depending on the IT infrastructure chosen of course. Some countries mentioned that IT costs for the provision of the datasets related to companies and company ownership cannot be distinguished from the overarching IT costs of the organisation and therefore precise figures cannot be provided (i.e. Poland). However, a few countries shared more precise information on their infrastructural costs for the setup of the data provision mechanisms:</p> <ul style="list-style-type: none"> <li>- <b>Czech Republic:</b> set up of the open data infrastructure (only) amounted to 10 000 euro. The maintenance costs of the infrastructure cannot be quantified as they are part of the operational costs of the data holder.</li> <li>- <b>Greece:</b> the running and maintenance of the infrastructure costs around 3.5 million euro per year (including not only the costs related to the servers and IT components but also the FTEs needed to ensure continuity of the infrastructure).</li> </ul>

<sup>332</sup> Unique identifiers of physical persons in their capacity of company managers, owners of companies, or shareholders in them, will be possible to the extent that privacy legislation in force permits it.

<sup>333</sup> Eurostat, High-value (HVDs) on Companies and company ownership, Feedback based on an ESS consultation.

		<ul style="list-style-type: none"> <li>- <b>Ireland:</b> the yearly IT budget corresponds to around 1.5 million euro to keep the system functioning. To this amount one should add the one-off investment which is currently being made to upgrade the entire API and IT infrastructure.</li> <li>- <b>Italy:</b> in the last few years around 4-5 million euro have been invested in improving the IT infrastructure and modernising it. These costs only relate to new initiatives and do not include the regular maintenance costs.</li> <li>- <b>Lithuania:</b> an investment of around 400 000 euro has been made for the development of the new open data portal which should also include some company datasets in the future.</li> <li>- <b>France:</b> INPI spends around 2-3 million euro per year in developing and maintaining the API infrastructure and underpinning IT system for the trade register.</li> <li>- <b>Slovenia:</b> the cost of IT infrastructure in Slovenia is of around 300 000 euro annually.</li> <li>- <b>Finland:</b> for the Finnish Patent and Registration Office, total infrastructural costs amount to 1.5 million euro per year.</li> </ul>
Data transformation costs	Costs related to data processing including data cleaning, preparation of metadata, aggregation, anonymisation, etc.	The data transformation costs are rarely available, as data transformation is considered business as usual by companies' registers and the related costs are too "hidden" in the budget of the concerned data holders. When specific open data initiatives have recently started, i.e. in Lithuania, costs of the overarching initiatives are available (around 2 million euro for the 2020-2023 period) but cover well beyond the thematic area of company and company ownership. For the Finnish Patent and Registration Office, cost of human resources for both data transformation and operational costs amount to 400 000 euro per year <sup>334</sup> .
Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	Similarly to the data transformation costs, the operational costs are rarely made explicit. In France, it is estimated that running the data services requires minimum 5-6 FTEs (for a total of more than 400 000 euro per year) <sup>335</sup> to INSEE and around 10 FTEs to INPI (more than 700 000 euro per year) <sup>336</sup> . As mentioned above, in Finland, costs for human resources for both data transformation and operations correspond to 400 000 euro <sup>337</sup> .
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	No precise figures could be collected on the other costs. It is widely acknowledged that, due to GDPR concerns for sharing personal data, legal advice is a cost driver but almost no country is able to indicate the magnitude of this cost, also because advice on data protection is most often sought internally (i.e. advise is asked to the data protection authority or the ministry of justice). The Finnish Patent and Registration Office shared that overhead costs amount to up 600 000 euro annually but it is unclear what these costs include.

As the table and the data collected suggest, data transformation and operational costs are not often quantified by countries considering that the collection and provision of these datasets is part of their traditional missions and data holders struggle even to count the FTEs which are dedicated to service provision. On the other hand, data on the infrastructural costs is generally more available: countries, which have made investments in new platforms (and APIs) and are running such new infrastructures, have a decent view on the costs that these entail. Finally, other costs, such as the costs of GDPR advice, are never really established.

<sup>334</sup> Stakeholders interview

<sup>335</sup> Indicative figures based on French cost of labour taken from Eurostat (2019 statistics) and a number of working days of around 250 per year. See: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages\\_and\\_labour\\_costs](https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages_and_labour_costs)

<sup>336</sup> Indicative figures based on French cost of labour taken from Eurostat (2019 statistics) and a number of working days of around 250 per year. See: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages\\_and\\_labour\\_costs](https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages_and_labour_costs)

<sup>337</sup> Stakeholders interview

It is also important to mention here that for many countries it is difficult to split their costs according to the structure presented above and, in these cases, only total figures for costs could be shared. This is the case in Estonia or Ireland for instance, which are spending respectively 2.6 and 2.7 million euro yearly to provide company data (all types of costs included). Finally, in many countries budgetary rules are such that costs for the data provision must match the revenue generated by the data selling: this is the case in Denmark, Estonia, Finland, Germany, Ireland and The Netherlands<sup>338</sup> for instance. In fact, the vast majority of data holders charging for the dataset do not have a mandate to generate revenue but rather to cover their data provision costs only. This means that all the revenue generated is reinvested in the service.

To understand how this cost structure would be affected by the implementation of the Directive one then needs to look at what could be (in general) the main budgetary implications of the recommended measures for publication those countries which are not yet close to them, as shown in the table below.

Table 23 – Recommended measure for publication for company datasets

<b>Recommended dimension for publication</b>	<b>Budgetary implication (little to none, low, medium, high)</b>
License: open license/no terms of use/no database right	Little: changing license (and especially if countries adopt readily available licenses such as CC0) would have little impact on countries' costs. Possibly, the adoption of such license could reduce the amount of resources needed to deal with legal matters and reuse rights. Small increase in the other costs (the costs of GDPR advice) could be possible in some countries depending on their data protection concerns and approaches. Furthermore, it might be required to invest in more advanced servers to cope with a possible increased amount of APIs calls per minute (if no restriction is imposed in this respect through terms of use). Nonetheless, these costs should remain relatively limited.
Format: XML/JSON	Medium: adapting the data provision to the XML/JSON formats would require a certain level of investments for countries which are currently using other formats. It is difficult though to provide an indicative idea of the magnitude of these costs also because they are strongly related to the costs of the modes of provision (see below).
Modes of provision: both API and bulk download	High: the establishment of APIs and bulk download would be the most impactful change in terms of costs for those countries which do not have one or the other (or both). Recent studies indicate that costs of APIs establishment range in between 30 000 euro and 2.5 million euro depending on the type of infrastructure and technical characteristics <sup>339</sup> and that in average an API set up costs 50 000 euro <sup>340</sup> . It is important to mention here that, compared to data from other thematic areas, company data are not dynamic and this makes the API less expensive. Nonetheless, the volume of these data remains considerable (i.e. datasets including six millions of companies in Italy, more than sixty million entities in France).

<sup>338</sup> In The Netherlands, a political decision was made to cover the overall costs (around 100 million euro) by the government, and the revenues generated, equally.

<sup>339</sup> Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, <https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en>

---

**Recommended dimension for publication**      **Budgetary implication (little to none, low, medium, high)**

---

Metadata and documentation: complete and in .csv format      Low: the cost of providing metadata and documentation in the .csv format can be considered as low if compared to the costs of setting up an API. IT literature considers that you need around a week of an FTE to create the documentation for an API and then one off costs for updating the documentation needs to be foreseen. For the metadata, there are no reliable estimations in terms of resources needed for company data<sup>341</sup> but, according to data holders, providing metadata would be more burdensome than providing documentation<sup>342</sup>.

---

Data linking: /      None: there is no recommended measure concerning data linking to be implemented by Member States.

---

Shared vocabularies and taxonomies: recommended      Little to none: adopting shared vocabularies and taxonomies could be costly for Member States but these are only recommended and therefore there could be little to no costs in the short term.

---

Traceability: recommended      Little to none: traceability is imposed only in very few countries and the adoption of the recommended measure would have limited to no impact on them.

---

Update frequency and timeliness: when available (real time)      High: the recommended frequency of publication would also be very costly to implement for those countries which do not yet work in real time or close to real time. Nonetheless, it must be clarified here that the PSI Directive does not affect the data validation processes of the data holders. That means that for those data holders imposing human validation of the company data, real time should be interpreted as “when the data has been validated by the responsible person and pushed in the system”. For countries, like Denmark, where no such validation exists and where back end and front end infrastructures are fully integrated, real time then means “as soon as the data is in the system”. For those two categories of countries, the costs of implementation would therefore be low as no modifications to their processes would be required. Nonetheless, a considerable amount of countries still works with daily, weekly or event monthly updates, and these Member States would have to change significantly their internal processes and underlying IT infrastructure to adapt.

---

Granularity: individual company and person      Little to none: the level of granularity indicated in the recommended measures is already provided by all data holders.

---

Key attributes: key identifiers      Little to none: all data holders possess key identifiers for companies and individual and a majority already makes them available.

---



---

<sup>340</sup> Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, <https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en>

<sup>341</sup> For estimations concerning metadata provision more in general see also the Mobility chapter.

<sup>342</sup> Stakeholders interviews

As the table clearly suggests a number of recommended measures would have limited to no impact on the budgets of data holders due to a) the fact that the measures are only recommended (i.e. use of shared vocabularies and taxonomies) or b) the fact that countries already naturally provide the datasets in that way and do not need to adapt (key attributes). However, there are some recommended measures that would lead to significant budgetary implications for countries having a different mode of provision today and notably those related to the establishment of APIs and the provision of near real time data. Furthermore, it must be reminded here that countries would have to provide the four data categories (see table Table 7 – Initial lists of HVDs under company and company ownership thematic area) for free, which constitute an important change for a majority of them. Based on these considerations, the table below identifies and summarises the possible expected costs linked to the implementation of the recommended measures (at the overall level).

Table 24 – Expected implementation costs of the recommended measures

<b>Cost components</b>	<b>Cost components description</b>	<b>Magnitude of costs (range)</b>
Infrastructural costs	Establishment of the API and bulk download, adaptation of the IT infrastructure to real time provision	In between 10 000 and 2.5 million euro (for setting up the infrastructure, without counting maintenance and for each datasets, if provided by different data holders) per year/ or 50 000 euro in average.
Data transformation costs	Costs related to data processing including data cleaning, preparation of metadata, aggregation, anonymisation, etc.	In between 4 and 10 FTEs (best estimates available)
Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	
(Lost) income for data supplier	(Share of) revenue related to the provision of the HVD	Approximately in between 30 000 euro and 60 million euro per year
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	Ad hoc/not quantifiable
Negative impact on competition	The estimated impact of competition distortion vis-à-vis private organisations active in the domain.	Not applicable/data holders would not suffer from competition distortion

Building on the data collection activities and on these general analysis of costs, we have made an attempt to establish the magnitude of costs (on a scale from low costs to very high costs<sup>343</sup>) for all EU countries for which we had sufficient information. The results of this assessment are provided below.

<sup>343</sup> Low costs = costs which could be accommodated within the current budget of the data holders, medium costs = costs which would entail budgetary discussions at the country level, high costs = costs that would require significant investments at the country level, very high = costs which countries might not be able to afford on a short term.

Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
Austria		X				Main costs would relate to the loss of revenue and some adaptations required to the IT infrastructure.
Belgium		X				Main costs would concern the setup of APIs and the loss of 450 000 of revenue from paying users.
Bulgaria		X				Main costs would concern the setup of APIs and the provision of related documentation.
Croatia					X	
Cyprus					X	
Czech Republic			X			Main costs would concern the setup of APIs and the changes in the timeliness of data.
Denmark	X					Main costs would concern the provision of metadata and the extension of the current data provision system to beneficial ownership information.
Estonia		X				Main costs would concern the loss of revenue and some infrastructural changes.
Finland	X					Main costs would concern the loss of revenue and some minor infrastructural adaptation.
France	X					France is already providing data as recommended by this report to a very large extent. It has also foreseen to move towards a new data provision



Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
						system integrating datasets from the agriculture and trade register by 2023 <sup>344</sup> .
Germany				X		The German system would have to be changed completely to adapt to the recommended modes of provision and make available this list of datasets.
Greece		X				Main costs would concern infrastructural changes. IT investments and investments in adoption of machine friendly format are currently being made.
Hungary			X			Main costs would concern the loss of revenue and the establishment of an API. Timeliness of data would also need to change.
Ireland		X				Main costs would be the loss of revenue (2.7 million per year). Investments in API are currently being made.
Italy				X		The Italian system would have to be changed completely and new legislation should be established changing the role of the Chamber of Commerce as data provider.
Latvia					X	
Lithuania			X			Main costs would concern the loss of revenue and the establishment of a new IT infrastructure including an API. The newly established open

<sup>344</sup> Loi n° 2019-486 du 22 mai 2019 relative à la croissance et la transformation des entreprises, Publiée au Journal Officiel du 23 mai 2019, [http://www.assemblee-nationale.fr/dyn/15/dossiers/Croissance\\_transformation\\_des\\_entreprises?etape=15-PROM](http://www.assemblee-nationale.fr/dyn/15/dossiers/Croissance_transformation_des_entreprises?etape=15-PROM)

Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
						data portal could be used for this purpose though, thus limiting the extra investment.
Luxembourg			X			Main costs would be the loss of revenue and the establishment of an API.
Malta		X				Main costs would concern the loss of revenue and the extension of the API to all reusers (now it is limited to public sector reusers only).
The Netherlands				X		Main costs would be the loss of the revenue (up to 50 million per year), as well as some minor adaptation in terms of metadata and documentation.
Poland	X					Main costs would concern limited loss of revenue and some minor adaptation in terms of metadata and documentation.
Portugal					X	
Romania					X	
Slovakia					X	
Slovenia			X			Main costs would concern the loss of revenue amounting to 11% of the budget of the institution (1 million euro yearly).
Spain					X	

<b>Country</b>	<b>Low costs</b>	<b>Medium costs</b>	<b>High costs</b>	<b>Very high</b>	<b>Unknown (based on information available)</b>	<b>Comments</b>
Sweden			X			Main costs would concern the loss of revenue and the need to set up an API and change the existing IT infrastructure.

As the table suggests, the costs of implementation of the recommended measures for the four datasets in scope would be low for a few frontrunner countries (Belgium, Denmark, Finland, France, Poland), medium for a considerable amount of countries (mainly those already having an IT infrastructure including APIs and that would be affected in terms of revenue), and high for those countries whose IT infrastructure functioning would need to change completely to adapt to the new modes of provision (i.e. Hungary, Slovenia, Luxembourg). Furthermore, there are also three countries (Italy, the Netherlands and Germany) for which the inclusion of these categories of datasets in the HVD list would lead to very high costs linked to the need of changing the entire system which is a) generating very high revenue (i.e. 50 million per year in the Netherlands and around 60million in Italy), and/or b) based on externalisation of the provision of these data to Chambers of Commerce and/or c) built on the assumption that company data should not be reusable. Therefore, it can be argued that a considerable part of the costs for implementation would be concentrated on around 1/3 of Member States, those for which the costs would be high or very high.

When looking at the analysis of costs for two particular countries (Finland and Slovenia) which were selected for the Cost Benefit Analysis, the general insights derived by the table above can be confirmed. In Finland only 15% of the Trade register's revenue comes from selling data and, more precisely, only 5.1% comes from charging for the datasets in scope of this analysis<sup>345</sup>. The rest comes from fees of stakeholders willing to change the status of their companies. The Trade register (PRH) is requested to cover its costs and does not receive any funding from the state. Finland's provision of the datasets in scope of the assignment is already quite close to what recommended by our analysis for the different dimensions of provision and therefore, as also indicated in the table above, the main costs would concern some infrastructural adaptation (minor) and the loss of income of PRH as data supplier. The table below provides the outcome of our analysis of the costs components of Finland for the Cost-Benefit Analysis. As the table shows, the aggregated costs obtain a total score of -4 and a weighted score of -1.05.

Cost components	Weight	Score	Weighed score
Infrastructural costs	0.30	-1	-0.30
Data transformation costs	0.15	-1	-0.15
Operational costs	0.15	0	0
Lost income for data supplier	0.30	-2	-0.6
Other costs	0.05	0	0
<b>Aggregated costs of HVD</b>		-4	-1.05

When we compare the assessment of the Finnish costs with the analysis of the costs for a country, Slovenia, which is less close to the recommended modes of provision, the difference in the cost magnitude appears very clearly.

Slovenia in fact presents significant differences compared to Finland concerning the provision of its company data. The Slovenian Business Register (PRS) provides two different datasets. In the first one, basic data<sup>346</sup> is available free of charge. This data is updated quarterly, is available for bulk download. On the other hand, the second dataset offers full access to all information and is charged for. Reusers can select their subscription model based on the frequency of the data (i.e. it can be refreshed either daily, monthly or yearly), which is translated in a different range of price. In the paying dataset, the data is available for bulk download or through API. In budget, the Slovenian Business Register had a financial

<sup>345</sup> The data used for the CBA are the latest available and refer to the 2018 budget of the agency.

<sup>346</sup> This refers to the full entity's name, the registration number, business address, legal address, and the registration entity.

envelop of around 9.4 million euros for 2019, of which 44% comes from the state budget, 33% from the fees charged for public services, and 11% is generated by the re-use of their data<sup>347</sup>. The provision of data in the Slovenia Business Register differs significantly from the recommendations presented in this report based on the analysis conducted. Therefore, the costs for this EU Member States will be more significant in comparison to other countries, such as Finland.

The table below displays the assessment for the costs components as part of the Cost-Benefits Analysis. As indicated in the table, the aggregated costs for Slovenia obtain a total score -7, which corresponds to a weighted score of -1.55.

Cost components	Weight	Score	Weighed score
Infrastructural costs	0.30	-2	-0.60
Data transformation costs	0.15	-1	-0.15
Operational costs	0.15	-1	0.15
Lost income for data supplier	0.30	-2	-0.6
Other costs	0.05	-1	-0.05
<b>Aggregated costs of HVD</b>		-7	-1.55

To conclude on the costs, the establishment of APIs and the loss of revenue are the biggest cost drivers for countries when adapting to the HVD rules. Nonetheless, the timeliness of data and the provision of real time information also have a very strong impact on the IT infrastructure needed and on the underlying operational processes.

### 3.1.2.3 Expected benefits

In general, benefits of making these four datasets available as HVDs were greatly discussed with reusers but much less so with data holders. In fact, except from the study carried out by Company House in 2019<sup>348</sup> and the older study on the data market carried out in Italy by the Chamber of Commerce<sup>349</sup>, data holders have not mandated specific analysis concerning the benefits of their data provision nor they implemented mechanisms to monitor reuse and related benefits. Most data holders were able to discuss individual examples of reuse, point at specific use cases and give insights on their revenue when applicable but they were incapable of mentioning figures or providing more general insights with regard to economic or social benefits. Nonetheless, based on the information available, it is possible to describe the five main expected benefits linked to the inclusions of the four datasets analysed in this chapter in the HVD list.

#### *Increase in reuse benefits*

There is a strong agreement on the fact that providing the four abovementioned datasets as HVDs would increase their overall number of users and therefore increase reuse benefits as defined in our value framework. Countries such as France and Denmark which started to provide these datasets through APIs and for free saw a very important surge in the number of reusers (i.e. from 12 full reusers to 1230 for INPI, thus increasing of 100 times<sup>350</sup> and the SIRENE database is the third most accessed datasets on the data.gouv.fr portal<sup>351</sup>). In Belgium, the CBE portal is one of the most visited public sector portal thanks to the individual search function of the database, and the newly established open datasets (despite

<sup>347</sup> Slovenian Business Register, Annual report 2019.

<sup>348</sup> <https://www.gov.uk/government/publications/companies-house-data-valuing-the-user-benefits>

<sup>349</sup> Stakeholders interviews

<sup>350</sup> Stakeholders interviews

<sup>351</sup> <https://www.insee.fr/fr/statistiques/4238594?sommaire=4238635>

containing only basic company information) was requested by 1890 different users in 2018 and 2232 in 2019 (+8%).

The same positive trend can be found in the United Kingdom<sup>352</sup> where the number of reusers multiplied by two (from several thousands to 1.5 million in 2019)<sup>353</sup> after the datasets started to be provided for free. The study carried out in 2019 by Company House even looked at the proportion between cost reduction and increase in the reuser base to establish the customers’ sensitivity to price. The elasticity of the demand for reuse of company data is very clearly shown by the picture below and confirms what found in previous studies: for each small decrease in the cost of data the number of reusers multiplies<sup>354</sup>.

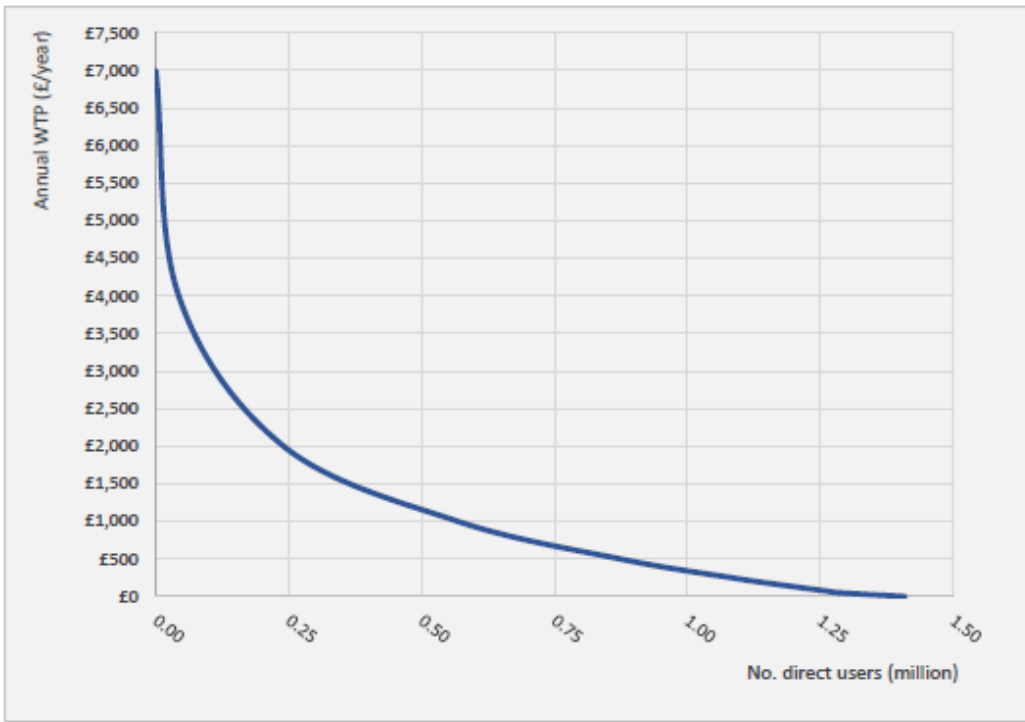


Figure 2 - Ratio between price and number of reusers for Company House Datasets

Source: Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

If one takes the study carried out by Company House and the experience of INPI as a reference, it can be argued that making companies datasets available through APIs and for free can increase the number of reusers 100 times or more. While the current number of reusers for each country and each datasets in scope is not available to the study team and while countries’ situation might vary based on the development of the data economy and reuser market, these insights give an idea of the magnitude of the increase in the number of reusers which countries can achieve by making the four datasets available as HVDs.

<sup>352</sup> Stakeholders interviews

<sup>353</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>354</sup> See for instance, The cost of Geospatial Open Data, Peter A. Johnson, Renee Sieber, Teresa Scassa, Monica Stephens, Pamela Robinson, Transaction in GIS, Wiley, January 2017, <http://onlinelibrary.wiley.com/doi/10.1111/tgis.12283/full> or See case studies on the Norwegian METNO case (meteorological data) and case study on the Dutch KNMI case (meteorological data), Study on the Pricing of Public Sector Information – POPSI Study, October 2011, Deloitte, <https://ec.europa.eu/digital-single-market/en/news/pricing-public-sector-information-study-popsis-models-supply-and-charging-public-sector>

What is also important to underline from a reuse perspective is that stakeholders from various countries all agreed that SMEs particularly benefit from easier access to company information, and that they are the group of reusers whose number increases the most. In France for instance, larger reusers saw an increase in the level of competition in their markets linked to smaller reusers entering into business thanks to the INSEE and INPI databases being made available for free and through API<sup>355</sup>. The bigger positive impact of facilitating access to information on smaller companies is also proven by the above-mentioned Company House study. This publication states that the distributional effects of charging for (Company House) information and data are likely to be disproportionate across users<sup>356</sup>. When charges apply in fact, less regular users drop out of the market and these tends to be SMEs<sup>357</sup>. Building on these evidence it seems clear that including the four datasets analysed by our study in the Directive would then open up the market to new players and allow smaller companies in particular to grow, which has direct economic benefits (described in the next section).

It must be noted here that the number of reusers of company information is actually increasing even across countries where these datasets are charged for (i.e. Italy or the Netherlands). This trend is due to the expansion of the data market and some KYC services in general but can be reinforced through public intervention. In this respect, the further opening of the four datasets would be very timely and very relevant in a market which was expanding (before COVID19) even under the previous rules.

#### Increase in the economic benefits

There is limited evidence available on the value of company data in Europe today and on the economic output linked to the reuse of such datasets. Therefore it is difficult to establish a precise baseline to identify potential benefits of the inclusion of certain datasets in the HVD list. Nonetheless, two main studies can be used as a reference to analyse the value and economic benefits of company data today:

- **The PWC study on the business information sector in Italy<sup>358</sup>**: this study quantified the value of the Italian business information sector and identified a constant growth in the last three years (from a value of 57 million euro in 2017 to 58 million euro in 2018 and 60 million euro in 2019)<sup>359</sup>. The business information sector constitutes 10% of the total Italian information sector (amounting to around 600 million euro) and the growth rate is at 4% CAGR.
- **The Company House 2019 study**: this analysis estimates that the value of company information for reusers is around 1.100 £ per year per reuser in the United Kingdom (approximately in between 0.6 billion £ and 1.7 billion £)<sup>360</sup>. The value of basic company information alone is estimated to be slightly lower (800 £ per year or in between 0.4 billion £ and 1.2 billion £)<sup>361</sup>. "In aggregate, the annual user benefits of CH data are estimated to be between £1 billion and £3 billion per year. This is likely to be an underestimate as it only includes benefits for Companies House Service (CHS) users"<sup>362</sup>.

What emerges from both these documents is that the company data hold significant economic value for reusers and constitute a considerable share of the information sector (up to 10% in Italy). Furthermore, the Italian analysis can help drawing a baseline for this country against which the effects of inclusion of the four datasets in the HVDs list can be measured.

---

<sup>355</sup> Stakeholders interviews

<sup>356</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>357</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>358</sup> Stakeholders interviews, the study was shared after the interview and is not publicly available

<sup>359</sup> Stakeholders interviews, the study was shared after the interview and is not publicly available

<sup>360</sup> The most valuable information within the datasets for reusers would be the financial information of companies.

<sup>361</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>362</sup> Valuing the user benefits of companies house data: Policy Summary, BEIS Research Paper Number 2019/015

The UK study holds actually “anticipatory value” as it can also be used to understand better what happens when company datasets are provided more freely and through an API. In fact, this analysis was completed after Company House adopted a more data friendly approach involving the removal of charges and the establishment of innovative data provision infrastructures. The study explains that **“more than half of the smaller intermediaries that access CH bulk data products have only been accessing these products since they became available free of charge.”** This suggests that access to free data has stimulated the development of new business opportunities<sup>363</sup>. If one makes a quick calculation, knowing that CH has 1.5 million users yearly out of which 1.4 million are low frequency users (which are assimilated to smaller users by the authors) and that the value of the data for each user is of 800-1.100€ per year, **it emerges that around 700 million £ (around 780 million euro) are generated yearly thanks to the publication of these datasets for free and through API.** The study further suggests that, “in the short to medium term (up to five years) it is reasonable to expect that year on year changes in aggregate benefit values will be driven by increases or decreases in user numbers rather than changes in unit benefit values<sup>364</sup>. The Company House study therefore implies that the value of company information provided for free and in an accessible way can correspond to several thousand millions of euro for an individual country. Unfortunately, this remains the only quantification of the economic effects of the data opening, as other countries which have transitioned towards open data friendly provision (i.e. France and Denmark) have not carried out similar analysis yet.

Besides generating new business opportunities as explained above, providing company and company ownership free of charge could also allow for savings. As explained by the PSI Alliance, the Ultimate Beneficial Owners initial checks in Austria cost around 17.048.520 EUR per year (repeated check would cost approximately the same). When extrapolating this figure at the EU level, UBO checks would require billions per years (according to the European Commission, there exist nearly 24 million SMEs in Europe<sup>365</sup>). These costs could be decreased dramatically by using open data as explained by the PSI Alliance. When the information included in the company register and the UBO register are available as HVDs (in a machine-readable format), the two databases could be combined in order to calculate “Standard ultimate beneficial owner” from the company register and compare it to the UBO register. If the two values display are the same, pre-completed forms can be provided to SMEs, reducing “dramatically” time and costs “to only a fraction of the initial figures<sup>366</sup>”.

Moreover, opening up company data could also have significant benefits in relation to trade credits. According to the European Central Bank statistics, trade credit is considerably important: 2.730 billion EUR in 2017 to 3.255 billion EUR in 2019. The Covid-19 crisis has clearly revealed the need for finance management in order to access to credit on a very tight timeframe, for which real-time and accurate credit information data is crucial.

Qualitative information on different aspects is available to complement the analysis of the economic benefits which would result from the inclusion of the four datasets in the list.

- *Increased market competition:* as previously discussed in this section, more availability of companies data fosters the development of a healthier and more competitive business environment and drives out unethical and undeforming companies more easily. This is also linked to the increased social benefits (see section below) and the improved capacity of companies to select their business partners, buyers and sellers due to the chance of performing better background checks. Traditional economic

---

<sup>363</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>364</sup> Valuing the user benefits of companies house data, Report 2: Direct Users, BEIS Research Paper Number 2019/015

<sup>365</sup> See : <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20181119-1>

<sup>366</sup> PSI Alliance and Compass Gruppe, Case Study regarding the release of Ultimate Beneficial Owner databases as HVDs.



literature suggests that the costs of information failure and imperfect competition within markets are very high and especially on consumers (see below).

- *Increased consumer benefits:* better availability of company datasets (all categories included) improve consumers' benefits as it decreases information asymmetries and enable consumers to make more informed choices. As explained in the section on benefits, many applications and services building on company datasets already exist to provide consumer with a better understanding of their buying choices. This allows consumers to reward what they consider positive companies' behaviours (i.e. companies having better environmental outlooks or paying their taxes in Europe and not in fiscal paradises) and sanction the opposite. Again, traditional economic literature highlights the importance of providing consumers with complete information so as the market demand and supply can better match.
- *Increased employment:* employment effects would stem indirectly from the reuse benefits and the possibility for more companies to enter the markets. As explained in the previous sections, availability of company datasets for free multiplies the number of reusers and especially SMEs thus allowing the directly creation of (qualified) jobs within the information society and the indirect creation of other types of jobs. No quantification of the specific employment impacts of open company data can be found but open data literature provides several examples concerning direct and indirect job creation effects.
- *Increased productivity and commercialisation:* there is an overwhelming consensus on the fact that making these four datasets available would allow businesses to establish new products and services and improve existing service and product portfolios. Where datasets were recently opened, i.e. in France, the business model of some existing intermediaries was disrupted by the multiplication of new services and products put on the market by entrants. While it is unclear whether some of these products and services are viable on the long term<sup>367</sup>, it is beneficial for the overall competition and for consumer to enable their establishment.
- *Increased trust and transaction costs:* by allowing better companies' check and more automated KYC and other procedure, making these datasets available would increase business trust and reduce transaction costs for all types of companies.

To conclude, it is impossible at this stage to put a number of the increased economic benefits which would derive from the inclusion of these datasets in the HVD list as many of the value drivers remain unquantifiable. Nonetheless, the experience of the frontrunner countries clearly show the exponential increase in the value of the information which emerged as a consequence of greater availability of company data and there is no reason to doubt that this anticipatory examples could not be indicative of what would happen in all other countries.

#### Increase in AI and innovation benefits

There is a strong agreement between stakeholders on the fact that the availability of the four datasets in scope of this analysis as HVD would result in direct AI and innovation benefits, although impossible to quantify. Citizens' innovation, private sector and public sector innovation would all be fostered by the greater possibilities of reuse of these datasets, in different ways.

From a private sector perspective, as already discussed in the previous two sections, providing these datasets for free and through APIs would result in increased number of reusers and increase economic output but it would also lead to more experimentation and innovation, and especially in the area of AI. Stakeholders agree that company datasets held potential for an uncountable number of AI applications and use cases, ranging from financial applications to environmental and social use cases<sup>368</sup>. The AI and

---

<sup>367</sup> Stakeholders interviews

<sup>368</sup> Stakeholders interviews

innovation potential of company data for business is just starting to be explored today and one of the reasons why some potential remains unexploited is the skewed and difficult access to data. “Without trustworthy data, applications like deep learning or machine learning (ML) will struggle to make sense of data and develop useful algorithms or insights”<sup>369</sup>. By including these datasets in the HVD list, public authorities would offer businesses the possibility to accelerate the development of AI applications in different sectors, by granting larger access to datasets which are considered as “horizontal” to many use cases and as pivotal for innovation.

For public sector authorities, greater availability and easier access (through API) to these datasets would also lead to innovation and adoption of artificial intelligence and big data analytics techniques. For instance, many countries have implemented fraud analytics mechanisms in the taxation, customs or social policy areas. For these countries, company data might constitute essential sources of information which needs to be fed to the algorithms but, according to stakeholders, tax/customs/social administrations do not always have easy access to these datasets and tend to use in-house equivalent data (i.e. companies data are extracted from customs declaration)<sup>370</sup>. This hampers the possibility of detecting trends and matching companies and individuals across systems. The example from the fraud analytics domain shows the importance for all public authorities to have easy access to, at least, basic company data and beneficial ownership information. Concerning the latter in particular, the approach taken by many Member States of allowing individual searches only (also for relevant public authorities) logically impedes the adoption of innovative data analytics and AI methods. There is a large potential for innovation within the public sector which could be untapped by inclusion of these datasets in the HVD lists.

Finally, citizens’ innovation would also be enhanced by greater availability of these datasets. Provided with more and more innovative applications and services based on companies data, citizens could greatly improve their interactions with companies and would be able to establish new forms of relationships based on less asymmetry of information.

To conclude, despite the overwhelming consensus on the fact that company datasets inclusion in the HVD list would generate enormous AI and innovation benefits, details on these possible positive effects remain quite lacking. Nonetheless, stakeholders agree on the fact that quantifying and establishing ex ante the AI and innovation potential of such datasets is impossible, as experimentation and serendipity are necessary to extract their value and reusers need to be able to “put them at work” before being able to develop new ideas and initiatives.

#### Increased social benefits

The types of company information analysed in this chapter already have a very strong potential in terms of social benefits and especially when looking at aspects such as crime fight, public engagement and understanding, and government accountability. Beneficial ownership information in particular are pivotal for liberal economies to ensure a sufficient level of transparency over business operations and also to implement anti-corruption and financial crimes rules. While these information were already made available to the competent authorities under the Anti-Money Laundering Directive 4, the revised Directive opens to any stakeholder the possibility to access these information. According to many reusers and especially NGOs and economic operators working on KYC services, this is a big step forward but yet not enough<sup>371</sup>. In fact, the Directive grants “consultation right” but does not look specifically at the reuse of these information and therefore many countries are setting up registers which are based on individual searches and do not offer API or bulk download services.

---

<sup>369</sup> <https://blog.opencorporates.com/2019/03/01/why-open-legal-entity-data-matters/>

<sup>370</sup> Stakeholders interviews

<sup>371</sup> <https://blog.opencorporates.com/2020/01/16/5aml-d-launches-close-but-no-ciqar/>

If beneficial ownership information were brought under the HVD rules, the social benefits would be immense and would translate first and foremost in greater control (from the society) on companies' activities and easier identification of companies' fraud and illicit actions. Many publications from OpenCorporates or other NGOs show how easier access to company data and beneficial ownership information in particular can help expose frauds more effectively. Furthermore, stakeholders underline how the existence of "firefly companies"<sup>372</sup> puts at risk traditional anti-money laundering and anti-corruption rules, which become too slow in such a fast change environment. Without open access to structured and up to date companies information, "many of those who use companies for illicit purposes will be able to sleep easy"<sup>373</sup>.

"Between 2007 and 2015, over \$200 billion in payments flowed through the non-resident portfolio of Danske Bank's branch in Estonia, and a report commissioned by the bank found many to be suspicious. This activity was eventually exposed by a whistleblower. Subsequently, investigators using identified many of the companies that were likely involved in the scandal. But the fact that the scheme happened for so long showed that the systems for detecting money laundering either were not working or were not able to utilise the data they needed to join the dots"<sup>374</sup>.

Given the current challenges to the enforcement of AML and anti-corruption rules, it is important to grant the civil society and other private organisations the possibility to not only access but also reuse and analyse these data. This would multiply the control power that can be exerted on bad intentioned companies and increase the trust and transparency of the overall business environment. Money laundering accounts for up to 1.2 percent of the EU's annual GDP, or around \$225.2bn (€197.2bn) in 2018, according to a 2017 report by Europol<sup>375</sup>. Providing the four analysed datasets as HVDs would allow all range of stakeholders to exert a better control on this phenomenon, thus recovering a percentage of these transactions. Finally, including the four datasets in the HVD list could help Europe progressing towards the United Nation Sustainable Development Goals as "one of the crucial requirements for sustainable development in the widest sense is making sure that we have effective, fair, transparent business markets that people can enter into and take risks"<sup>376</sup>.

While fighting companies' crime is very important from a societal perspective and even more in a post COVID19 situation where trust in the business environment must be rebuilt, citizens' trust in governments and good levels of governments' transparency and accountability are also essential and these can be supported by listing these four datasets as HDV. As already mentioned in this report, company information combined with procurement data in particular allow to make procurement processes open and transparent thus increasing citizens trust in public authorities. Globally, according to the UN's Office on Drugs and Crime and the Organisation for Economic Co-operation and Development (OECD), 10-25% of the value of public contracts is lost to corruption<sup>377</sup>. In Europe, corruption in procurement and, more generally, opaque procurement procedures cost around 5 billion euro per year according to a 2016 analysis<sup>378</sup>. Studies suggest that better availability of procurement data entail a) an increase in the number of bidders, b) an increase in the levels of competition for the bidding procedures, c) an increase in the chances for new

<sup>372</sup> <https://medium.com/@opencorporates/fireflies-and-algorithms-the-coming-explosion-of-companies-9d53cdb8738f>

<sup>373</sup> <https://blog.opencorporates.com/2020/01/16/5amlid-launches-close-but-no-cigar/>

<sup>374</sup> <https://blog.opencorporates.com/2020/01/16/5amlid-launches-close-but-no-cigar/>

<sup>375</sup> From suspicion to action - converting financial intelligence into greater operational impact, 2017, Europol, <https://www.europol.europa.eu/publications-documents/suspicion-to-action-converting-financial-intelligence-greater-operational-impact>

<sup>376</sup> <https://blog.opencorporates.com/2019/03/01/why-open-legal-entity-data-matters/>

<sup>377</sup> <https://www.weforum.org/agenda/2018/10/how-transparency-can-help-grow-the-global-economy/>

<sup>378</sup> The Cost of Non- Europe in the area of Organised Crime and Corruption, Rand Europe, 2016, [https://www.europarl.europa.eu/ReqData/etudes/STUD/2016/579319/EPRS\\_STU%282016%29579319\\_EN.pdf](https://www.europarl.europa.eu/ReqData/etudes/STUD/2016/579319/EPRS_STU%282016%29579319_EN.pdf)

players to enter the market (+8.7% of new vendors), and d) contracts value falling of 8% in average<sup>379</sup>. Although these figures refer to procurement data, practitioners insist on the fact that open procurement requires as well a good availability of company data and that “governments should facilitate the link between public procurement data and further datasets<sup>380</sup>”. For instance, “data on company ownership and the data on individuals involved in public procurement could be linked in order to measure and eventually control the risks of favouritism”<sup>381</sup>. Therefore, another social benefit of including these company datasets in the HVDs list would concern the improved governments’ transparency and accountability through improved and more open procurement processes. Once more, the COVID19 crisis exacerbates the importance of this use case from a social benefit perspective: emergency spending of governments has radically increased to face the pandemic but procurement process must remain as transparent and as controlled as possible for citizens to trust their governments<sup>382</sup>. NGOs have already looked at how companies and procurement data combined can help monitor governments’ spending during the crisis<sup>383</sup>.

Although not quantifiable, the social benefits of including these datasets in the HVDs list would as important as all other benefits and should not be underestimated in the current political context and considering the impact that the COVID19 crisis could have on business environment and public trust.

#### Increased environmental and climate change benefits

As argued by several stakeholders and by DG FISMA in particular, information contained in companies’ accounts and reports hold a very high potential value from an environmental and climate change perspective<sup>384</sup>. These documents, which are not yet available for free and through APIs in all Member States, can be used to analyse the environmental impact of companies and their sustainability outlook, thus giving stakeholders (incl. investors, civil society including NGOs, financial intermediaries, sustainable credit rating agencies, insurers, businesses, and consumers) to have sufficient and reliable information from financial and non-financial companies on their climate, environmental and social risks and impacts.

For example, consumer would have better information for choosing their product and services. Although some apps already exist building on such type of data and informing consumers about companies’ environmental actions, the lack of data availability and the lack of machine readable format for these datasets currently limits further expansion of this use case. If companies accounts and reports were to be included in the list of HVD and provided for free in machine readable format, this would lead to environmental and climate change benefits as consumers’ information and possibility of choosing “green” options would increase. Furthermore, public sector authorities would also benefit from increased availability of information concerning companies’ environmental actions, which in fact would facilitate decisions in the context of the COVID 19 recovery plans for instance<sup>385</sup>.

#### Cost-benefit comparison

When comparing the costs and benefits of making these datasets available, it emerges clearly that the benefits to society and reusers greatly exceed the costs borne by the data holders. The Cost Benefit Analysis case studies particularly support this statement. The table below illustrates the categories of

---

<sup>379</sup> Duguay, Raphael and Rauter, Thomas and Samuels, Delphine, The Impact of Open Data on Public Procurement (November 22, 2019). Available at SSRN: <https://ssrn.com/abstract=3483868> or <http://dx.doi.org/10.2139/ssrn.3483868>

<sup>380</sup> <https://digiwhist.eu/an-implementers-guide-for-open-public-procurement-data/>

<sup>381</sup> <https://digiwhist.eu/an-implementers-guide-for-open-public-procurement-data/>

<sup>382</sup> <https://www.open-contracting.org/what-is-open-contracting/covid19/>

<sup>383</sup> See examples from Colombia, Moldova, Ukraine and Paraguay on the Open Contracting website: <https://www.open-contracting.org/what-is-open-contracting/covid19/>

<sup>384</sup> Stakeholders interviews

<sup>385</sup> As mentioned earlier in this report, some countries have established some conditionalities for companies receiving funding under COVID 19 specific recovery plan. Many of these conditionalities concern environmental actions and CO2 emission reduction.

benefits identified for both Finland<sup>386</sup> and Slovenia, their weight and how do they compare with the costs analysed.

Table 25 – Benefits assessment (Finland)

Benefit components	Weight	Benefit indicators	Score	Weighed score
Economic	0.40	<ul style="list-style-type: none"> <li>Consumer benefits [X]</li> <li>Economic output [X]</li> <li>Employment [X]</li> </ul>	+3	+ 1.20
Environmental	0.05	<ul style="list-style-type: none"> <li>Environment management [X]</li> </ul>	+1	+ 0.05
Innovation & AI	0.10	<ul style="list-style-type: none"> <li>Public sector innovation [X]</li> </ul>	+2	+ 0.20
Public services and public administration	0.10	<ul style="list-style-type: none"> <li>Public sector procurement [X]</li> </ul>	+3	+ 0.30
Re-use	0.25	<ul style="list-style-type: none"> <li>Demand for information [X]</li> <li>Trust and confidence in information [X]</li> <li>Volume and range of information [X]</li> </ul>	+3	+ 0.75
Social	0.10	<ul style="list-style-type: none"> <li>Public engagement and understanding, and government accountability [X]</li> </ul>	+2	+ 0.20
<b>Aggregated benefits for Finland</b>			<b>+14</b>	<b>+2.70</b>

Table 26 – Benefits and costs (Finland)

Benefits and costs for Finland	Score
Aggregated benefits of HVD	+2.70
Aggregated costs of HVD	-1.05
<b>Overall impact</b>	<b>+1.65</b>
<b>Benefit/cost ratio</b>	<b>+2.5</b>

Table 27 – Benefits assessment (Slovenia)

Benefit components	Weight	Benefit indicators	Score	Weighed score
Economic	0.30	<ul style="list-style-type: none"> <li>Consumer benefits [X]</li> <li>Economic output [X]</li> <li>Employment [X]</li> </ul>	+3	+ 0.9
Environmental	0.05	<ul style="list-style-type: none"> <li>Environment management [X]</li> </ul>	+1	+ 0.05
Innovation & AI	0.10	<ul style="list-style-type: none"> <li>Entrepreneurialism and private sector innovation [X]</li> <li>Public sector innovation [X]</li> </ul>	+1	+ 0.1
Public services and public administration	0.10	<ul style="list-style-type: none"> <li>Public sector procurement [X]</li> </ul>	+1	+ 0.1
Re-use	0.40	<ul style="list-style-type: none"> <li>Demand for information [X]</li> <li>Trust and confidence in information [X]</li> <li>Volume and range of information [X]</li> </ul>	+3	+ 1.2
Social	0.05	<ul style="list-style-type: none"> <li>Public engagement and understanding, and</li> </ul>	+2	+ 0.1

<sup>386</sup> The data used for the CBA are the latest available and refer to the 2018 budget of the agency.

government accountability [X]		
<b>Aggregated benefits for Slovenia</b>	<b>+11</b>	<b>+2.45</b>

Table 28 – Benefits and costs assessment (Slovenia)

<b>Benefits and costs for Slovenia</b>	<b>Score</b>
Aggregated benefits of HVD X	+2.45
Aggregated costs of HVD X	-1.55
<b>Overall impact</b>	<b>+0.9</b>
<b>Benefit/cost ratio</b>	<b>+1.5</b>

As the table suggests, the overall impact of providing these datasets as HVD in Finland is positive (+1.65) and for each unit spent in providing these datasets there is a +2.5 unit of benefits generated (as shown by the cost ration). Furthermore, even if for Slovenia where the costs required for the implementation would be higher, the overall impact would remain positive (+0.9) and there would be a positive cost/benefit ratio, although smaller (for each unit spent there would be 1.5 units of benefits generated).

This conclusion is entirely in line with all available analysis concerning cost benefits of opening specific datasets, including in the company and company ownership thematic area. As such, it is not surprising to conclude that, even when costs of implementing the Directive would be relatively high on data holders and especially for a few of them, the benefits for society would be greater. This needs to be taken into account for the development of relevant policy options.

### 3.1.3 Recommended policy options

The result of the cost benefit analysis suggests that, **from a pure cost-benefit perspective and when looking at how to improve the EU data economy, all four datasets in scope of this analysis should be considered as high-value datasets under the PSI Directive.** In fact, the economic and societal benefits of such a policy choice would exceed the costs of implementation for the Member States and would bring great benefits to the data economy at the EU level. Nonetheless, our analysis also highlights that costs of implementation would be particularly concentrated on a number of countries which would then have a lower cost-benefit ratio than the others and would be more strongly affected from the changes. To limit costs and facilitate countries' transition, several options could be envisaged:

1) **Shorten the list of datasets in scope:** if a decision on shortening the list of datasets needs to be taken, our team recommends to consider first the datasets on beneficial ownership and insolvency as possible candidates to be taken out. In fact, the provision of basic company data and company documents and accounts as per our recommended measures would require more limited adaptations for the Member States and therefore lower costs. This is also partially due to the fact that these datasets are normally held and provided by the same data holders: hence, investments and loss of revenue would concern one institution per country and not several. Beneficial ownership information on the other hand are not always provided by the same institutions in charge of company basic information and documents. Furthermore, for many countries beneficial ownership registers do not exist yet (although need to be established by this year) and the timing of the implementation of the PSI Directive would come to disrupt some newly established datasets and platforms. Concretely, there is a risk that countries invested time and efforts into making their beneficial ownership registers compliant with the AML Directive (which allows charging and is conceived on an individual search basis function) for then being obliged to change their IT infrastructures

and charging models only a few months later. Concerning insolvency data, the main concern for Member States and the main cost driver would not be revenue or timing of implementation (as these datasets are already provided for free in all countries) but rather the IT infrastructure. Insolvency registers in fact are very rarely compatible with the recommended measures of this report and countries would have to invest into new portals and establishment of APIs. This could also constitute a duplication of efforts for those countries in which insolvency registers are managed by different authorities than business/company registers. For all these reasons, if shortening the list is absolutely necessary, the insolvency and beneficial ownership datasets would be the best candidate for exclusion. Of course, from a reusers perspective and when looking at the use cases, these datasets are however as important (and even more important for KYC) than company basic information and company documents and accounts.

2) **Shorten the list of data fields:** on top or besides shortening the list of datasets in scope, it could be envisaged to limit the list of data points, within each dataset (i.e. basic information, company documents and accounts etc.). The table below provides an example of how a shorter list would look like, with the deleted data points in red. This would help reducing the impact on Member States which could still sell the other data points now included in the list (i.e. annual accounts). This approach is actually very similar to what already has been done by many countries, such as Slovenia and Belgium, where an open dataset is provided including some limited data points and, alongside, paying customers get access to the full datasets. Nonetheless, this option would a) greatly limit the benefits discussed in Section 3.1.3.3, b) require further negotiation on which data points should be in or out, as countries have different preferences and c) probably be not entirely in line with the PSI Directive spirit concerning the HVD.

Table of Boxes | Micro-level Analysis of High-value Datasets and Recommended Policy Options

Basic information		Company documents and accounts	Company ownership		Company insolvency status
Non personal	Personal		Non personal	Personal	
<ul style="list-style-type: none"> <li>- Name of the company (full version; in different languages when applicable)</li> <li>- Company status (active, resolved, in liquidation, reconstruction, merger ...)</li> <li>- Founding date</li> <li>- Cessation date (if applicable);</li> <li>- Historical names</li> <li>- Addresses (i.e. legal, visiting postal)</li> <li>- Legal form</li> <li>- Identifiers (registration number / company identifier / the valid VAT identification number / phone number / e-mail address);</li> <li>- Data from VIES</li> <li>- Member State where registered</li> <li>- NACE code (of the predominant and secondary activities and the code's source)</li> <li>- <b>Number of employees</b></li> <li>- <b>Turnover</b></li> <li>- <b>Capital</b></li> <li>- <b>Detailed information on branches (including the features presented elsewhere under "basic information")</b></li> </ul>	<ul style="list-style-type: none"> <li>- (Name(s) of company legal representative(s)</li> <li>- Name of company's directors )</li> <li>- <b>The appointment, termination of office and particulars of the persons who either as a body constituted pursuant to law or as members of any such body: are authorised to represent the company in dealings with third parties and in legal proceedings; it shall be apparent from the disclosure whether the persons authorised to represent the company may do so alone or are required to act jointly; take part in the administration, supervision or control of the company</b></li> <li>- <b>All changes (to individual companies and list of companies dissolved), and date of the last update.</b></li> </ul>	<ul style="list-style-type: none"> <li>- Legal entities</li> <li>- Accounting documents, which include:                             <ul style="list-style-type: none"> <li>o consolidated financial statements (incl. the list of resident and foreign affiliates and subsidiaries, their countries, and unique identifiers),</li> <li>o non-financial statements, management reports,</li> <li>o transfer prices reports (e.g. as in the country-by country reports of BEPS Directive (2016/1164)); and</li> <li>o <b>other reports (e.g. financial reports, audit reports, corporate governance</b></li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>- Share (percentage) of ownership, and nature and extent of Beneficial Interest held (in shareholding and/or voting rights) as well as legal ownership</li> <li>- <b>Capital links between companies</b></li> <li>- <b>All changes, and date of the last update</b></li> </ul>	<ul style="list-style-type: none"> <li>- Name of the owner</li> <li>- <b>Month and Year of birth</b></li> <li>- <b>Nationality</b></li> <li>- <b>Owner identifier</b></li> <li>- <b>Names of shareholders</b></li> <li>- <b>Country of residence of the shareholders/owners</b></li> </ul>	<ul style="list-style-type: none"> <li>- Type of insolvency proceeding</li> <li>- Time limit for lodging claims;</li> <li>- Date of closing main insolvency proceedings</li> <li>- The court before which the decision opening insolvency proceedings is to be lodged</li> <li>- All changes (i.e. to individual companies and list of companies dissolved), and date of the last update</li> </ul>



<ul style="list-style-type: none"> <li>- All changes (to individual companies and list of companies dissolved), and date of the last update</li> </ul>		<ul style="list-style-type: none"> <li>reports);</li> <li>- Detailed data on branches (including the features presented elsewhere under “company documents and accounts”)</li> <li>- Intra-group transactions</li> <li>- Date of the last update of the reports;</li> <li>- Other companies documents which are provided to the authority (i.e. companies’ meeting minutes)</li> </ul>			
--	--	--	--	--	--

3) **Lower the requirements for the recommended measures for publications:** while the requirements related to APIs cannot be withheld and the provision of data for free is mandated by the Directive, it could be decided to lower the bar for some recommended measures for publications in order to contain costs. This could mean for instance, in terms of timeliness, to allow countries to provide these datasets on daily basis or when available for them *instead of real time*. Changing this recommended measures and making it “lighter” would allow a few countries to maintain the current data provision modes and simply add an API on top instead of having to rethink the way they collect and make available data completely. However, this would have a rather limited impact for the majority of countries which already work in real time or close to real time. Furthermore, for all other recommended measures for publication lowering the bar would not have significant cost reduction consequences on Member States.

Besides these options for lowering costs of implementation for Member States, there are two other important aspects which need to be taken into account by the legislator in the negotiations:

- **The definition of “companies”** and therefore the material scope of the PSI Directive (HVD rules), and more concretely which types companies should be included: on this specific point, there might be a misalignment between what reusers wish and what would be possible under current legislation. For reusers, the material scope of the PSI Directive (high-value datasets rule) for company datasets should be as large as possible: this would entail that any type of legal entity (including NGOs, trusts but also public authorities) which performs commercial activities should be covered by the Directive and countries should make available as HVD whatever they have in their different “company” registers. Unfortunately however, the situation across countries in terms of a) what each company register contains and b) how many company/legal entities registers exist varies greatly. This could make it difficult to find a formulation that on the one hand does not oblige countries to collect new data (i.e. if a country has no register of trusts) as it is not currently stipulated in the Directive, and on the other covers all datasets already available in all countries as desired by reusers. Based on the list of data fields, it should be clear where the information can be found in each of the Member States. Alternatively, as some stakeholders suggested, it would be possible to make a list of registers in scope of the PSI Directive (for the HVD rules on company and company ownership) similar to what is done under the Procurement Directive<sup>387</sup>. Nonetheless, this approach would not be consistent with the approach taken for the other high-value datasets and would also be burdensome and leave some margin of manoeuvre to Member States negotiating the inclusion or exclusion of certain registers in the list. For this reason, a more horizontal approach should be found but this would require further legal analysis of the matter.
- **The sensitivity of the disclosure of personal information for a number of Member States:** as highlighted by many data holders, strong guidance on the rules for the disclosure of basic company data of personal nature is required for Member States to feel confident about their inclusion in the high-value datasets list. In the absence of guidance from the European Commission<sup>388</sup> or the European Data Protection Supervisor, a plethora of different interpretation on the possibility to disclose directors’ names or beneficial owners information will continue to exist and this will hamper the reusability of data at the European level. A consensus on the rules applicable to personal (company) data should be found and, from this perspective, specific terms of use/license could be allowed to reassure data holders about the GDPR compliance of the PSI Directive (although this is strongly opposed by reusers).

---

<sup>387</sup> Directive 2014/24/EU of the European and of the Council of 26 February 2014 on public procurement and repealing Directive 2004/18/EC, <https://www.publicprocurement.be/sites/default/files/documents/dir201424ue-en.pdf>, see in particular Annex XI which lists the registers across Member States involved in public procurement related processes.

<sup>388</sup> It should be noted, however, that the Implementing Act might not be the most appropriate legal instrument for this purpose. This should be further assessed by the (legal) services of the European Commission.

These three recommendations have been taken into account to shape the policy options designed for the company and company ownership thematic area, and presented in the next sub-sections.

Box 1 – Validation workshop results: company and company ownership, overall appreciation of policy intervention options

During the validation workshop organised on 28 July 2020, participants were requested to indicate whether they agreed or disagreed with the three proposed policy options. The company and company ownership options received the following appreciations (86 Respondents): Agree: 71% and Disagree: 29%.

### 3.1.3.1 Lower intensity intervention

This first option is a lower intensity intervention, which implies that only the bare minimum of the datasets presented in Deliverable 2&3 would be considered as HVDs. Out of the four main datasets presented, the lower intensity intervention would only include the first two, i.e. basic information, and company documents and accounts. The reasoning behind this choice is that the data fields within these two datasets are normally held and provided by the same data holders: hence, investments and loss of revenue would concern one institution per country and not several.

Nevertheless, the list of data fields under these two datasets, i.e. basic information and company documents, would be shortened. This implies that, although all two categories would be included as HVDs, some data fields would be left out. This approach would help reducing the impact on Member States which could still sell the other data points now included in the list. This approach is actually very similar to what already has been done by many countries, such as Slovenia and Belgium, where an open dataset is provided including some limited data points and, alongside, paying customers get access to the full datasets. Nonetheless, this option would a) greatly limit the benefits, b) require further negotiation on which data points should be in or out, as countries have different preferences and c) probably be not entirely in line with the PSI Directive spirit concerning the HVD (as it would keep some data fields of high value inaccessible). Table 6 below provides a suggestion on how this option could be articulated (note: the data fields out of scope are highlighted in grey boxes).

Beneficial ownership information on the other hand are not always provided by the same institutions in charge of company basic information and documents. Furthermore, for many countries beneficial ownership registers do not exist yet (although need to be established by this year) and the timing of the implementation of the PSI Directive would coincide with the development of these datasets and platforms. Therefore, this opportunity should be seized in order to ensure that the newly established datasets and platform concerning beneficial ownership data (compliant with the AML Directive) are aligned (in terms of IT infrastructures for example) with the PSI Directive.

Concerning insolvency data, the main concern for Member States and the main cost driver would not be revenue or timing of implementation (as these datasets are already provided for free in all countries) but rather the IT infrastructure. Insolvency registers in fact are very rarely compatible with the recommended measures of this report and countries would have to invest into new portals and establishment of APIs. This could also constitute a duplication of efforts for those countries in which insolvency registers are managed by different authorities than business/company registers.

For all these reasons, the first policy option for the company and company ownership thematic area only includes basic information, company documents and accounts, and non-personal data related to company ownership, leaving out of scope company ownership personal data and company insolvency status.

This material scope of the first policy option is illustrated in the table below.

Table 29 – Company and company ownership – Scope of the low intensity intervention

Basic information		Company documents and accounts	Company ownership		Company insolvency status
Non personal	Personal		Non personal	Personal	
<ul style="list-style-type: none"> <li>- Name of the company (full version; in different languages when applicable)</li> <li>- Company status (active, resolved, in liquidation, reconstruction, merger ...)</li> <li>- Founding date</li> <li>- Cessation date (if applicable)</li> <li>- Historical names</li> <li>- Addresses (i.e. legal, visiting postal)</li> <li>- Legal form</li> <li>- Identifiers (registration number / company identifier / the valid VAT identification number / phone number / e-mail address)</li> <li>- Data from VIES</li> <li>- Member State where registered</li> <li>- NACE code (of the predominant and secondary activities and the code's</li> </ul>	<ul style="list-style-type: none"> <li>- (Name(s) of company legal representative(s);</li> <li>- Name of company's directors)</li> <li>- <i>The appointment, termination of office and particulars of the persons who either as a body constituted pursuant to law or as members of any such body are authorised to represent the company in dealings with third parties and in legal proceedings; it shall be apparent from the disclosure whether the persons authorised to represent the company may do so alone or are required to act jointly; take part in the administration, supervision or control of the company</i></li> <li>- All changes (to</li> </ul>	<ul style="list-style-type: none"> <li>- Legal entities</li> <li>- Accounting documents, which include: <ul style="list-style-type: none"> <li>o Consolidated financial statements (incl. the list of resident and foreign affiliates and subsidiaries, their countries, and unique identifiers),</li> <li>o non-financial statements,</li> <li>o management reports,</li> <li>o transfer prices reports (e.g. as in the country-by-country reports of BEPS Directive (2016/1164)); and</li> <li>o other report (e.g. financial reports, audit reports, corporate governance reports);</li> </ul> </li> <li>- Detailed data on branches (including the</li> </ul>	<ul style="list-style-type: none"> <li>- Share (percentage) of ownership, and nature and extent of Beneficial Interest held (in shareholding and/or voting rights) as well as legal ownership</li> <li>- Capital links between companies</li> <li>- <i>All changes, and date of the last update</i></li> </ul>	<ul style="list-style-type: none"> <li>- <i>Name of the owner</i></li> <li>- <i>Month and Year of birth</i></li> <li>- <i>Nationality</i></li> <li>- <i>Owner identifier</i></li> <li>- <i>Names of shareholders</i></li> <li>- <i>Country of residence of the shareholders/owners</i></li> </ul>	<ul style="list-style-type: none"> <li>- <i>Type of insolvency proceeding</i></li> <li>- <i>Time limit for lodging claims</i></li> <li>- <i>Date of closing main insolvency proceedings</i></li> <li>- <i>The court before which the decision opening insolvency proceedings is to be lodged</i></li> <li>- <i>All changes (i.e. to individual companies and list of companies dissolved), and date of the last update</i></li> </ul>

<p>source)</p> <ul style="list-style-type: none"> <li>- Number of employees</li> <li>- Turnover</li> <li>- Capital</li> <li>- Detailed information on branches (including the features presented elsewhere under "basic information")</li> <li>- All changes (to individual companies and list of companies dissolved), and date of the last update</li> </ul>	<p><i>individual companies and list of companies dissolved), and date of the last update</i></p>	<p>features presented elsewhere under "company documents and accounts")</p> <ul style="list-style-type: none"> <li>- Intra-group transactions</li> <li>- Date of the last update</li> <li>- Other companies documents which are provided to the authority (i.e. companies' meeting minutes)</li> </ul>			
--	--	--	--	--	--

It should be noted that this policy option raises the question on the sensitivity of the disclosure of personal information. From the reusers' perspective, the provision of personal data is necessary as it complements the non-personal data, and increases the utility and value of the data overall. On the other hand, as highlighted by stakeholders during interviews, the inclusion of personal data in the HVDs list is of concern for some Member States, as they fear a misuse of it as there exist serious legal and practical implications of providing (personal) data for re-use in full compliance with GDPR. Member States have different traditions and approaches to the use of personal data. Therefore, clear indication and guidance should be provided on the provision of this type of data. Besides an opinion of the European Data Protection Supervisor on the issue, an agreement on the provision of these data should be achieved at EU level. Otherwise, a plethora of different interpretation on the possibility to disclose directors' names (or beneficial owners' information) will continue to exist and this will hamper the reusability of data at the European level. To cope with this challenge, specific measures concerning the specific terms of use and license are foreseen in this policy option. These rules applicable to the provision of personal (company) data should allow to reassure data holders about the GDPR compliance of the PSI Directive.

In terms of measures for publication of company datasets, it was found that they differ to a very large extent across Member States, and despite this, users are pretty much aligned on what they need and how data should be published for them to exploit it. The discussion with reusers pointed at very clear measures for publication for the datasets in scope and, in fact, these measures have been largely adopted already by the countries which are considered as best practices in terms of company data provision (i.e. France, Ireland, Denmark and Finland in the EU, and the United Kingdom outside the EU<sup>389</sup>). This low intervention proposes the minimum measures to ensure the reusability of the company datasets in scope of this option. These measures are summarised in the table, which is followed by a more detailed description of each of them.

Table 30 – Recommended measures for publication of company datasets- Low intensity intervention

<b>Dimensions</b>		<b>Basic information (non-personal and personal)</b>	<b>Company documents and accounts</b>	<b>Company ownership (non-personal)</b>
<i>Openness-data specification</i>	<i>License (terms of use)</i>	CC-BY 4.0 (or equivalent open license) Terms of use concerning personal data and registration No database right		
	<i>Format</i>	XML		
	<i>Machine-readability</i>	Mandatory		
	<i>Availability of API, bulk download</i>	Both API and bulk download		
<i>Documentation</i>	<i>Metadata (dataset content description)</i>	Complete (*.csv document available)		

<sup>389</sup> The analysis of OpenCorporates on the accessibility of company datasets across European Countries confirms that France, Denmark, Finland and Ireland are in the top five of the ranking, together with Bulgaria, Cyprus (which could not be reached out by the study team) and Latvia (which could not be reached out by the study team), see: <https://opencorporates.files.wordpress.com/2020/06/eu-company-data-state-of-the-union.pdf>

Dimensions		Basic information (non-personal and personal)	Company documents and accounts	Company ownership (non-personal)
	<i>Data linking</i>	No specific recommendation		
	<i>Documentation (incl. structure and semantics)</i>	Complete and web-available		
	<i>Shared vocabularies</i>	Not mandatory		
	<i>Taxonomies</i>	Not mandatory		
Completeness	<i>Traceability</i>	Not necessary		
	<i>Update frequency and timeliness</i>	When available (min. weekly)		
	<i>Granularity</i>	Individual company level (plus identifier)		
	<i>Key attributes</i>	Company code for disambiguation		

As the table suggests, the same recommended modes of provision apply to the two categories of datasets in scope of the low intervention. The justifications for each of these recommended measures are the following:

- Concerning **licenses and terms of use**, the low intensity intervention would require the use of an open license, and impose some terms of use. While CC-BY 4.0 license is preferred by reusers, open national licenses could still be accepted (although this could hamper the cross border reuse of these datasets). From this perspective, CC-BY 4.0 should be used whenever possible. It is also important to highlight here that data holders should not have the possibility to claim datasets rights over the entire datasets, as this would not be coherent with an open data approach. Although an open license would be required, this low intervention suggests to impose some terms of use. These would target the use of the personal data. As indicated above, data holders have different approaches and traditions to the disclosure of this type of information. These terms of use would aim to reassure those data holders, and prevent the misuse of personal data. For example, it could be envisaged to impose terms of use in certain cases in order to forbid the reuse of Directors' names for marketing purposes (as is currently the case in Denmark).
- Format** wise, there is a strong consensus amongst stakeholders on the relevance of the XML formats whenever applicable as it is both human and machine readable and it is already in use in many countries, and not only in the open data frontrunner countries. Therefore, the low intervention would require (only, as there other formats available – see subsequent interventions) the XML formats.
- While provision of these datasets through APIs would be mandatory under the PSI Directive for HVDs, **both APIs and bulk download should be made possible** in order to foster the reuse of these datasets. This is because these two provision modes suit different use cases: APIs allow the provision



of real time services (i.e. for companies such as Altares which offer KYC services) while bulk download allows to carry out analysis (i.e. NGOs analysing business transparency in a country) and can be used to feed and train Artificial Intelligence system. Providing these datasets in API only would therefore limit the number of use cases which could build on these datasets. Therefore, the two categories of datasets within the scope of the low intervention would be available both via APIs and bulk download. Concerning the APIs, section 3.5 provides more details on horizontal considerations to be taken into account.

- Although not always provided today, **metadata (complete and in csv format) and complete documentation (web available)** are considered as absolutely necessary for the reusability of these datasets. As many countries only allow individual search of information and do not provide API or consent bulk download of the data, reusers are now used to live without them<sup>390</sup>. Nonetheless, from an HVD perspective and when API and bulk download will be the norm, the absence of metadata and full documentation could have profound consequences on the extent to which the data can be reused. Hence, already the low-intensity intervention requires the provision of accurate metadata and complete and web available documentation.
- Concerning **data linking**, no specific recommendations can be made based on the data available. This topic does not seem to be a top priority for reusers at the moment.
- **Shared vocabularies and taxonomies** would not be mandatory in this low intervention. It was often mentioned during the interviews that a lot of work is still required from the stakeholders' community to agree and progress on such standards and that it would be too much complicated for data holders to deal with this question right now. Considering the costs and efforts that would go into finding an agreement on this matter, it is suggested to recommend their use when possible without fully imposing them on data holders.
- Views on **traceability** differ: for reusers imposing traceability of data would mean adding a lot of burden to their activities, especially today when datasets are combined from countless sources and in the context of AI applications. According to some data holders however (i.e. in Italy or Slovenia), traceability is necessary to ensure that the end consumer of the data recognises the validity of the information. As practices varies in this domain and only a minority of data holders oblige reusers to trace back the data to their database, it is recommended not to impose any form of traceability and leave it up to the data holder to ask for this feature, when needed.
- From a reuser's perspective and to facilitate the reusability of the data, the question of **update, frequency and timeliness** of data is pivotal and it was made very clear that non-timely data hold no social and economic value whatsoever. Therefore the timeliness of the information is a pre-condition to qualify these datasets as high value. In fact, "we are entering a world where there will be an explosion in the number, speed and complexity of companies, brought about by companies being incorporated by computer programs... this will bring a phenomenon of what 'firefly companies' – firms that exist for mere hours, minutes, even seconds, and corporate networks that change every day, driven by programmatic company formation"<sup>391</sup>. In this sense, the lower intensity intervention suggests that data is provided when available (and not less frequent than weekly). This option therefore won't require data holders to provide the data on real time, but only when the data is available. Nevertheless, it should be noted that this flexibility would have a rather limited impact for the majority of countries which already work in real time or close to real time.
- Concerning the **granularity of the datasets**, this is not a particularly difficult topic in the area of the company data. In the scope of the company thematic area, the granularity required is at the individual company, and this is how the data is provided today. These data points refer to the identifiers listed in the first column (i.e. Basic information – Non personal) of Table 29 – Company and company

---

<sup>390</sup> A country which already today provides good metadata and documentation is Finland.

<sup>391</sup> <https://blog.opencorporates.com/2020/03/05/oecd-anti-corruption-integrity-forum-can-aml-survive-in-a-fireflies-world/>

ownership – Scope of the low intensity intervention: registration number / company identifier / tax identification number / phone number / e-mail address.

- Finally, in terms of **key attributes**<sup>392</sup>, it is very important to underline the role that identifiers play for the disambiguation of companies in this domain. For these datasets to be valuable, reusers must be able to disambiguate between homonyms. As a majority of countries provide unique identifiers and these can be used for disambiguation, it is suggested to oblige data holders to make these available as key attributes.

### 3.1.3.2 Higher intensity intervention

The second option in the company and company ownership thematic area is the higher intensity intervention.

As its name indicates, this intervention is slightly more ambitious than the previous one, both in terms of datasets to be included as HVDs, and measures for publication. In this case, the categories of datasets on company ownership (personal data) and company insolvency status would also be included in the scope of the intervention. In other words, the higher intensity intervention implies that all four categories of datasets would be included in the HVDs list.

As explained above (in the low intensity intervention), insolvency registers in fact are very rarely compatible with the recommended measures of this report and countries would have to invest into new portals and establishment of APIs. However, including company ownership and company insolvency status would unleash an added value for reusers, who could access these datasets and reuse the information (of high relevance for KYC services).

---

<sup>392</sup> A key attribute refers to the unique characteristic of the entity (e.g. the company code is an attribute of the entity company).

Table 31 – Company and company ownership – Scope of the higher intensity intervention

Basic information		Company documents and accounts	Company ownership		Company insolvency status
<i>Non personal</i>	<i>Personal</i>		<i>Non personal</i>	<i>Personal</i>	
<ul style="list-style-type: none"> <li>- Name of the company (in different languages when applicable)</li> <li>- Company status (active, resolved, in liquidation, reconstruction, merger ...)</li> <li>- Founding date</li> <li>- Cessation date (if applicable)</li> <li>- Historical names</li> <li>- Addresses (i.e. legal, visiting postal)</li> <li>- Legal form</li> <li>- Identifiers (registration number / company identifier / the valid VAT identification number / phone number / e-mail address)</li> <li>- Data from VIES</li> <li>- Member State where registered</li> <li>- NACE code (of the predominant and secondary activities and the code's</li> </ul>	<ul style="list-style-type: none"> <li>- (Name(s) of company legal representative(s)</li> <li>- Name of company's directors)</li> <li>- The appointment, termination of office and particulars of the persons who either as a body constituted pursuant to law or as members of any such body are authorised to represent the company in dealings with third parties and in legal proceedings; it shall be apparent from the disclosure whether the persons authorised to represent the company may do so alone or are required to act jointly; take part in the administration, supervision or control of the company</li> </ul>	<ul style="list-style-type: none"> <li>- Legal entities</li> <li>- Accounting documents, which include: <ul style="list-style-type: none"> <li>o consolidated financial statements (incl. the list of resident and foreign affiliates and subsidiaries, their countries, and unique identifiers)</li> <li>o non-financial statements</li> <li>o management reports</li> <li>o transfer prices reports (e.g. as in the country-by country reports of BEPS Directive (2016/1164)); and</li> <li>o other report (e.g. financial reports, audit reports, corporate governance reports)</li> </ul> </li> <li>- Detailed data on branches (including the</li> </ul>	<ul style="list-style-type: none"> <li>- Share (percentage) of ownership, and nature and extent of Beneficial Interest held (in shareholding and/or voting rights) as well as legal ownership</li> <li>- Capital links between companies</li> <li>- All changes, and date of the last update</li> </ul>	<ul style="list-style-type: none"> <li>- Name of the owner</li> <li>- Month and Year of birth</li> <li>- Nationality</li> <li>- Owner identifier</li> <li>- Names of shareholders</li> <li>- Country of residence of the shareholders/owners</li> </ul>	<ul style="list-style-type: none"> <li>- Type of insolvency proceeding</li> <li>- Time limit for lodging claims</li> <li>- Date of closing main insolvency proceedings</li> <li>- The court before which the decision opening insolvency proceedings is to be lodged</li> <li>- All changes (i.e. to individual companies and list of companies dissolved), and date of the last update</li> </ul>

<p>source)</p> <ul style="list-style-type: none"> <li>- Number of employees</li> <li>- Turnover</li> <li>- Capital</li> <li>- Detailed information on branches (including the features presented elsewhere under “basic information”)</li> <li>- All changes (to individual companies and list of companies dissolved), and date of the last update</li> </ul>	<ul style="list-style-type: none"> <li>- All changes (to individual companies and list of companies dissolved), and date of the last update</li> </ul>	<p>features presented elsewhere under “company documents and accounts”)</p> <ul style="list-style-type: none"> <li>- Intra-group transactions</li> <li>- Date of the last update</li> <li>- Other companies documents which are provided to the authority (i.e. companies’ meeting minutes)</li> </ul>			
--	--	--	--	--	--

In terms of the measures for publication, the higher intensity intervention builds on the measures previously explained for the lower intensity option. The table below provides an overview of these measures, highlighting in blue those that are new in comparison to the medium intensity intervention.

Table 32 - Recommended measures for publication of company datasets- Higher intensity intervention

Dimensions		Basic information (non-personal and personal)	Company documents and accounts	Company ownership	Company status
Openness-data specification	License (terms of use)	CC-BY 4.0 No terms of use No database right			
	Format	XML - Json			
	Machine-readability	Mandatory			
	Availability of API, bulk download	Both API and bulk download			
Documentation	Metadata (dataset content description)	Complete (*.csv document available), DCAT-AP			
	Data linking	No specific recommendation			
	Documentation (incl. structure and semantics)	Complete and web-available			
	Shared vocabularies	ISA <sup>2</sup> Core Vocabularies			
	Taxonomies	Recommended			
Completeness	Traceability	Not necessary			
	Update frequency and timeliness	Real time (minimum daily for insolvency data)			
	Granularity	Individual company level (plus identifier)	Individual owner (plus identifier)	Individual company level (plus identifier)	
	Key attributes	Company code for disambiguation	Beneficial owner code for disambiguation	Company code for disambiguation	

As displayed in the table above, the medium intensity intervention suggests the same recommended modes of provision to the four categories of datasets in scope, except for the aspects concerning the completeness of information for which some nuances must be made. The justifications for each of these recommended measures are the following:

- Following the reasoning presented for the previous interventions concerning the **license and terms of use**, open licenses would be required in this option as well. Again, CC-BY 4.0 would be the norm; however, equivalent national open licenses would not be accepted. The multiplication of non-harmonised national open license could possibly hamper the cross border reuse of these datasets, reducing the potential of these HVDs. Besides the open license, the higher intensity intervention would not impose any terms of use, as they would defeat the purpose of HVDs. Any restriction in terms of license or terms of use in fact hampers the economic and social value of the data as it is detrimental to the reuse of the information. Therefore, the high intensity would not impose any.
- **Format** wise, the higher intensity intervention would accept not only XML formats (as in the lower intensity intervention), but also the JSON format. The latter is particularly appreciated by reusers, is also already used in Ireland but does not seem to be in use in any other country at the moment.
- As in the previous intervention, these HVDs would be available via **APIs and bulk download** in the higher intensity intervention. As explained above, APIs and bulk download would foster the reuse of these datasets, covering two types of use cases (i.e. real time services for APIs, and analysis and training AI systems for bulk download). For further details on horizontal consideration on the APIs, see section 3.5.
- **Metadata (complete and in csv format) and complete documentation (web available)** are a must have in order to ensure the reusability of these datasets. As in the previous intervention, the provision of accurate metadata and complete and web available documentation would become mandatory in the medium intensity intervention. The higher intensity intervention would suggest to use the machine-readable DCAT-AP in addition to .csv. The DCAT-AP specification is only included in the higher intensity intervention as Member States do not have it in place currently.
- Concerning **data linking**, no specific recommendations can be made based on the data available. This topic does not seem to be a top priority for reusers at the moment.
- **Shared vocabularies and taxonomies** would be mandatory in this higher intensity intervention. Although agreeing on the controlled vocabularies would require some efforts, it would ensure the interoperability across Member States. Particularly, the higher intensity intervention suggests to start from using the ISA<sup>2</sup> Core Vocabularies (Business), which describes the fundamental characteristics of a legal entity (e.g. its identifier, and activities).<sup>393</sup>
- Concerning the **traceability**, the higher intensity intervention suggests to not impose any form of traceability. This would allow for some flexibility, as data holders would decide whether to ask for this feature, when deemed necessary.
- From a reuser's perspective and to facilitate the reusability of the data, the question of **update, frequency and timeliness** of data is pivotal and it was made very clear that non-timely data hold no social and economic value whatsoever. Therefore the timeliness of the information is a pre-condition to qualify these datasets as high value. In fact, "we are entering a world where there will be an explosion in the number, speed and complexity of companies, brought about by companies being incorporated by computer programs... this will bring a phenomenon of what 'firefly companies' – firms that exist for mere hours, minutes, even seconds, and corporate networks that change every day, driven by programmatic company formation"<sup>394</sup>.
- In terms of update, frequency and timeliness of data, the higher intensity intervention suggests that the data is provided in real time, meaning that changes in the database should appear to reusers as soon as validated by the data holders. There are some differences across the categories of datasets in terms of when such changes could occur though: for instance, basic company information can change at any time. There are countries in which you can open and close a company within the same day (i.e. Estonia). However, in many countries companies' accounts are only provided once a year (i.e. in

<sup>393</sup> See: <https://joinup.ec.europa.eu/release/core-business-vocabulary/100>

<sup>394</sup> <https://blog.opencorporates.com/2020/03/05/oecd-anti-corruption-integrity-forum-can-aml-survive-in-a-fireflies-world/>

Denmark it would be during the month of May) and therefore this is the period in which most of the changes to the database would occur. Changes to company ownership and insolvency datasets could also happen daily. Reusers are adamant that the datasets should be then provided in “real time” and that change logs should be made available for them to be able to only consider changes since their last retrieval of the data.

- The higher intensity intervention would require the **granularity** at the individual company (as in the lower intensity intervention), as well as at physical person level (for the company ownership – personal data points).
- Lastly, in terms of **key attributes**, this higher intensity intervention would require the use of identifiers to allow reusers to disambiguate between homonyms. These identifiers should thus be available as key attributes.

## 3.2 Geospatial

This section presents the micro-level assessment for the thematic area of geospatial data. It illustrates the current state of play of the provision of these datasets. Furthermore, it provides the recommended measures for publication together with the costs and benefits of including these datasets as high-value datasets under the PSI Directive. Lastly, it details the three policy options proposed for this thematic area.

### 3.2.1 As-is situation: how Member States provide these datasets today

There are major differences in terms of how Member States publish geospatial data. These differences were also confirmed through our desk research and stakeholders’ interviews.

When it comes to licence and terms of use, the desk research results showed that more than half of the countries (57%) use one type of open data licence for the datasets in scope. This percentage varies across the type of datasets from 75% (21 countries) in the case of administrative units to 39% (11 countries) in the case of cadastral parcels. CC BY licence seems to be the most frequent across Member States. However, there are differences between the uses of this licence for different datasets. The highest frequency of use is in administrative units (12 countries), while the lowest in addresses and cadastral parcels (eight countries). A peculiar case is the Austrian case (Tyrol<sup>395</sup>), where the licence has been slightly modified to allow the import of data into OpenStreetMap (CC Plus<sup>396</sup>). The CC0 is another type of licence used for data, especially for administrative units (five countries) and toponyms and addresses (four countries). In general, the access to the datasets is in most of the cases free of charge. However, research findings show that this is not always the case and for some datasets the licences are not open and include restrictions for commercial reuse or modification. Often the data provided under customised restrictive licences (including under the payment of fees) are the cadastral data (10 out of 28 countries analysed), buildings (nine countries) and addresses (seven countries). In some cases, the restrictions are limitations related to privacy concerns (GDPR) or to the creation of official documents. In other cases, a registration for the download is required even if the licence is open (e.g. Sweden). There are four cases (Italy, Malta, Poland, Romania) where the access to the data is in read-only mode and in some cases the reuse is regulated by the CC-BY-NC-ND (Italy and Malta).

When it comes to the availability of APIs and bulk downloads, the most frequent type of web service used is Web Feature Service (WFS), with differences across types of datasets (administrative units: 19 countries, place names: 18 countries, addresses: 15 countries, buildings: 18 countries and cadastral parcels: 13 countries). Another option is the Web Map Service (WMS) (administrative units: 16 countries, place names: 14 countries, addresses: nine countries, buildings: 14 countries and cadastral parcels: 16 countries), and REST APIs to a lower scale (administrative units: four countries, place names: five

<sup>395</sup> <https://www.data.gv.at/katalog/dataset/a0535d6d-4c34-4524-9591-e9e51e3d28c4>

<sup>396</sup> <https://wiki.creativecommons.org/wiki/CCPlus>

countries, addresses: four countries, buildings: six countries and cadastral parcels: four countries). WFS and WMS are typically and respectively used to implement INSPIRE Download and View Services. The bulk download is also an available feature across Member States (administrative units: 18 countries, place names: 15 countries, addresses: 11 countries, buildings: 14 countries and cadastral parcels: seven countries). Countries might provide only one or multiple options regarding APIs and bulk download use. For example, in the Netherlands, data is available through REST APIs, Web Feature Service (WFS), SPARQL endpoint (linked data) and bulk download. But the Netherlands seems to be a unique case. The most used web service protocols are those defined by the OGC (Open Geospatial Consortium), e.g. WMS and WFS, chosen by INSPIRE. For the bulk downloads, the most common formats are GML (administrative units: 13 countries, place names: 16 countries, addresses: 13 countries, buildings: 14 countries and cadastral parcels: 12 countries), ESRI Shapefile and MapInfo (administrative units: 15 countries, place names: 12 countries, addresses: eight countries, buildings: 11 countries and cadastral parcels: six countries). The latter two, while not open formats, are popular due to their diffusion over the years.

Regarding the documentation (incl. structure and semantics), all the revised countries use the INSPIRE documentation. The information is mostly available through the National Geoportals (where usually you can find a section called INSPIRE data where there are the sources organized with the taxonomy of the INSPIRE Annexes) by using the standards defined by INSPIRE and exposing the metadata through the CSW protocol (Catalog Service for the Web).

However, not all these characteristics are common for all the selected datasets, and an overview of the current provisions for all the datasets across Member States is provided below.



Table 33 - Current modalities for provision of geospatial data

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
Austria	Administrative units	CC BY 4.0	Y	ESRI Shapefile	Y	WFS and bulk download	INSPIRE and GeoDCAT-AP API	- documentation in the webpage and a PDF with the information written in the national official language and English
	Place Names	CC BY 4.0	Y	ESRI Shapefile	Y	WFS and bulk download	INSPIRE and GeoDCAT-AP API	- documentation in the webpage and a PDF with the information written in the national official language and English
	Addresses	Access by payment	N	CSV	Y	API	INSPIRE and GeoDCAT-AP API	- documentation in the webpage and a PDF with the information written in the national official language and English
	Buildings	CC BY 4.0	Y	ESRI Shapefile	Y	WFS and bulk download	INSPIRE and GeoDCAT-AP API	- documentation in the webpage and a PDF with the information written in the national official language and English
	Cadastral Parcels	Access by payment	N	?	?	?	INSPIRE	?
Belgium	Administrative units	No Commercial	Y	ESRI Shapefile	Y	bulk download	INSPIRE and GeoDCAT-AP API	- documentation in the webpage and a PDF with the information written in the national official language and English
	Place Names	No Commercial	Y	GML	Y	bulk download	INSPIRE and GeoDCAT-AP API	- documentation in the webpage and a PDF with the information written in the national official language and English

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
	Addresses	No Commercial	Y	Geopackage	Y	bulk download	INSPIRE and GeoDCAT-API	- documentation in the webpage and a PDF with the information written in the national official language and English
	Buildings	No Commercial	Y	ESRI Shapefile	Y	bulk download	INSPIRE and GeoDCAT-API	- documentation in the webpage and a PDF with the information written in the national official language and English
	Cadastral Parcels	restrictive license: reuse is allowed but the products have no official value without the permission of the government	Y	?	Y	WFS and WMS	INSPIRE and GeoDCAT-API	- documentation in the webpage and a PDF with the information written in the national official language and English
Bulgaria	Administrative units	No limitation	Y	-	Y	WFS - for registered users and WMS	INSPIRE	documentation under INSPIRE directive in national language
	Place Names	-	-	-	-	-	-	-
	Addresses	-	-	-	-	-	-	-
	Buildings	-	-	-	-	-	-	-
	Cadastral Parcels	personal use	N	-	Y	web services	?	- documentation on the official website in national language

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
Croatia	Administrative units	CC-BY 4.0 registration required	- Y	-	y	WFS and WMS	INSPIRE	- documentation on the official website in national language
	Place Names	CC-BY 4.0 registration required	- Y	-	y	WMS and WFS	INSPIRE	- documentation on the official website in national language
	Addresses	CC-BY 4.0 registration required	- Y	-	y	WMS and WFS	INSPIRE	- documentation on the official website in national language
	Buildings	CC-BY 4.0 registration required	- Y	-	y	WMS	INSPIRE	- documentation on the official website in national language
	Cadastral Parcels	CC-BY 4.0 registration required	- Y	-	y	WMS	INSPIRE	- documentation on the official website in national language
Cyprus	Administrative units	CC-BY 4.0	Y	GML	Y	WMS, ESRI RestAPI, WFS, Soap	INSPIRE and GeoDCAT-API	- documentation on the official website in national language and English
	Place Names	CC-BY 4.0	Y	GML	Y	WMS, ESRI RestAPI, WFS, Soap	INSPIRE and GeoDCAT-API	- documentation on the official website in national language and English
	Addresses	CC-BY 4.0	Y	GML	Y	WMS, ESRI RestAPI, WFS, Soap	INSPIRE and GeoDCAT-API	- documentation on the official website in national language and English

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
	Buildings	CC-BY 4.0	Y	GML	Y	WMS, WFS, ESRI RestAPI, Soap	INSPIRE and GeoDCAT-API	- documentation on the official website in national language and English
	Cadastral Parcels	CC-BY 4.0	Y	GML	Y	WMS, WFS, ESRI RestAPI, Soap	INSPIRE and GeoDCAT-API	- documentation on the official website in national language and English
Czechia	Administrative units	free access	Y	GML		WMS, WFS, ESRI RestAPI, Soap	INSPIRE	- documentation on the official website in national language and English
	Place Names	free access	Y	GML	Y	WMS, WFS, ESRI RestAPI, Soap	INSPIRE	- documentation on the official website in national language and English
	Addresses	free access	Y	GML	Y	WMS, WFS, ESRI RestAPI, Soap	INSPIRE	- documentation on the official website in national language and English
	Buildings	free access	Y	GML	y	WMS, WFS, ESRI RestAPI, Soap	INSPIRE	- documentation on the official website in national language and English
	Cadastral Parcels	free access	Y	GML, DGN, SHP, DXF	y	WMS, WFS, ESRI RestAPI, Soap	INSPIRE	- documentation on the official website in national language and English
Denmark	Administrative units	like CC0 with a login access	Y	ESRI Shapefile and MapInfo	y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English

State	Dataset	License (terms of use)	Free of charge	of	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
	Place Names	like CC0 with a login access	Y		ESRI Shapefile and MapInfo	y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Addresses	like CC0 with a login access	Y		GML	y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Buildings	like CC0 with a login access	Y		GML	y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Cadastral Parcels	like CC0 with a login access	Y		GML	y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
Estonia	Administrative units	national data (attribution license)	open license	Y	GML MapInfo Esri Shapefile DXF DGN	Y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Place Names	national data (attribution license)	open license	Y	GML	Y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Addresses	CCO		Y	CVS and GML	Y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Buildings	national data	open license	Y	GML	Y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English

State	Dataset	License (terms of use)		Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
		(attribution license)					download		official languages and English
	Cadastral Parcels	national data (attribution license)	open license	Y	GML MapInfo Esri Shapefile DXF DGN GeoPackage	Y	WFS and bulk download	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
Finland	Administrative units	CC-BY 4.0		Y	GML	Y	bulk download and WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Place Names	CC-BY 4.0		Y	GML	Y	bulk download and WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Addresses	CC-BY 4.0		Y	GML	Y	bulk download and WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Buildings	CC-BY 4.0		Y	GML	Y	bulk download and WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Cadastral Parcels	CC-BY 4.0		Y	GML MapInfo Esri Shapefile	Y	WFS and WMS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
France	Administrative units	national data (attribution)	open license	Y	ESRI Shapefile	Y	WFS, WMS, CartoAPI, bulk	INSPIRE	documentation in all the official national language

State	Dataset	License (terms of use)		Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
		license)					download		
	Place Names	national data (attribution license)	open license	Y	ESRI Shapefile	Y	WFS, WMS, CartoAPI, bulk download	INSPIRE	documentation in all the official national language
	Addresses	national data (attribution license)	open license	Y	ESRI Shapefile, CSV, JSON	Y	WFS, WMS, CartoAPI, bulk download	INSPIRE	documentation in all the official national language
	Buildings	End License	User	Y (for some profiles)	ESRI Shapefile	Y	WFS, WMS, CartoAPI, bulk download	INSPIRE	documentation in all the official national language
	Cadastral Parcels	End License	User	Y (for some profiles)	ESRI Shapefile	Y	WFS, WMS, CartoAPI, bulk download	INSPIRE	documentation in all the official national language
Germany	Administrative units	national data	open license	Y	GML, Shapefile	Esri Y	WFS, WMS and bulk download	INSPIRE	documentation according INSPIRE in national language and English
	Place Names	national data	open license	Y	GML, Shapefile	Esri Y	WFS, WMS and bulk download	INSPIRE	documentation according INSPIRE in national language and English
	Addresses	national data	open license or End User	Y/N	GML, Shapefile	Esri Y	WFS, WMS and bulk download	INSPIRE	documentation according INSPIRE in national language and English

State	Dataset	License (terms of use)	Free of charge	of	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
		License							
	Buildings	national data End License	open license or User	Y/N	GML, Shapefile	Esri Y	WFS, and download	INSPIRE	documentation according INSPIRE in national language and English
	Cadastral Parcels	national data End License	open license or User	Y/N	GML, Shapefile	Esri Y	WFS (registered users)	INSPIRE	documentation according INSPIRE in national language and English
Greece	Administrative units	CC-BY 3.0	Y		ESRI Shapefile	Y	bulk download, WFS and WMS	INSPIRE	documentation in national language
	Place Names	-	-	-	-	-	-	-	-
	Addresses	-	-	-	-	-	-	-	-
	Buildings	CC-BY 3.0	Y		ESRI Shapefile	Y	bulk download, WFS and WMS	INSPIRE	documentation in national language
	Cadastral Parcels	End License	User	N	?	N	N	?	documentation in national language
Hungary	Administrative units	(e) intellectual property rights	Y		GML	Y	WFS, and	INSPIRE	documentation according INSPIRE in English and national language



State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
						download		
	Place Names	(e) intellectual property rights	Y	GML	Y	WFS, and bulk download	INSPIRE	documentation according to INSPIRE in English and national language
	Addresses	-	-	-	-	-	-	-
	Buildings	open access - no derivative work	Y	GML	N	WMS	INSPIRE	documentation according to INSPIRE in English and national language
	Cadastral Parcels	commercial	N	-	N	WMS	INSPIRE	documentation according to INSPIRE in English and national language
Ireland	Administrative units	CC BY 4.0	Y	GML	Y	WFS, and bulk download	INSPIRE	documentation under INSPIRE directive in national language
	Place Names	custom license	N	GML	Y	WFS or RestAPI	INSPIRE	documentation under INSPIRE directive in national language
	Addresses	-	-	-	-	-	-	-
	Buildings	custom license	N	GML	Y	WFS or RestAPI	INSPIRE	documentation under INSPIRE directive in national language
	Cadastral Parcels	CC-BY 4.0	Y	GML	Y	WFS	INSPIRE	documentation according to INSPIRE national languages

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
Italy	Administrative units	CC-BY 4.0	Y	ESRI Shapefile	Y	WFS and bulk download	According to the INSPIRE requirements / GeoDCAT-AP API	- documentation in the webpage with the information written in the national official language
	Place Names	CC BY-SA 4.0	Y	N//A	Y	WFS	According to the INSPIRE requirements	- documentation in the webpage with the information written in the national official language
	Addresses	CC0 or CC-BY or CC-BY-SA or national open data license	Y	CSV, GeoJSON, KML, ESRI Shapefile	Y	WFS	According to the national directive for the house numbers or INSPIRE	- documentation in the webpage with the information written in the national official language
	Buildings	CC0 or CC-BY or CC-BY-SA or national open data license	Y	CSV, GeoJSON, KML, ESRI Shapefile	Y	WFS or ESRI RestAPI	According to the INSPIRE requirements / GeoDCAT-AP API	- documentation in the webpage with the information written in the national official language
	Cadastral Parcels	CC-BY-NC-ND	No, the users are free to see the map of the cadastral parcels	-	N	WMS	According to the INSPIRE requirements	- documentation in the webpage with the information written in the national official language
Latvia	Administrative	End User	Y	ESRI Shapefile	Y	WMS or bulk	INSPIRE	documentation in the webpage with the information written in the national

State	Dataset	License (terms of use)	Free of charge	of	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
	units	License			and GML		download		official language
	Place Names	End License	User	Y	ESRI Shapefile and GML		WMS or bulk download	INSPIRE	documentation in the webpage with the information written in the national official language
	Addresses	End License	User	N	ESRI Shapefile and GML	Y	WFS and bulk download	INSPIRE	documentation in the webpage with the information written in the national official language
	Buildings	End License	User	N	ESRI Shapefile and GML	Y	WFS and bulk download	INSPIRE	documentation in the webpage with the information written in the national official language
	Cadastral Parcels	End License	User	N	ESRI Shapefile	N	WMS	INSPIRE	documentation in the webpage with the information written in the national official language
Liechtenstein	Administrative units	End License	User	N	ESRI Shapefile and GML	Y	WMS	According to the INSPIRE requirements	documentation in all the official national language
	Place Names	End License	User	N	ESRI Shapefile and GML	Y	WMS	According to the INSPIRE requirements	documentation in all the official national language
	Addresses	End License	User	N	ESRI Shapefile and GML	Y	WMS	According to the INSPIRE requirements	documentation in all the official national language
	Buildings	End	User	N	CSV, INTERLIS	Y	WMS	According to the INSPIRE	documentation in all the official

State	Dataset	License (terms of use)			Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
		License				DXF , ESRI Shapefile			requirements	national language
	Cadastral Parcels	End License	User - no commercial	N		INTERLIS , DXF , ESRI Shapefile	Y	WMS	According to the INSPIRE requirements	documentation in all the official national language
Lithuania	Administrative units	End License	User - no commercial	Y		ESRI File Geodatabase , ESRI Shapefile	Y	WMS and Arcgis RestAPI	According to the INSPIRE requirements	documentation in all the official national language and English
	Place Names	End License	User - no commercial	Y		GML, ESRI Shapefile	Y	WMS and RestAPI	According to the INSPIRE requirements	documentation in all the official national language and English
	Addresses	End License	User - no commercial	Y		GML ESRI Shapefile	Y	WMS and Arcgis RestAPI	According to the INSPIRE requirements	documentation in all the official national language and English
	Buildings	End License	User - no commercial	Y		ESRI File Geodatabase , ESRI Shapefile	N	WMS and Arcgis RestAPI	According to the INSPIRE requirements	documentation in all the official national language and English
	Cadastral Parcels	End License	User - no commercial	Y		GML ESRI Shapefile	N	WMS and Arcgis RestAPI	According to the INSPIRE requirements	documentation in all the official national language and English
Luxembourg	Administrative units	CC0		Y		ESRI Shapefile, GeoJSON,	Y	bulk download and WMS	According to the INSPIRE	documentation in all the official national languages and English

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
				GML			requirements	
	Place Names	CC0	Y	ESRI Shapefile, GeoJSON, GML	Y	bulk download and WMS	According to the INSPIRE requirements	documentation in all the official national languages and English
	Addresses	CC0	Y	ESRI Shapefile, GeoJSON, CSV	Y	bulk download and geocode restAPI	According to the INSPIRE requirements	documentation in all the official national languages and English
	Buildings	CC0	Y	GeoJSON and GML	Y	bulk download and WMS	According to the INSPIRE requirements	documentation in all the official national languages and English
	Cadastral Parcels	CC0	Y	GeoJSON and GML	Y	bulk download and WMS	According to the INSPIRE requirements	documentation in all the official national languages and English
Malta	Administrative units	CC-BY-NC-ND	Y	ESRI Shapefile and GML	y	bulk download	INSPIRE	- documentation in the webpage with the information written in the national official language
	Place Names	CC-BY-NC-ND	Y	ESRI Shapefile and GML	y	bulk download	INSPIRE	- documentation in the webpage with the information written in the national official language
	Addresses	the data is not present on the national data	?	?	?	?	?	?

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)	
	sources								
	Buildings	CC-BY-NC-ND	Y	the data is available only as map	?	WMS	INSPIRE	- documentation in the webpage with the information written in the national official language	
	Cadastral Parcels	CC-BY-NC-ND	No, the users are free to see the map of the cadastral parcels	?		WMS	INSPIRE	- documentation in the webpage with the information written in the national official language	
Netherlands	Administrative units	CC-BY 4.0	Y	GML	Y	- bulk download - WFS - RestAPI - SPARQL end point	INSPIRE	- documentation in the webpage with the information written in the national official languages and English	
	Place Names	CC-BY 4.0	Y	GML	Y	- bulk download - WFS - RestAPI - SPARQL end point	INSPIRE	- documentation in the webpage with the information written in the national official languages and English	
	Addresses	CC-BY 4.0	Y	GML	Y	- bulk download - WFS - RestAPI - SPARQL end	INSPIRE	- documentation in the webpage with the information written in the national official languages and English	

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
						point		
	Buildings	CC-BY 4.0	Y	GML	Y	- bulk download - WFS - RestAPI - SPARQL end point	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Cadastral Parcels	CC-BY 4.0	Y	GML	Y	- bulk download - WFS - RestAPI - SPARQL end point	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
Poland	Administrative units	free use national act	- Y	GML, ESRI Shapefile, RDF	Y	WFS and WMS	INSPIRE	documentation in all the official national language
	Place Names	free use national act	- Y	GML, ESRI Shapefile, Excel, RDF	Y	WFS and WMS	INSPIRE	documentation in all the official national language
	Addresses	free use national act	- Y	GML, ESRI Shapefile, Excel, RDF	Y	WFS and WMS	INSPIRE	documentation in all the official national language
	Buildings	free use national act	- Y	GML	Y	WFS and WMS	INSPIRE	documentation in all the official national language

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
	Cadastral Parcels	free use to view the data	N	-	Y	WMS	INSPIRE	documentation in all the official national language
Portugal	Administrative units	CC-0	Y	ESRI Shapefile	Y	bulk download	INSPIRE	documentation in all the official national language
	Place Names	CC-0	y	GML	Y	bulk download, WFS and WMS	INSPIRE	documentation in all the official national language
	Addresses	CC-0	Y	GML	Y	bulk download	INSPIRE	documentation in all the official national language
	Buildings	CC-BY	Y	GML, ESRI Shapefile	Y	bulk download	INSPIRE	documentation in all the official national language
	Cadastral Parcels	CC-BY or CC0	Y	GML, ESRI Shapefile	Y	bulk download	INSPIRE	documentation in all the official national language
Romania	Administrative units	national open data license	Y	GeoDB, ESRI Shapefile, DXF, DWG, DGN	Y	bulk download, WFS and WMS	INSPIRE	documentation in all the official national language
	Place Names	national open data license	Y	GeoDB, ESRI Shapefile, DXF, DWG, DGN	Y	bulk download, WFS and WMS	INSPIRE	documentation in all the official national language
	Addresses	free use to view	Y	?	Y	only view	INSPIRE	documentation in all the official



State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
		the data						national language
	Buildings	national data license	open Y	GeoDB, ESRI Shapefile, DXF, DWG, DGN	Y	bulk download, WFS and WMS	INSPIRE	documentation in all the official national language
	Cadastral Parcels	free use to view the data	Y	unknown	N	WMS	INSPIRE	documentation in all the official national languages and English
Slovakia	Administrative units	unknown	unknown	unknown	?	WFS and WMS	INSPIRE	minimal documentation for the INSPIRE metadata
	Place Names	unknown	unknown	unknown	?	WFS and WMS	INSPIRE	minimal documentation for the INSPIRE metadata
	Addresses	unknown	unknown	unknown	?	WFS and WMS	INSPIRE	minimal documentation for the INSPIRE metadata
	Buildings	unknown	unknown	unknown	?	WFS and WMS	INSPIRE	minimal documentation for the INSPIRE metadata
	Cadastral Parcels	unknown	unknown	unknown	?	WFS and WMS	INSPIRE	minimal documentation for the INSPIRE metadata
Slovenia	Administrative units	CC-BY 2.5	Y	ESRI Shapefile	Available	bulk download, WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Place Names	CC-BY 2.5	Y	ESRI	Available	bulk download,	INSPIRE	documentation in the webpage with the information written in the national

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
				Shapefile		WFS		official languages and English
	Addresses	CC-BY 2.5	Y	ESRI Shapefile	Available	bulk download, WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Buildings	CC-BY 2.5	Y	ESRI Shapefile	Available	bulk download, WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Cadastral Parcels	CC-BY 2.5	Y	ESRI Shapefile	Available	bulk download, WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
Spain	Administrative units	CC-BY 4.0	Y	Available	WFS	INSPIRE	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Place Names	CC-BY 4.0	Y	Available	WFS	INSPIRE	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Addresses	CC-BY 4.0	Y	Available	WFS	INSPIRE	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Buildings	CC-BY 4.0	Y	Available	WFS	INSPIRE	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Cadastral Parcels	restrictive license: reuse is	Y	Available	WFS	INSPIRE	INSPIRE	- documentation in the webpage with the information written in the national

State	Dataset	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Documentation (incl. structure and semantics)
		allowed but the products have no official value without the permission of the government						official languages and English
Sweden	Administrative units	CC0	Y	ESRI Shapefile and MapInfo	Available	bulk download ftp registered user	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Place Names	CC0	Y	ESRI Shapefile and MapInfo	Available	bulk download ftp registered user	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Addresses	payment	N	GML	Available	WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Buildings	payment	N	GML	Available	WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English
	Cadastral Parcels	payment	N	GML	Available	WFS	INSPIRE	- documentation in the webpage with the information written in the national official languages and English

State	Dataset	Taxonomies	Traceability	Timeliness	Granularity	Key attributes
Austria	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	Annual update		
	Place Names	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update		
	Addresses	INSPIRE	National geodata Catalogue and Dedicated service	Daily		
	Buildings	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update		
	Cadastral Parcels	INSPIRE	National geodata Catalogue and Dedicated service	Daily		
Belgium	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	Annual update	boroughs municipalities districts provinces regions national sea-frontiers	borders - national identification code - identification code of the upper administrative level - official name - short name - abbreviation - name in other languages - coordinate reference system used by the national government
	Place Names	INSPIRE	National GeoCatalogue Portal	Periodical update	administrative uninhabited waterways remarkable buildings and infrastructures	units places - name - type - specific attribute
	Addresses	INSPIRE	National GeoCatalogue Portal	Daily	- national coverage	coordinates x,y house number suffix of the number name of the street zip code name of the municipality national identify code of the municipality last update

							type of position
	Buildings	INSPIRE	National Catalogue and Geodata Catalogue	Open Data National	Periodical update	- national coverage	footprint of the building destination of use (if available) height
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal		Daily	- national coverage	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs
Bulgaria	Administrative units	INSPIRE	National GeoCatalogue Portal		Periodical update	only national boundary	
	Place Names	-	-		-	-	-
	Addresses	-	-		-	-	-
	Buildings	-	-		-	-	-
	Cadastral Parcels	?	Dedicated service		Daily	national coverage	?
Croatia	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue		Periodical update	national coverage	city, municipality, settlement, delivery area of the post office, local self-government units, protected and protected areas, cadastral municipality, cadastral area at sea, statistical circle, census circle, street, square and building with corresponding house numbers
	Place Names	INSPIRE	National geodata Catalogue and open data catalogue		Periodical update	national coverage	
	Addresses	INSPIRE	National geodata Catalogue and open data catalogue		Daily	national coverage	
	Buildings	INSPIRE	National geodata Catalogue and open data catalogue		Periodical update	national coverage	

	Cadastral Parcels	INSPIRE	National geodata Catalogue and open data catalogue and dedicated service	Daily	national coverage	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs
Cyprus	Administra tive units	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update	national coverage	(a) District (b) Municipality / Community (c) Parish (d) Postal Sector which were used to record the characteristics of the population as in the 2011 Population Census.
	Place Names	INSPIRE	National geodata Catalogue and open data catalogue	When required	national coverage	geographical names and localities
	Addresses	INSPIRE	National geodata Catalogue and open data catalogue	When required	4% on the nation	according the INSPIRE directive
	Buildings	INSPIRE	National geodata Catalogue and open data catalogue	When required	national coverage	Volume number of floors type of roofs type of building footprint height
	Cadastral Parcels	INSPIRE	National geodata Catalogue and open data catalogue	When required	not completed	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs
Czechia	Administra tive units	INSPIRE	National geodata Catalogue		national coverage	
	Place Names	INSPIRE	National geodata Catalogue	Periodical update	national coverage	
	Addresses	INSPIRE	National geodata Catalogue	Periodical update	national coverage	NO geometry, entrance levels
	Buildings	INSPIRE	National geodata Catalogue	Periodical update	national coverage	NO geometry, entrance levels
	Cadastral Parcels	INSPIRE	National geodata Catalogue	Daily	national coverage	- geometry of cadastral parcels - type of particle - particle code

						- references to the administrative area to which the parcel belongs
Denmark	Administrative units	INSPIRE	National geodata Catalogue	Annual update	Administrative division Municipality Subdivision county boundaries Administrative boundary	- name
	Place Names	INSPIRE	National geodata Catalogue	Periodical update	administrative uninhabited waterways remarkable buildings and infrastructures	units places - name - type - specific attribute
	Addresses	INSPIRE	National geodata Catalogue	Daily	- national coverage	house number suffix of the number name of the street zip code name of the municipality national identify code of the municipality last update type of position
	Buildings	INSPIRE	National geodata Catalogue	Periodical update	- national coverage	footprint of the building
	Cadastral Parcels	INSPIRE	National geodata Catalogue	Daily	- national coverage	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs
Estonia	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue and dedicated service	Monthly	national coverage	all the administrative units
	Place Names		National GeoCatalogue Portal	Monthly	national coverage	
	Addresses	INSPIRE	National geodata Catalogue, open data catalogue and dedicated website	Monthly	national coverage	
	Buildings	INSPIRE	National GeoCatalogue Portal	Monthly	national coverage	
	Cadastral	INSPIRE	National geodata Catalogue	Daily	national coverage	"- geometry of cadastral parcels

	Parcels		and open data catalogue				<ul style="list-style-type: none"> <li>- type of particle</li> <li>- particle code</li> <li>- references to the administrative area to which the parcel belongs"</li> </ul>
Finland	Administrative units	INSPIRE	National geodata Catalogue	Annual update	<ul style="list-style-type: none"> <li>boroughs</li> <li>municipalities</li> <li>districts</li> <li>provinces</li> <li>regions</li> <li>national</li> <li>sea-frontiers</li> </ul>	borders	<ul style="list-style-type: none"> <li>- national identification code</li> <li>- identification code of the upper administrative level</li> <li>- official name</li> <li>- short name</li> <li>- abbreviation</li> <li>- name in other languages</li> <li>- coordinate reference system used by the national government</li> </ul>
	Place Names	INSPIRE	National geodata Catalogue	Periodical update	<ul style="list-style-type: none"> <li>administrative uninhabited waterways</li> <li>remarkable buildings and infrastructures</li> </ul>	units places	<ul style="list-style-type: none"> <li>- name</li> <li>- type</li> <li>- specific attribute</li> </ul>
	Addresses	INSPIRE	National geodata Catalogue	each 4 months	- national coverage		<ul style="list-style-type: none"> <li>latitude and longitude (wgs84)</li> <li>house number</li> <li>suffix of the number</li> <li>name of the street</li> <li>zip code</li> <li>name of the municipality</li> <li>national identify code of the municipality</li> <li>last update</li> <li>type of position</li> </ul>
	Buildings	INSPIRE	National geodata Catalogue	Periodical update	- national coverage		footprint of the building
	Cadastral Parcels	INSPIRE	National geodata Catalogue	Daily	- national coverage		<ul style="list-style-type: none"> <li>- geometry of cadastral parcels</li> <li>- type of particle</li> <li>- particle code</li> <li>- references to the administrative area to which the parcel belongs</li> </ul>
France	Administrative units	INSPIRE	National geodata Catalogue, open data catalogue and dedicated website	Monthly	national coverage		all the administrative units
	Place	INSPIRE	National geodata Catalogue,	each 4 months	national coverage		place names



	Names		open data catalogue and dedicated website			
	Addresses	INSPIRE	National geodata Catalogue, open data catalogue and dedicated website	each 4 months	national coverage	latitude and longitude (wgs84) house number suffix of the number name of the street zip code name of the municipality national identify code of the municipality last update type of position
	Buildings	INSPIRE	National geodata Catalogue, open data catalogue and dedicated website	When required	national coverage	footprint of the building heights
	Cadastral Parcels	INSPIRE	National geodata Catalogue, open data catalogue and dedicated website	When required	national coverage	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs"
Germany	Administrative units	INSPIRE	National GeoCatalogue Portal	Periodical update	national coverage	NUTS
	Place Names	INSPIRE	National GeoCatalogue Portal	Periodical update	national coverage	INSPIRE
	Addresses	INSPIRE	National GeoCatalogue Portal	variable	the availability (as open data or not) depends on each regions/cities	INSPIRE
	Buildings	INSPIRE	National GeoCatalogue Portal	variable	the availability (as open data or not) depends on each regions/cities	INSPIRE
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	variable	the availability (as open data or not) depends on each regions/cities	INSPIRE
Greece	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update	national coverage	regions, subregions and municipalities
	Place Names	-	-	-	-	-

	Addresses	-	National GeoCatalogue Portal and open data catalogue	Undefined	only for the city of Kalamaria in ccby 3.0	-
	Buildings	INSPIRE	National GeoCatalogue Portal and open data catalogue	Undefined	Buildings of the wider public administration, organized into categories according to the purpose of their use	latitude, longitude, kind of use, name
	Cadastral Parcels	?	Dedicated service	When required	national coverage	?
Hungary	Administrative units	INSPIRE	National GeoCatalogue Portal	when needed	national coverage	border, country, region, settlement, subregion
	Place Names	INSPIRE	National GeoCatalogue Portal	Periodical update	not completed	geographical name place name location name
	Addresses	-	-	-	-	-
	Buildings	INSPIRE	National GeoCatalogue Portal	Periodical update	national coverage	building footprints
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	Daily	national coverage	parcel number building cadastre land parcel cadastral map
Ireland	Administrative units	INSPIRE	National GeoCatalogue Portal	When required	national coverage	border, country, region, settlement, subregion
	Place Names	INSPIRE	National GeoCatalogue Portal	When required	national coverage	geographical name place name location name
	Addresses	-	-	-	-	-
	Buildings	INSPIRE	National GeoCatalogue Portal	Periodical update	national coverage	
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	Daily	national coverage	parcel number building cadastre land parcel cadastral map

Italy	Administrative units	INSPIRE	National Catalogue	Open Data	Annual update	Macro-regions regions provinces municipalities	- national identification code - identification code of the upper administrative level - official name - short name - abbreviation - name in other languages (if available) - coordinate reference system used by the national government
	Place Names	INSPIRE	National GeoCatalogue Portal		Periodical update	national coverage	1) Settlement names (cities, villages and individual objects) 2) Domain names 3) Mountain names (mountains, mountains and valleys) 4) Glacier names 5) Water names (rivers, streams, lakes and ponds) 6) Reed names
	Addresses	According to the national directive for the house numbers and INSPIRE	National Catalogue and Geodata Catalogue	Open Data	Periodical update	very complex situation without complete national coverage	latitude and longitude in WGS84 house number suffix of the number name of the street zip code name of the municipality national identify code of the municipality last update type of position
	Buildings	INSPIRE	National Catalogue and Geodata Catalogue	Open Data	Periodical update	- very complex situation without complete national coverage - the provincial capital cities are present	footprint of the building
	Cadastral Parcels	INSPIRE	Dedicated server		Daily	- national coverage excluded autonomous provinces	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs

Latvia	Administrative units	INSPIRE	National geodata Catalogue	Periodical update	national coverage	<ul style="list-style-type: none"> <li>• Villages</li> <li>• Land territory of Latvia</li> <li>• County</li> <li>• Parish</li> <li>• Cities</li> </ul>
	Place Names	INSPIRE	National geodata Catalogue	Periodical update	national coverage	
	Addresses	INSPIRE	National geodata Catalogue	Periodical update	national coverage	
	Buildings	INSPIRE	National geodata Catalogue	Periodical update	national coverage	
	Cadastral Parcels	INSPIRE	National geodata Catalogue		national coverage	parcel number building cadastre land parcel cadastral map
Liechtenstein	Administrative units	INSPIRE	National GeoCatalogue Portal and dedicated service	When required	national coverage	Political and administrative boundaries
	Place Names	INSPIRE	National GeoCatalogue Portal and dedicated service	When required	national coverage	INSPIRE
	Addresses	INSPIRE	National GeoCatalogue Portal and dedicated service	When required	national coverage	INSPIRE
	Buildings	INSPIRE	National GeoCatalogue Portal and dedicated service	The tracking takes place in coordination with the building permit procedure	national coverage	Each building > 6 square meters has a unique and unique identification number (GEID) and a geocoded address. House numbers with street names and property numbers as well as metric coordinates allow exact geographic localization of the buildings.
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal and dedicated service	Daily	national coverage	parcel number building cadastre land parcel cadastral map
Lithuania	Administrative units	INSPIRE	National GeoCatalogue Portal and dedicated service	When required	national coverage	Political and administrative boundaries

	Place Names	INSPIRE	National GeoCatalogue Portal and dedicated service	Annual update	national coverage	geographical name place name location name
	Addresses	INSPIRE	National GeoCatalogue Portal and dedicated service	Annual update	national coverage	minimal information
	Buildings	INSPIRE	National GeoCatalogue Portal and dedicated service	Annual update	national coverage	
	Cadastral Parcels					
Luxembourg	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	When required	national coverage	country, the districts, the cantons and the municipalities.
	Place Names	INSPIRE	National geodata Catalogue	When required	national coverage	
	Addresses	INSPIRE	National geodata Catalogue and open data catalogue	When required	national coverage	
	Buildings	INSPIRE	National geodata Catalogue	footprint, kind of use	national coverage	footprint, destination of use
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	cadastral parcels	national coverage	parcel number building cadastre land parcel cadastral map
Malta	Administrative units	INSPIRE	National geodata Catalogue	Annual update	Administrative division Municipality Subdivision county boundaries Administrative boundary	- name
	Place Names	INSPIRE	National geodata Catalogue	Periodical update	administrative uninhabited waterways remarkable buildings and infrastructures	units places - name - type - specific attribute
	Addresses	?	the data is not present on the national data sources	?	?	?
	Buildings	INSPIRE	National geodata Catalogue	Periodical update	- national coverage	footprint of the buildings
	Cadastral	INSPIRE	National geodata Catalogue	Daily	- national coverage	- geometry of cadastral parcels

	Parcels		and dedicated service				<ul style="list-style-type: none"> <li>- type of particle</li> <li>- particle code</li> <li>- references to the administrative area to which the parcel belongs</li> </ul>
Netherlands	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	Annual update		<ul style="list-style-type: none"> <li>boroughs</li> <li>municipalities</li> <li>districts</li> <li>provinces</li> <li>regions</li> <li>national borders</li> <li>sea-frontiers</li> </ul>	<ul style="list-style-type: none"> <li>- national identification code</li> <li>- identification code of the upper administrative level</li> <li>- official name</li> <li>- short name</li> <li>- abbreviation</li> <li>- name in other languages</li> <li>- coordinate reference system used by the national government</li> </ul>
	Place Names	INSPIRE	National GeoCatalogue Portal	Periodical update		<ul style="list-style-type: none"> <li>administrative uninhabited waterways</li> <li>remarkable buildings and infrastructures</li> </ul>	<ul style="list-style-type: none"> <li>- name</li> <li>- type</li> <li>- specific attribute</li> </ul>
	Addresses	INSPIRE	National Catalogue and Geodata Catalogue	Open Data National	Daily	- national coverage	<ul style="list-style-type: none"> <li>latitude and longitude (wgs84)</li> <li>house number</li> <li>suffix of the number</li> <li>name of the street</li> <li>zip code</li> <li>name of the municipality</li> <li>national identify code of the municipality</li> <li>last update</li> <li>type of position</li> </ul>
	Buildings	INSPIRE	National Catalogue and Geodata Catalogue	Open Data National	Periodical update	- national coverage	footprint of the building
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	Daily		- national coverage	<ul style="list-style-type: none"> <li>- geometry of cadastral parcels</li> <li>- type of particle</li> <li>- particle code</li> <li>- references to the administrative area to which the parcel belongs</li> </ul>
Poland	Administrative units	INSPIRE	National GeoCatalogue Portal	Periodical update		national coverage	<ul style="list-style-type: none"> <li>registration precincts;</li> <li>municipalities</li> </ul>

	Place Names	INSPIRE	National GeoCatalogue Portal	Periodical update	national coverage	
	Addresses	INSPIRE	National GeoCatalogue Portal		Periodical update	national coverage
	Buildings	INSPIRE	National GeoCatalogue Portal	Periodical update	national coverage	footprint of the building
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	Periodical update	national coverage	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs
Portugal	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	Annual update	national coverage	Administrative division Municipality Subdivision county boundaries Administrative boundary
	Place Names	INSPIRE	National geodata Catalogue and open data catalogue	continuous	national coverage	place names
	Addresses	INSPIRE	National geodata Catalogue and open data catalogue	continuous		house number suffix of the number name of the street zip code name of the municipality national identify code of the municipality
	Buildings	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update	national coverage	buildings footprint
	Cadastral Parcels	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update	national coverage	parcel number building cadastre land parcel cadastral map
Romania	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update	national coverage	Administrative division Municipality Subdivision county boundaries Administrative boundary
	Place Names	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update	national coverage	

	Addresses	INSPIRE	National geodata Catalogue and open data catalogue	Daily	national coverage	house numbers
	Buildings	INSPIRE	National geodata Catalogue and open data catalogue	Periodical update	national coverage	building footprint, destination of use
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal and dedicated service	Daily	national coverage	parcel number cadastre land parcel cadastral map
Slovakia	Administrative units	INSPIRE	National GeoCatalogue Portal	?	the data in view appears completed but not in download	minimal INSPIRE requirements
	Place Names	INSPIRE	National GeoCatalogue Portal	?	the data in view appears completed but not in download	minimal INSPIRE requirements
	Addresses	INSPIRE	National GeoCatalogue Portal	?	the data in view appears completed but not in download	minimal INSPIRE requirements
	Buildings	INSPIRE	National GeoCatalogue Portal	?	the data in view appears completed but not in download	minimal INSPIRE requirements
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	?	the data in view appears completed but not in download	minimal INSPIRE requirements
Slovenia	Administrative units	INSPIRE	National geodata Catalogue	Annual update	Administrative division Municipality Subdivision county boundaries Administrative boundary	- name
	Place Names	INSPIRE	National geodata Catalogue	Periodical update	administrative units uninhabited places waterways remarkable buildings and infrastructures	- name - type - specific attribute
	Addresses	INSPIRE	National geodata Catalogue	each 4 months	- national coverage	house number suffix of the number name of the street zip code name of the municipality national identify code of the



							municipality last update type of position
	Buildings	INSPIRE	National geodata Catalogue	Periodical update		- national coverage	footprint of the building
	Cadastral Parcels	INSPIRE	National geodata Catalogue	Daily		- national coverage	- geometry of cadastral parcels - type of particle - particle code - references to the administrative area to which the parcel belongs
Spain	Administrative units	INSPIRE	National geodata Catalogue and open data catalogue	Annual update		boroughs municipalities districts provinces regions national borders sea-frontiers	- national identification code - identification code of the upper administrative level - official name - short name - abbreviation - name in other languages - coordinate reference system used by the national government
	Place Names	INSPIRE	National GeoCatalogue Portal	Periodical update		administrative uninhabited waterways remarkable buildings and infrastructures	units places - name - type - specific attribute
	Addresses	INSPIRE	National Catalogue and Geodata Catalogue	Open Data National Daily		- national coverage	latitude and longitude (wgs84) house number suffix of the number name of the street zip code name of the municipality national identify code of the municipality last update type of position
	Buildings	INSPIRE	National Catalogue and Geodata Catalogue	Open Data National Periodical update		- national coverage	footprint of the building
	Cadastral Parcels	INSPIRE	National GeoCatalogue Portal	Daily		- national coverage	- geometry of cadastral parcels - type of particle

						<ul style="list-style-type: none"> <li>- particle code</li> <li>- references to the administrative area to which the parcel belongs</li> </ul>
Sweden	Administrative units	INSPIRE	National geodata Catalogue	Annual update	Administrative division Municipality Subdivision county boundaries Administrative boundary	<ul style="list-style-type: none"> <li>- name</li> </ul>
	Place Names	INSPIRE	National geodata Catalogue	Periodical update	administrative uninhabited waterways units places remarkable buildings and infrastructures	<ul style="list-style-type: none"> <li>- name</li> <li>- type</li> <li>- specific attribute</li> </ul>
	Addresses	INSPIRE	National geodata Catalogue	daily	- national coverage	<ul style="list-style-type: none"> <li>house number</li> <li>suffix of the number</li> <li>name of the street</li> <li>zip code</li> <li>name of the municipality</li> <li>national identify code of the municipality</li> <li>last update</li> <li>type of position</li> </ul>
	Buildings	INSPIRE	National geodata Catalogue	Periodical update	- national coverage	<ul style="list-style-type: none"> <li>footprint of the building</li> </ul>
	Cadastral Parcels	INSPIRE	National geodata Catalogue	Daily	- national coverage	<ul style="list-style-type: none"> <li>- geometry of cadastral parcels</li> <li>- type of particle</li> <li>- particle code</li> <li>- references to the administrative area to which the parcel belongs</li> </ul>

### 3.2.2 To be situation: recommended measures for publication

This section presents the recommended measures for publication, as well as the expected costs and benefits of including these datasets under the scope of the PSI Directive as HVD.

#### 3.2.2.1 Recommended measures for publication

The table below summarises the recommended measures for publication for the five categories of datasets, which have been considered in scope for this analysis. In the table we have defined the appropriate modes of provision to be applied in all countries.

Table 34 – Recommended measures for publication for geospatial data

	Description	Administrative Units	Place Names	Addresses	Buildings	Cadastral Parcels
	<i>License and terms of use</i>	CC0 / CC-BY 4.0	CC0 / CC-BY 4.0	CC0 / CC-BY 4.0	CC0 / CC-BY 4.0	CC0 / CC-BY 4.0
	<i>Format</i>	GML / GeoPackage / GeoJSON / INSPIRE recommendations	GML / GeoPackage / CSV / GeoJSON / INSPIRE recommendations	GML / GeoPackage / CSV / GeoJSON <sup>397</sup> / INSPIRE recommendations	GML / GeoPackage / GeoJSON / INSPIRE recommendations	GML / GeoPackage / GeoJSON / INSPIRE recommendations
Openness	<i>Machine-readability</i>	Recommended and already available	Recommended and already available	Recommended and already available	Recommended and already available	Recommended and already available
	<i>Availability of API, bulk download</i>	Bulk download	Bulk download	Bulk download	Bulk download	Bulk download
		WFS, WMS, CSW	WFS, WMS, CSW	WFS, WMS, CSW	WFS, WMS, CSW	WFS, WMS, CSW
		REST APIs (e.g. OGC API, ArcGIS RestAPI, Carto API)	RestAPI (e.g. OGC API ArcGIS RestAPI, Carto API)	RestAPI (e.g. OGC API ArcGIS RestAPI, Carto API)	RestAPI (e.g. OGC API ArcGIS RestAPI, Carto API)	RestAPI (e.g. OGC API ArcGIS RestAPI, Carto API)
	SPARQL end point	SPARQL end point	SPARQL end point	SPARQL end point	SPARQL end point	
Documentation	<i>Metadata (dataset content description)</i>	INSPIRE	INSPIRE	INSPIRE	INSPIRE	INSPIRE
	<i>Documentation</i>	INSPIRE / GeoDCAT-AP	INSPIRE / GeoDCAT-AP	INSPIRE / GeoDCAT-AP	INSPIRE / GeoDCAT-AP	INSPIRE / GeoDCAT-AP
	<i>Data linking</i>	To be considered	To be considered	To be considered	To be considered	To be considered
	<i>Shared vocabularies/taxonomies</i>	INSPIRE	INSPIRE	INSPIRE	INSPIRE	INSPIRE
Co	<i>Traceability</i>	National geodata Catalog and open	National geodata Catalog and open	National geodata Catalog and open	National geodata Catalog and open	National geodata Catalog and open

<sup>397</sup> Please refer to the work done in 2018-2019 within INSPIRE to define a GeoJSON encoding, with some specific transformation rules (from the original UML models) done for Addresses: <https://github.com/INSPIRE-MIF/2017.2/blob/master/GeoJSON/ads/simple-addresses.md>

	data catalog	data catalog	data catalog	data catalog	data catalog
<i>Update frequency and Timeliness</i>	Annual update	When necessary	When necessary	When necessary	When necessary
<i>Granularity</i>	boroughs municipalities districts provinces regions national borders sea-frontiers	National coverage	National coverage; Level of scale (min. 1:10000) <sup>398</sup>	National coverage; Level of scale (min. 1:10000)	National coverage; Level of scale (min. 1:10000)
<i>Key attributes</i>	National identification code; identification code of the upper administrative level; official name; short name abbreviation; names in multiple languages;	Name; name in multiple languages; category; latitude and longitude (INSPIRE)	Latitude and longitude; house number; suffix of the number; name of the street; zip code; name of the municipality; national identification code of the municipality; last update type of position	footprint of the building; height; entrances; levels; type of destination.	geometry of cadastral parcels; type of particle, particle code references to the administrative area to which the particle belongs

The table suggests that similar recommended modes of provision can be applied to the five categories of datasets in scope.

The preferred type of licence is the CC0, alternatively the CC BY 4.0 is considered. The CC0 licence (Public Domain licence) is often recommended due to its advantages as no reserved rights apply to data and works released under this type of licence. Therefore, others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law. Furthermore, according to our findings on all analysed countries, the CC0 and public domain licences are highly used (e.g. in the Netherlands). In the case of the CC BY licence, the content can be distributed, remixed, adapted, and built upon, even commercially, as long as the user credits the original creation. The recommendation of the CC-BY licences is additional to INSPIRE and highly desirable.

In terms of accuracy, a level of scale is specified. Regarding the Administrative Units and Place names, it may depend on what has to be represented. The suggestion for the other geospatial high-value datasets is to guarantee a scale of 1:10000, whether 1:5000 should not be possible. This is the scale interval that can be adapted to the needs of the use cases. For example, for the Priority Geospatial Datasets for European Commission needs, recommended scales are between 1:5000 and 1:10000. This aspect is also subject to some restrictions due to national legislations. For all spatial datasets, the recommendation is to use the formats required by INSPIRE. In general, the use of open formats, not software specific formats, is highly desirable. While the format Shapefile is currently extensively used, this format is not open format but

<sup>398</sup> The level of scale is here added.

proprietary<sup>399</sup>. For example, using formats such as GML / GeoPackage / CSV / GeoJSON increases the machine-readability of data. Allowing and providing APIs is required for all the datasets in scope. For example, the new OGC APIs, which are gradually being introduced by the Open Geospatial Consortium (OGC)<sup>400</sup>, and the use of Vector Tiles are two possible options for this. The provision of bulk download is also within the recommended options. The metadata should match the INSPIRE requirements as the transformation to geoDCAT-AP can be done automatically from this basis. It is important that the data are also displayed in geoDCAT-AP in order to integrate them automatically into national geo-portals. While the distribution through services such as Web Map Service (WMS) is useful for viewing the data, providing data in machine-readable formats, via service or via bulk download, remain highly recommended provisions.

The provision of documentation in pdf and html should be available, accompanied by at least a version in an international language (e.g. in English) if multilingual versions are not feasible. Additionally, availability of national documentation in RDF or comparable formats to be used in machine-to-machine connections is also desirable. The metadata should be published in the national (INSPIRE) discovery service. In general, the geospatial metadata is more elaborated<sup>401</sup> than the ones published by most of the open data catalogues. Therefore, to prevent loss of information, geospatial metadata should be primarily published in geospatial catalogues, from where they can be automatically harvested by open data portals. Once available on both the National INSPIRE portal and the open data portals, it can be further accessed and retrieved by all kind of portals, websites etc. (e.g. the EU INSPIRE geo-portal and the EU open data portal). The geoDCAT-AP plays an extremely important role in facilitating this harvesting process. This standard is also an important factor in the process of finding data by using dedicated search engines (e.g. dataset search of Google).

When it comes to traceability, the INSPIRE Directive fully covers this aspect as it asks to provide clear information about the sources of data published. In these conditions, the publication of the data in the national geo-catalogue and its discovery service also guarantees that the traceability requirements stated in the study is fulfilled.

Since it's not realistic to recommend the increase of the update frequency and timeliness more often than is the case on national level or the same for all Member States, the recommendation is to require availability of frequent updates in relation to the type of dataset (e.g. for the cadastre the update is daily, for the administrative units is on annual basis).

---

<sup>399</sup> The ESRI Shapefile format is a proprietary format, although some parts are publicly documented. It is a format that has been very successful becoming a de facto standard. However, it remains an old format as it is composed of several files each with the same name and with a different extension connected with a specific function (.shp for geometries, .dbf for attributes, .shx to associate attributes with geometries, .prj for identify the projection, .lyr to define the representation style ...). Not all the specifications of these formats are known (e.g. .lyr). The lack of a single file can create problems in reuse. Furthermore, the format which is used to store the attributes (.dbf) is a format with different limits (e.g. column names are made up of a maximum of 10 characters). Over time, much more robust formats have emerged that do not require this fragmentation in different files and with a open and public documentation.

<sup>400</sup> <https://ogcapi.ogc.org/>

<sup>401</sup> Please see the definition of GeoDCAT-AP provided on the INSPIRE web portal available at <https://inspire.ec.europa.eu/good-practice/geodcat-ap>  
"GeoDCAT-AP is an extension to the "DCAT application profile for European data portals" (DCAT-AP) for the representation of geographic metadata. To achieve this, GeoDCAT-AP defines transformation rules from INSPIRE / ISO metadata to DCAT-AP, which is currently the de facto standard metadata interchange format across European data catalogues. These transformation rules can be run on top of the existing INSPIRE infrastructure, without requiring any internal modification, and can be directly used to expose metadata available via a CSW (Catalog Service for the Web) in GeoDCAT-AP."

### 3.2.2.2 Expected costs

While formulating the recommendations on the modes of provision of the datasets in scope to be made available as HVD, the debate on the expected costs that Member States have to face when adapting to the PSI HVD conditions has been controversial. The table of costs we adopted to better explain the types of costs that data holders bear today for providing the datasets, has been used to get more information on how much the request of opening up some datasets under HVD conditions is sustainable. Particular attention should be focused on cadastral parcels, for example. The business model of the cadastral agencies is based on the revenue coming from the sale of the data. The stakeholders encountered some difficulties in providing figures on the specific costs. However, insights from the interviews suggest that the highest costs concern the frequency of update, accuracy, and scale.

The table below aims to provide a comprehensive overview of the main cost drivers for the provision of data.

Table 35 – Expected cost categories for geospatial data

Cost category	Description	Insights from the data collection
Infrastructural costs	Costs related to infrastructural investments such as portals, APIs, Servers (cloud), etc.	Infrastructural costs vary across countries depending on the existing IT infrastructure, but also on the country's size. In this case, the total costs for the national geo-portals depend on the responsible data holder, which is not the same in all the revised countries. It's difficult to separate these costs from the general IT costs of the organisation as a whole. In this case, many of the stakeholders mentioned that precise figures couldn't be provided only for particular datasets provisions.
Data transformation costs	Costs related to data processing including data cleaning, preparation of metadata, aggregation, anonymisation, etc.	The data transformation costs are difficult to quantify, as data transformation costs are affected by the nature of the dataset and data holders consider the related activities as part of their usual work. When the specific open data initiatives are already in place, the amounts might be available, but these costs cover more than just the datasets in scope of the analysis.
Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	The operational costs are rarely made explicit. The cost for the personnel could vary. It might be possible to have more details on the extra employees needed but frequently these types of activities are already part of the current workload of the permanent employees.
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	No precise figures could be collected on the other costs.

To understand how this cost structure would be affected by the implementation of the Directive one then needs to look at what could be (in general) the main budgetary implications of the recommended measures for publication those countries which are not yet close to them, as shown in the table below.

Table 36 – Main budgetary implication of the recommended measures for publication

Recommended dimension for publication	Budgetary implication (little to none, low, medium, high)
License and terms of use: CC BY/CC0	Little to none.

Format: adding an open format	Low to Medium. Adapting the data provision to recommended formats would require a minimum level of investments for the countries that do not yet provide an open format.
Modes of provision: both API and bulk download	Medium to High.
Metadata and documentation: recommended	Low.
Data linking: to be considered	Low to Medium.
Shared vocabularies and taxonomies: recommended	Low to Medium.
Traceability: recommended	Medium.
Update frequency and timeliness: annual, quarterly or monthly based on the dataset's specificity	Medium.
Granularity: national or regional level <b>(depending on the dataset)</b>	Low to Medium.
Key attributes	Low.

As the table suggests, a number of recommended measures would have different level on impact on the budgets of data holders due to the fact that countries already provide the datasets in the recommended modality and do not need to adapt or the adjustments are rather small and does not involve high costs. The recommendations that might have budgetary implications for countries are mostly related to the mode of provision (API and bulk download), metadata provision (where needed), and data and metadata versioning. However, often the changes and adjustments might refer to broader activity in datasets production than to only a few ones. This exercise could be more valuable if we consider costs per single dataset.

Table 37 – Magnitude of costs

Cost components	Cost components description	Magnitude of costs (range)
Infrastructural costs	Establishment of the API and bulk download, adaptation of the IT infrastructure to real time provision	Initial investment (one time only) depending on the solution, in between 250 000 and 3 000 000 EUR.  For further developments, depending on the country and the size, an example is the costs for data storage device: 450 000 EUR (once off)
Data transformation costs	Costs related to data processing including data cleaning, preparation of metadata, aggregation, anonymisation, etc.	In between 100 000 and 200 000 EUR (yearly)
Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	In between 150 000 and 350 000 EUR (yearly)
(Lost) income for data supplier	(Share of) revenue related to the provision of the HVD	Depending on country
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	N/A
Negative impact on competition	The estimated impact of competition distortion vis-à-vis private organisations	N/A

---

active in the domain.

---

Relying on the available information and on the previously mentioned general analysis of costs, we attempted to estimate the size of expenses (on a scale from low costs to very high costs) for all the Member States for which the pertinent data was accessible. The result of this evaluation is provided below.



Country	Cost impacts					Comments
	Low	Medium	High	Very high	Unknown	
Austria	x					Written feedback provided.
Belgium	x					The datasets in scope have several different producers, from federal and regional levels. There are several portals in Belgium: four geo-portals, two open data portals. The costs provided are too aggregated. We couldn't get a real estimation of the costs. The cost for the geo-portal of Flanders is around 200 000/yearly (infrastructural costs).
Bulgaria					x	N/A
Croatia					x	N/A
Cyprus					x	The estimated costs for the provision of the HVD under the PSI conditions have not been provided.
Czech Republic					x	Feedback provided but no specific information on costs is available.
Denmark	x					Denmark provides the datasets as recommended by this report. The costs will concern mainly the system's update and maintenance.
Estonia	x					Estonia provides all the datasets in scope as open data. A current estimate of costs is not provided, but mainly concerns data collecting and updating costs, infrastructure costs, distributing costs, archiving costs. A minor increase of costs will be related to the translation to English and publishing.
Finland	x					Finland provides the datasets as recommended by this report. The costs will concern mainly the system's update and maintenance.
France					x	No specific figures provided.
Germany	x					Not significant figures on the extra costs from the BKG (The Federal Agency for Cartography and Geodesy), which is responsible for the development and operation of the GeoPortal.Bund. BKG cooperates closely with the Bund and Länder administrations concerning management and further extension.
Greece					x	Written feedback provided. This specific information is not available.
Hungary					x	N/A
Ireland		x				Feedback provided.
Italy	x					The organization and its infrastructure is already enabled for HVD open geospatial data distribution. No additional costs or procedures are needed.

Latvia			x	N/A	
Lithuania			x	No extra costs provided. They involve different categories of costs, and it's difficult to get them separated by datasets (many datasets live in one information system) and accumulate over time. State Enterprise Centre of Registers would lose income from selling data. Financial flows would change, and perhaps also requirements for the datasets (frequency of update, quality).	
Luxembourg	x			x	<i>Geospatial Datasets in scope already provided as open data.</i>
Malta		x			<i>There's a plan to open the datasets in scope, not yet implemented.</i>
The Netherlands	x				Data is provided up to standard for what is required within government (including INSPIRE requirements). There is no additional data cleaning. Costs for provision of datasets, including INSPIRE compliant services, are approx. € 15M/year.  The requirement to provide data through APIs will incur higher costs compared to providing data via downloads. They are currently considering the introduction of a 'high volume user' fee in cases where APIs (including OGC-services) are used in on-line applications.
Poland				x	<i>No specific data on cost impacts provided.</i>
Portugal		x			No estimates on extra costs have been provided, however the poor national coverage could affect the cost impacts.
Romania				x	N/A
Slovakia				x	<i>No specific information on cost impacts has been provided.</i>
Slovenia				x	No answer provided on cost impacts.
Spain	x				No extra costs provided, although Spain is trying to estimate the economic value of their web services comparing with other cartography agencies and private companies who are charging for similar web services.
Sweden	x				No extra costs since Sweden meets all the requirements. But the business model and the way the Authority is financed would have to change. For example, the Lantmäteriet's yearly income, which is approximately 90 million SEK today, would be zero if these datasets will be made available under HVD conditions.

We have looked at the analysis of costs for two particular countries (Italy and Sweden), which were selected for the CBA analysis.

For assessment of the benefits and using the framework developed with the study, we assessed the expected benefits through the indicators included in different macro-economic areas, considering also our desk research results and the inputs received from different stakeholders. Within the analysis we will refer to the overall table below as then reference for the benefits related to the four datasets together (for both use cases considered).

Benefit components	Weight	Benefit indicators	Score	Weighed score
Economic	0.3	<ul style="list-style-type: none"> <li>· Competition [X]</li> <li>· consumer benefits [X]</li> <li>· Economic output [X]</li> <li>· Employment [X]</li> <li>· Productivity and commercialisation [X]</li> </ul>	3	0.9
Environmental	0.3	<ul style="list-style-type: none"> <li>· Environment management [X]</li> </ul>	3	0.9
Innovation & AI	0.05	<ul style="list-style-type: none"> <li>· Citizen innovation [X]</li> <li>· Public sector innovation [X]</li> <li>· Entrepreneurialism and private sector innovation [X]</li> <li>· AI [X]</li> </ul>	2	0.1
Public services and public administration	0.25	<ul style="list-style-type: none"> <li>· Public services management [X]</li> </ul>	2	0.5
Re-use	0.05	<ul style="list-style-type: none"> <li>· Demand for information [X]</li> </ul>	2	0.1
Social	0.05	<ul style="list-style-type: none"> <li>· Disease prediction and prevention [X]</li> <li>· Mobility efficiency [X]</li> <li>· Mobility planning</li> </ul>	3	0.15
<b>Aggregated benefits of HVD geospatial</b>			<b>15</b>	<b>2.65</b>

In case of Italy, the organization and its infrastructure are already enabled for HVD open geospatial data distribution. Therefore, there are not envisaged further additional costs for compliance with the recommendations. However, based on the information provided by the stakeholders regarding current and possible future costs for the organization, we developed the cost impact in the table below.

Cost components	Weight	Cost indicators	Score	Weighed score
Infrastructural costs	0.3	Variation* of the Sum of: <ul style="list-style-type: none"> <li>· Portals</li> <li>· APIs</li> <li>· Servers (cloud)</li> </ul> = 25% of PSB budget	-3	-0.9
Data transformation costs	0.2	Variation* of the Sum of: <ul style="list-style-type: none"> <li>· Data cleaning</li> <li>· Preparation of metadata</li> <li>· Aggregation</li> <li>· Anonymisation</li> </ul> = -6.25% of PSB budget	0	0
Operational costs	0.15	Variation* of the Sum of: <ul style="list-style-type: none"> <li>· Updating data</li> <li>· Replying to users' requests</li> <li>· Correcting mistakes</li> </ul> = -25% of PSB budget	0	0
Lost income for data	0.25	Not available	0	0

supplier				
Other costs	0.05	Not applicable	0	0
Negative impact on competition	0.05	Not applicable	0	0
<b>Aggregated costs of HVD geospatial</b>			<b>-3</b>	<b>-0.9</b>

From the data, it results that the possible highest impact will be in the infrastructure area, where a significant investment might be required. However, this is an only-once type of investments for the organization. In addition, the other types of costs are assumed to be lower than the current costs for the organisation. It is worth mentioning that the costs provided refer to the geospatial data as a whole, and it is not database specific. Considering the different ways different datasets are currently available, for certain data the provision under HVD requirements might result in a higher impact (e.g. cadastral parcels) compared to other of data (e.g. administrative units).

<b>Benefits and costs</b>	<b>Score</b>
Aggregated benefits of HVD geospatial	2.65
Aggregated costs of HVD geospatial	-0.9
<b>Overall impact</b>	<b>1.75</b>
<b>Benefit/cost ratio</b>	<b>2.94</b>

The results above show that the cost/benefit balance for Italy is positive, with an overall impact of +1.75. In the same time, the cost/benefit ratio shows that for each unit spent in providing these datasets there is an increase of +2.94 units of benefits generated. However, these results should be considered carefully, as they are more linked to the aggregate effect of all datasets in scope rather than for each on specific one.

Another particular use case is Sweden, where the impact of providing the datasets in scope as high-value datasets will impact on the business model of the organisation rather than the costs for data provision. From the input provided by the stakeholders, the organisation will lose approximately 90 million SEK yearly from changing the data provisions (see table below).

<b>Cost components</b>	<b>Weight</b>	<b>Cost indicators</b>	<b>Score</b>	<b>Weighed score</b>
Infrastructural costs	0.3	Variation* of the Sum of: · Portals · APIs · Servers (cloud) = not significant	0	0
Data transformation costs	0.2	Variation* of the Sum of: · Data cleaning · Preparation of metadata · Aggregation · Anonymisation = not significant	0	0
Operational costs	0.15	Variation* of the Sum of: · Updating data · Replying to users' requests · Correcting mistakes = not significant	0	0

Lost income for data supplier	0.25	Variation* of the (share of) revenues related to the provision of the HVD = 16.3% of PSB budget	-2	-0.5
Other costs	0.05	Not applicable	0	0
Negative impact on competition	0.05	Not applicable	0	0
<b>Aggregated costs of HVD geospatial</b>			<b>-2</b>	<b>-0.5</b>

<b>Benefits and costs</b>	<b>Score</b>
Aggregated benefits of HVD Geospatial	2.65
Aggregated costs of HVD Geospatial	-0.5
<b>Overall impact</b>	<b>2.15</b>
<b>Benefit/cost ratio</b>	<b>5.30</b>

In this case, the cost/benefit analysis shows a higher positive result. The overall balance is also positive (+2.15) as in Italy case, and the benefit/cost ratio shows is even higher. The results show that for each unit spent in providing these datasets, there are +5.3 units of benefits generated. But, as in the previous case, these results should be considered carefully, as they are more linked to the aggregate effect of all datasets in scope rather than for each on specific one.

The two cases show that the benefits of opening up the geospatial data override the costs implied by the full compliance process. It is worth mentioning that the lack of more detailed information doesn't allow a more detailed analysis by type of datasets, as due to the current provision some significant differences might arise when different conditions apply.

### 3.2.3 Recommended policy options

There is wide consensus that geospatial data are those that offer the widest set of opportunities for reuse, namely because of their combinability with other datasets. The number of use cases is almost infinite, and the recent Covid-19 crisis only served to reinforce the strategic importance of wide availability of high quality georeferenced data to track the epidemics and control the reopening.

Because of their wide reusability, it is very difficult to select the highest value datasets. The analysis points out that the selection of datasets presented is instrumental, as these datasets can be considered as the tip of the iceberg, the initial effort to open up which can initiate a virtuous cycle that fosters organisational and cultural change. All domains in Annex I of the INSPIRE directive should be ultimately available as open data.

For the same reason, there are many overlaps between geospatial and other domains. Administrative units are also addressed by Statistics; transport network falls under the Mobility domain; orthophotos are part of Earth observation and environment. It is necessary to consider the cross-domain nature of datasets.

It is also clear that many countries already offer the data free of charge. Interviewed stakeholders confirm the strong demand for geospatial data also from industry players from the country, hence strengthening the economic argument for opening up.

The estimated change and additional cost of opening up is, broadly speaking, considered feasible. The main costs lie in the update of the technological infrastructure to offer API access, the cost in ensuring data quality, and the loss in revenues for some specific categories. In those countries where data are not free, this loss of revenue is much higher than the technological investment to be expected. Critical cases are represented by cadastral parcels, buildings and addresses (in five countries on the total number of the revised countries), which generate a market with a very significant volume. However, the available evidence clearly points out that the economic benefits from reuse are of an order of magnitude greater than the revenue loss.

When it comes to the modalities of publication, any choice should be in line with the INSPIRE directive. The preliminary steps to be undertaken by the Member states regard the release of the first two above-mentioned datasets, at least Administrative Units and Place Names, under the CC0 licence. The webservices and API development should follow the next INSPIRE recommendations as much as possible<sup>402</sup> (e.g. the use of the OGC-API). The implementation of geDCAT-AP is crucial for the datasets that can be easily harvested on the national open data portals and indexed by search engines. Opening up the INSPIRE attributes, and provision of the documentation in English as well are two key actionable items that should be part of the recommendations. A good example is related to the house numbers and its georeferencing: small municipalities often don't cover all the house numbers and it's due to the lack of skills, time and resources.

The next sub-sections present the different policy options designed for the geospatial thematic area. For this thematic area, the development of the policy options concern both the number of datasets to be included as high-value datasets, and the set of the measures used for the publication of the datasets (e.g. the formats available, the licences and terms of use, the key attributes, and the level of scale (granularity)). The approaches are the result of the monitoring of the geospatial datasets in scope, available across all Member States. The five datasets in scope to be considered as HVDs are the following: Administrative Units, Place Names, Addresses, Buildings, Cadastral Parcels. For all spatial datasets, the recommendation is to follow the requirements defined in the INSPIRE directive (vocabularies, metadata, formats, interoperability services, etc.). The key attributes related to these HVDs vary across countries, the recommendations fit the current and future feasibility.

Box 2 – Validation workshop results: geospatial, overall appreciation of policy intervention options

During the validation workshop organised on 28 July 2020, participants were requested to indicate whether they agreed or disagreed with the three proposed policy options. The geospatial thematic area options received the following appreciations (65 Respondents): Agree: 94% and Disagree: 6%.

<sup>402</sup> The geospatial data storage and distribution services required by INSPIRE offer all the mandatory webservices for the directive and several other forms of distribution. The most popular products are ESRI ArcGIS (proprietary) and Geoserver (open source). Both offer Rest API services (already included or available as extensions). The ESRI product offers its own Rest API, widespread and integrated also in ESRI's services as client. Geoserver offers the OGC-API extension which follows the specifications of the Open Geospatial Consortium. Therefore, in the creation of the geospatial services required by the INSPIRE directive, the products acquired are also able to provide Rest API. A case in itself is that of France which has developed its own service, released in open source mode, called Carto-API.

### **3.2.3.1 Lower intensity intervention**

This lower intensity intervention implies that only limited requirements of data and minimum changes of the current available publication options will apply for datasets considered as HVDs. When it comes to the geospatial thematic area, the number of datasets in scope as HVDs are five to four, including Administrative units, Place Names, Addresses, Buildings, Cadastral Parcels. Based on the desk research and the data holders' interviews, the Cadastral Parcels dataset appears to be not free of charge in most of the Member States, revealing the complexity affecting its release under minimum HVD requirements. The lower-intensity intervention proposes a set of minimum adjustments' measures to ensure the reusability of the geospatial datasets in scope. The measures are summarised in the table, which is followed by a more detailed description of each of them.

Table 38 – Geospatial - Scope of the lower intensity intervention

	<b>Description</b>	<b>Administrative units</b>	<b>Place Names</b>	<b>Addresses</b>	<b>Buildings</b>	<b>Cadastral parcels</b>
<b>Openness</b>	<b>License and terms of use</b>	CC-BY 4.0				
	<b>Format</b>	GeoPackage; GeoJSON; INSPIRE requirements.	GeoPackage; CSV; GeoJSON; INSPIRE requirements.	GeoPackage; CSV; GeoJSON; INSPIRE requirements.	GeoPackage; GeoJSON; INSPIRE requirements.	GeoPackage; GeoJSON; INSPIRE requirements.
	<b>Machine-readability</b>	Mandatory				
	<b>Availability of API, bulk download</b>	Bulk download; INSPIRE distribution services; RestAPI (e.g. OGC API, ArcGIS RestAPI, Carto API).	Read-only mode (WMS service defined by INSPIRE).			
<b>Documentation</b>	<b>Metadata (dataset content description)</b>	INSPIRE				
	<b>Documentation (incl. structure and semantics)</b>	INSPIRE; GeoDCAT-AP.				
	<b>Data linking</b>					
	<b>Shared vocabularies/taxonomies</b>	INSPIRE				
<b>Completeness</b>	<b>Traceability</b>	National Geodata Catalog and/or open data catalog				
	<b>Update frequency and timeliness</b>	Annual update	When necessary	When necessary	When necessary	When necessary
	<b>Granularity</b>	From municipalities to countries.	National coverage.	Partial National coverage (e.g. most populated cities).	Partial National Coverage (e.g. most populated cities); Level of scale 1:5000.	National coverage; Level of scale 1:5000.
	<b>Key attributes</b>	National identification code; identification code of the upper administrative level; official name; country code; name in multiple languages (only for countries with more than one official	Name; name in multiple languages (only for countries with more than one official language); category; latitude and longitude (INSPIRE)	Latitude and longitude; house number; suffix of the number; name of the street; name of the municipality; national identification	Footprint of the building; entrances; floors; type of use.	Geometry of cadastral parcels; type of particle; particle code; references to the administrative area to which the particle belongs.



---

language).

code of the  
municipality;  
last update.

---

As the table suggests, the some of the recommended modes of provision apply to all datasets in scope, while others are more specific (e.g. update/timeliness, granularity, key attributes). The justifications for each of these recommended measures are the following:

- Concerning **licences and terms of use**, the recommended licence for the lower intensity intervention is the Creative Commons Attribution 4.0 International - CC-BY 4.0. This recommended option will have no impact since this type of licences is widely used across Member States (MS). In the same time, this type of licence is also preferred by re-users. In the case of the Member States that have already implemented less restrictive licences (for example the CCO) or similar ones, this recommendation has not to be considered a requirement.
- **APIs and bulk download**: Regarding the cadastral parcels - the recommendation is to ensure at least access through WMS service as defined by INSPIRE, in read-only mode, due to the difficulties in the release of this dataset. The WMS remains a data product which allows the creation of new georeferenced data and the view of the geo-data as a whole. The WMS protocol provides feature information (as XML) by identifying a point on a map. For the other datasets, the download through API, bulk download, OGC services listed by INSPIRE should be guaranteed.
- When it comes to **formats**, an option for the publication of datasets is to follow INSPIRE requirements. According to the stakeholders' suggestions, the recommendation is to use Geopackage and GeoJSON, which are two relatively recent open and low-cost formats. GeoJSON is commonly used, the re-users have shown a strong preference for new open standards like Geopackage. Using these formats increases the machine-readability of the data. The proposal regarding the use of dedicated services for automatic conversion is added.
- **Granularity**: concerning Addresses and Buildings datasets the recommendation is to guarantee at least a partial coverage (e.g. most populated cities). Based on our interviews, the findings already show a diversified situation across the MS and main issues are related to the data ownership at local level, the costs of management and the frequency of update that strongly affect the lack of a full national coverage. However, the full national coverage is recommended for the rest of the datasets. The level of scale to be guaranteed for the Buildings and Cadastral parcels is 1:5000 or better. The increase of detail in the level of scale corresponds to a raise in the costs of the dataset production.
- The **metadata** should match the INSPIRE requirements as the transformation to geoDCAT-AP can be done automatically from this basis. It is important that the data are also displayed in geoDCAT-AP in order to integrate them automatically into national open data portals.
- When it comes to **key attributes**, the values identified correspond to the common characteristics available across the majority of the Member States, for each type of dataset in scope. This is due to the implementation of the INSPIRE directive. Therefore, the impact on the organisations will remain rather low and limited. The interviewed stakeholders highlighted the relevance of having names in multiple languages as a key attribute, because it significantly improves the reuse. Regarding the Buildings dataset, entrances and floors are recommended in the list of the key attributes, as based on these elements it's possible to calculate the height of the building. The type of use could be generic and very basic. The recommended attributes for the Cadastral Parcels are the basic ones needed to release cadastral data with respect to the GDPR and to guarantee a good level of reusability.

### 3.2.3.2 Higher intensity intervention

The higher intensity intervention is the most far-reaching intervention in terms of the measures for publication. The higher intensity intervention builds on the measures previously explained for the lower intensity option. The table below provides an overview of these measures.

Table 39 – Geospatial - Scope of the high intensity intervention

	<b>Description</b>	<b>Administrative units</b>	<b>Place Names</b>	<b>Addresses</b>	<b>Buildings</b>	<b>Cadastral parcels</b>
<b>Openness</b>	<b>License and terms of use</b>	CC0				
	<b>Format</b>	GeoPackage; GeoJSON; INSPIRE requirements.	GeoPackage; CSV; GeoJSON INSPIRE requirements.	GeoPackage; CSV; GeoJSON; INSPIRE requirements.	GeoPackage; GeoJSON; INSPIRE requirements.	GeoPackage; GeoJSON; INSPIRE requirements.
	<b>Machine-readability</b>	Mandatory				
	<b>Availability of API, bulk download</b>	Bulk download; INSPIRE distribution services; RestAPI (e.g. OGC API, ArcGIS RestAPI, Carto API).				
<b>Documentation</b>	<b>Metadata (dataset content description)</b>	INSPIRE				
	<b>Documentation (incl. structure and semantics)</b>	INSPIRE / GeoDCAT-AP				
	<b>Data linking</b>					
	<b>Shared vocabularies/taxonomies</b>	INSPIRE				
	<b>Traceability</b>	National geodata Catalog and open data catalog.				
<b>Completeness</b>	<b>Update frequency and timeliness</b>	Annual update	Annual update	Annual update	Annual update	Continuous update (close to real-time).
	<b>Granularity</b>	From municipalities to countries; sea-frontiers.	From municipalities to countries; sea-frontiers.	From municipalities to countries; sea-frontiers.	From municipalities to countries; sea-frontiers.	National coverage; Level of Scale 1:5000 or beyond (1:2000).
	<b>Key attributes</b>	National identification code; identification code of the upper administrative level; official name; country code; name in multiple languages (only for countries with more than one official language).	National identification code; identification code of the upper administrative level; official name; country code; name in multiple languages (only for countries with more than one official language).	National identification code; identification code of the upper administrative level; official name; country code; name in multiple languages (only for countries with more than one official language).	National identification code; identification code of the upper administrative level; official name; country code; name in multiple languages (only for countries with more than one official language).	Geometry of cadastral parcels; type of particle; particle code; references to the administrative area to which the particle belongs.

The main differences with the lower intensity intervention refer to the licence, the APIs and few changes on granularity and key attributes options for the datasets in scope:

- Concerning **licences and terms of use**, the recommendation is the use of CC0. According to our research, this type of licence is already adopted for the datasets in scope across several Member States. However, it continues to be not applied because it implies legal issues of compatibility and raises a lot of scepticism in terms of lack of attribution and, charging of responsibility.
- When it comes to the availability through **APIs** and **bulk download**, the option for SPARQL Endpoint was initially considered for this higher-intensity policy option, but no longer available in this version. This would imply a more accurate implementation in the use of sharing vocabularies and metadata. This is already available in the Netherlands, for example. Also, the distribution and the download through APIs option is extended to the Cadastral Parcels.
- For **granularity**, the recommended level of scale is 1:5000 or beyond, for buildings and cadastral parcels datasets. In the case of Cadastral parcels, the range of the level of scale of the existing datasets varies from 1:10000 to 1:500. In most of the countries, the data is available at a scale of 1:2000. Concerning administrative units and place names, the scale depends on what needs be visualized.
- **Traceability:** The issues related to the interoperability between the geo-catalogues and open data catalogues can be solved by using the GEO-DCAT AP and DCAT-AP 1.0. However, the actual implementation of the open data catalogues encompasses the use of DCAT-AP 2.0, which results to be incompatible with the GEO-DCAT AP. This policy option should support a resolution of these challenges.
- **Key Attributes:** Concerning the Addresses in this intervention, the zip code is here added as a key attribute, although this information is privately owned in most of the cases and, could be significantly hard to obtain. The attribute of the height is very relevant in terms of value and reuse scenarios. This information is costly for data holders, if not collected from the beginning. As defined by INSPIRE, the value of the height should be available however, could be “voidable”. The recommendation here entirely follows the INSPIRE requirements. Alternatively, the number of floors (levels) should be guaranteed. The type of use should be as detailed as possible. In this case, please see the INSPIRE recommendations, the information on the types of building are defined in their Annex III. The Cadastral Parcels’ key attributes remain unchanged compared to the previous policy intervention.

### 3.3 Meteorological data

This section presents the micro-level assessment for the thematic area of meteorological data. It illustrates the current state of play of the provision of these datasets. Furthermore, it provides the recommended measures for publication together with the costs and benefits of including these datasets as high-value datasets under the PSI Directive. Lastly, it details the three policy options proposed for this thematic area.

#### 3.3.1 As-is situation: how Member States provide these datasets today

National weather services are key elements of the public sector, with ‘protecting life and property’ as core public task. All MS collect such data. NWSs routinely work together, and data is shared with colleague institutions routinely and on an ongoing basis. All NWSs therefore collect most of the data in scope, and all of them also already have the ability to share such data they collect.

Differences do exist in (technical) capabilities, and not all NWSs collect all data within scope of this analysis (e.g. Ireland does not collect its own lightning strike data).

In the 1990's it was common for NWSs to be commercially active, but that has changed gradually since the late nineties. This is reflected in that over time, regularly also influenced by the PSI Directive, by now 8 MS publish open data, and do no longer use licenses or fees for the data itself (though marginal costs for provision and service level agreements may still apply).

The table below provides a high level overview of all information gathered on today's modes of provision by NWSs across Member States. The table remarks on the differences for the datasets in scope where possible. Furthermore, the table indicates the data gaps and the type and scale of information missing. In particular, concerning the topics of data linking, shared vocabularies and taxonomies, the data collection did not allow to gather much information. These topics seemed to be less relevant for data holders and almost never came up in the interviews with an input from Member States experts. For these reasons, these characteristics of data provision are marked as Not Applicable in the table below, although they are shortly discussed in Section 3.3.2 – Recommended measures for provision. Finally, the table indicates with an asterisk next to the name of the country (\*) those Member States for which information were only gathered through desk research and input was not received directly from Member States counterparts.

	Openness			Documentation				Completeness					
	License (terms of use)	Free of charge	of Machine-readability	Availability of API, bulk download	Metadata	Data Linking	Documentation	Shared vocabularies	Traceability	Timeliness	Granularity	Key attributes	
Austria	Contracts	True costs, limited free of charge	partly	partly	?	N/A	?	N/A	N/A	Near RT	?	No radar or lightning data (3 <sup>rd</sup> parties)	
Belgium*	Terms of use?	Some INSPIRE datasets free of charge, others paid	Yes	WFS	partly	N/A	?	N/A	N/A	Near hourly, daily	RT, Per station, nationally	Radar, observations, alerts, NWP, climate	
Bulgaria	?	yes	No	no	?	N/A	?	N/A	N/A	?	?	Observations data, other ?	
Croatia	Attribution for website, terms of use	no,	Yes	no	?	N/A	?	N/A	N/A	?	Obs 10mins per station, radar per station hourly	Observations, radar, climate, NWP, alerts	
Cyprus	?	Yes, observations and climate data at marginal costs	Yes	Api to be replaced, download	?	N/A	?	N/A	N/A	RT observations	?	Observations, alerts, climate data	
Czech Republic	Non-open license	No, some web viewable	No	no	?	N/A	?	N/A	N/A	?	Radar composite 10 mins	Observations, radar, lightning, alerts, NWP	
Denmark	Terms of use (opening 2020-2024)	No, yes from 2020	Yes from 2020	Yes from 2020, bulk t.b.d.	Yes from 2020	N/A	planned	N/A	N/A	Highest available	Highest available	All	
Estonia	?	Marginal costs (observations, climate),	Yes	no	Spatial metadata	N/A	?	N/A	N/A	(near) RT	various	Observations, radar, climate, NWP, alerts	

		no (alerts), yes (radar, lightning, NWP)											
Finland	No licenses	Yes, services charged	yes	Yes, INSPIRE req's	as yes	N/A	?	N/A	N/A	Various	various	Observations, radar, climate, NWP, alerts	
France(*)	Open national attribution, and non-open licenses	Some most at marginal costs (RT observations, radar NWP)	Yes	Some through WFS	yes	N/A	?	N/A	N/A	Various	various	All	
Germany	attribution	yes	Yes	ftp	no	N/A	some	N/A	N/A	Various	various	Observations, radar, climate, NWP, alerts	
Greece	?	No, below marginal costs	?	no	?	N/A	?	N/A	N/A	Various	Various		
Hungary	Terms of use	No, above marginal costs	Yes	?	?	N/A	?	N/A	N/A	Various	various		
Ireland	Attribution	Yes, marginal service costs may apply	Yes	ftp (radar), api for parts (NWP points)	?	N/A	?	N/A	N/A	Various, not all released	Various, not all released	No lightning data	
Italy	Terms of use	No, mostly marginal costs charged	yes	no, ftp	?	N/A	?	N/A	N/A	Various	Various	National meteo part of armed forces, some data on regional level	
Latvia	Terms of use	no	Yes	no	?	N/A	?	N/A	N/A	?	?	Observations and forecasts	
Lithuania	Non-open	Free non-commercial use	Yes	View and bulk download	?	N/A	?	N/A	N/A	?	?	Observations data	

Luxembourg	Yes	yes	Yes	downloads		N/A		N/A	N/A	RT observations, alerts, hourly validated observations		Observations, alerts, climate, radar. NWP
Malta	No	some	Yes	ftp (radar)	?	N/A	?	N/A	N/A	RT radar, hourly observations, monthly climate data	?	Not a public sector body
Netherlands	CC0	Yes, services at marginal costs	yes	Bulk download ftp, API planned	yes	N/A	yes	N/A	N/A	RT and archive obs. and radar, validated climate data	Per station (obs. & radar), NWP 2.5km grid	Observations, climate, radar, NWP, alerts
Poland	Terms of use	Yes for non-commercial use	Yes	API planned		N/A		N/A	N/A	?	Per station (observ, radar) and composite (radar)	Observations, climate, radar, alerts, NWP
Portugal(*)	?	no	?	?	?	N/A	?	N/A	N/A	?	?	?
Romania	terms of use	No	?	?	?	N/A	?	N/A	N/A	Various	Various	Observations, climate, radar, alerts, NWP
Slovakia	terms of use	No, above marinal cost charges	Yes	no	?	N/A	?	N/A	N/A	?	?	Observations, climate, radar, alerts, NWP
Slovenia	open license	Yes	yes	API under development	?	N/A	?	N/A	N/A	Various	Various	Observations, radar, climate, NWP, alerts
Spain	terms of use	marginal costs mostly,	Yes	API for selected free data	?	N/A	?	N/A	N/A	Various	Various	



		some free of charge		products								
Sweden	CC BY	Yes, some at marginal costs	Yes	Mixed API, bulk download	N/A	?	N/A	N/A	Various	Various	Observations, radar, climate, NWP, alerts	

A few considerations are connected to the table above and should be taken into account when identifying options for the future:

- **Almost all countries rely on licenses and terms of use to regulate their relation with re-users for at least part of their data provision.** While open licenses (whether international such as Creative Commons 0 or national such as the Etalab license) are in use in a number of MS, they sometimes are used next to closed licenses (e.g. France), or applied to only parts of the datasets, or terms of use and service level agreements are used for provisioning real time data and more complete datasets for instance. Frequently, a log in is required to access the data or use APIs and FTP facilities. This creates a mosaic of different conditions if you would e.g. set out to combine data from multiple MS.
- **Charging for data provision or services is a common practice.** Thirteen MS charge for basic observations data, usually mandated to do so. Eight MS do not charge for data at all. However in the latter case, marginal costs for various modes of data provision may still apply. While prices are mostly made transparent, as well as calculation models regularly, what is not made transparent is the connection to actual (marginal) costs. In some cases, e.g. the Dutch NWS, marginal costs based prices for services have been set nine years ago, and not revised since, leading us to assume there no longer is a correlation between such pricing and current actual marginal costs.
- **Common data formats are in use.** Due to the international character of the field, and the routine exchange of data between NWSs, WMO standardised formats and open standard scientific data formats are common across MS. This also means that **machine readability is widespread**, for the sake of international exchange.
- **At least seven MS have already set up APIs for the provision of meteorological data, and a similar number provides bulk download.** Three countries (Denmark, Netherlands, Poland) are in the process of creating APIs, while Cyprus is renewing their existing API. In responses we received from MS worries about the capabilities needed and the costs of developing APIs are very common. Perceived lack of capabilities and resources have led to postponing API creation (e.g. Ireland), or the decision to not go it alone but try to collaborate at European level (e.g. Luxembourg). Both dataholders and re-users have also remarked on how not all data types, especially when real time delivery is important, lend themselves easily to being incorporated into an API, and that download is then preferable. It leads some (e.g. Netherlands) to plan for a file based API at firsts, that allows for the selection of data files, but not subsets.
- **Long-listed data sets exist in most countries, except for lightning strike data.** The data core to the work of NWSs, observations, climate, alerts, and model output exist in all countries, while not every country has the technical capability for radar, or lightning strike detection. In such latter cases they may rely on data shared by another NWS (e.g. Ireland for lightning), or on data being gathered by third parties (e.g. Austria). As lightning strike data is only gathered in a minority of MS it is dropped from the list of data sets within scope.

### 3.3.2 To be situation: recommended measures for publication

The table below summarises the recommended measures for publication for the categories of datasets which have been considered.

<b>Dimensions</b>		<b>Observations Data (weather stations)</b>	<b>Climate data</b>	<b>Radar data</b>	<b>Weather alerts</b>	<b>NWP model output</b>
<i>Openness-data specification</i>	<i>License (terms of use)</i>	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right
	<i>Format</i> <sup>403</sup>	BUFR, NetCDF, ascii (for RT), json (for hourly) <sup>404</sup>	NetCDF, JSON	HDF5, JSON	XML (CAP and/or RSS-Atom) <sup>405</sup>	GRIB (or NetCDF), JSON
	<i>Machine-readability</i>	Available	Available	Available	Available	Available
	<i>Availability of API, bulk download</i> <sup>406</sup>	API and/or bulk download	API and/or bulk download	API and/or bulk download	API and/or bulk download	API and/or bulk download
<i>Documentation</i>	<i>Metadata (dataset content description)</i>	Complete (xml, *.csv document available)	Complete ( xml, *.csv document available, or included in the data)	Complete ( xml, *.csv document available, or included in the data)	Complete ( xml, *.csv document available)	Complete ( xml, *.csv document available, or included in the data)
	<i>Data linking</i>	/	/	/	/	/
	<i>Documentation (incl.</i>	Complete and web-available,	Complete and web-available,	Complete and web-available,	Complete and web-available.	Complete and web-available,

<sup>403</sup> Recommended formats are current open and WMO standards and/or optimized for keeping data volumes low. Over time these standards may change, so are suggested as currently advisable

<sup>404</sup> While BUFR and NetCDF are recommended standards in the field, ascii is common for automated stations, and both ascii and json (for hourly data) allow a broader group of re-users easy access.

<sup>405</sup> While CAP is standard, RSS-Atom are more easily accessible to a broader group of re-users.

<sup>406</sup> Most of these datasets, with the exception of hourly observations data in json, are often more suited for bulk download or API-based file retrieval, not for API data subset selection

	<i>structure and semantics)</i>					
	<i>Shared vocabularies</i>	/	/	/	/	/
	<i>Taxonomies</i>	/	/	/	/	/
<i>Completeness</i>	<i>Traceability</i>	/	/	/	NWS or national open data portal	/
	<i>Update frequency and timeliness</i>	Every 5-10 minutes in real time for automated stations, hourly unvalidated for the last 24hrs	Validated hourly (or better temporal resolution) published at least daily and daily average observations data ; historic data	Near real time in 5 minute intervals (or available shortest interval)	As issued, or hourly	Every 6hrs, or better temporal resolutions, from the last 24hrs
	<i>Granularity</i>	Per weather station, full temporal resolution	Per weather station, full temporal resolution	Per radar station in the MS, and national composite	Alerts, 24hrs or more ahead	48hrs ahead or more in 1hr steps, national, at 2.5km/best available grid
	<i>Key attributes</i>	All observation variables measured	All validated observation variables measured	Reflectivity, Backscatter polarisation, Precipitation, wind and echo-tops		Deterministic, and/or ensembles if available, for meteorologically relevant parameters and levels

As the table suggests, the same recommended modes of provision apply to the four categories of datasets in scope, except for the aspects concerning the completeness of information for which some nuances must be made. The justifications for each of these recommended measures are the following:

- Re-users and data holders we interacted with in this domain agree concerning **licenses and terms of use** that open licenses must be the default and terms of use should be avoided by default, with the possible exception of demanding attribution of source. Creative Commons 0 (CC0) licenses are the preferred default, certainly from the perspective of re-users, with CC BY as a possible alternative where attribution of source is demanded. Any further restriction in terms of license or terms of use reduces the potential socio-economic value of re-use as it creates barriers to entry for new reuse of the information, as well as friction and inefficiencies for existing re-use of the information. Countries like the Netherlands use CC0 for all their data, whereas some such as Ireland use CC BY but attach specific conditions for how to attribute. Other countries such as Germany or France apply their own license, but equivalent to CC BY. Other more restrictive licenses exist as well. While CC0 licenses are preferred by re-users, open national licenses could still be deemed acceptable with one strong caveat: the use of meteorological data is an international market, and as such data sources from different MS are regularly combined and re-used. Even in the case of national open licenses being used, that are non-harmonised across the EU, will cross-border re-use be hindered, as it quickly becomes confusing how to e.g. provide a variety of attributions within different products and services. Currently one of the most heard complaints from re-users is the diversity in licensing conditions of for instance observations data across MS. Therefore from the perspective of enabling re-use, CC0 should be used whenever possible.
- **The suggested formats** are close to the current state of play, and chosen for practicality. They should not be seen as limiting more advanced technology choices, but as a **minimum level which MS as well as European or international collaborative meteorological membership organisations are free to decide to move beyond**. While BUFR is an accepted WMO standard for the exchange of observations data, it takes an encoding step, making it often easier to provide real time observations data in ascii. Where timeliness is key that makes an important difference. Using json for non-real time observations data, next to e.g. BUFR, like ascii is more appreciated by re-users outside of the meteo domain, for its ease of re-use and combination of human and machine readability. For other, more voluminous types of data such as radar and NWP model output, HDF5 and NetCDF are suggested as these are open standards already in use, which are designed to reduce the file size for large amounts of scientific data. As otherwise these data volumes might unduly increase the need for storage and bandwidth, and thus introduce marginal costs, choosing a less voluminous format is a way to lower barriers for re-use. Practicality should inform the use of formats, and such considerations will likely change over time with shifts in technology, data volume and bandwidth needs. In that vein both data holders and re-users warned to not be overly specific in specifying formats or data content, "*don't create another INSPIRE*", and the PSI Directive should likely not presume to limit the allowed data formats, other than suggesting a minimum of following at least current practice in the field.
- While under the PSI Directive providing high value data sets through **APIs** would be mandatory, there is a common concern that the HVD for the meteo thematic sector isn't always well suited for provision through APIs. MS that already have APIs do not have them for all their available data. At least not where such an API would mean the possibility of selecting subsets of the data provided (e.g. for subsets of radar data). For instance were one to make it possible to select a specific variable from a range of automated observation stations this would introduce delays for the provision of real time data, as intermediate steps and/or transformations would be needed to make such selections possible. This is different for non-real time data, where timeliness is not a defining factor for re-use value. The use of APIs thus likely will sometimes mean the possibility of more easily specifying which entire files to retrieve (as e.g. the Netherlands is planning), essentially an easy interface to bulk download, but

not subsets of data in e.g. Json format. Additionally bulk downloads serve different use cases such as for analysis and as training sets. These concerns are voiced by both dataholders, from a fear of costs for development and transformation, and by re-users from fearing a reduction in use value for real-time data or not being able to download data in bulk and explore at their own leisure. Simpler APIs may also mean less likelihood of the data holder serving as the back-end of every form of re-use which would increase the provision costs for a data holder.

- Although not always provided today, **metadata (complete and in csv format) and complete documentation (web available)** are considered necessary for the reusability of datasets. The absence of metadata and full documentation specifically is a barrier to re-use by new entrants and re-users not directly involved in the meteo field. To be able to understand what is contained in data and explore its usefulness such metadata and documentation is key. Hence, the provision of accurate metadata and complete and web available documentation should become mandatory (from a scientific point of view NWSs often already provide documentation concerning the methods used for collecting or modelling data). It should be noted that some of the suggested formats, such as HDF5 and NetCDF are suited for metadata to be included in the data. Additionally open standards to describe the functioning of APIs, e.g. OpenAPI are useful inclusions into the documentation as well.
- Concerning **data linking**, no specific recommendations can be made based on the data available. The availability of basic data across the MS is the primary key interest of re-users. This is also true for **shared vocabularies and taxonomies**, where no specific recommendations emerge from the gathered data. The meteorological dataholders are highly collaborative amongst each other, and no need was formulated to mandate specific vocabularies and taxonomies, beyond acknowledging the existing efforts of the European level networks and the WMO efforts.
- Views on **traceability** are similar to those on the need to provide attribution based on the license, where combining data from a variety sources and across multiple MS can quickly become a challenge to correctly and viably link back to the original source of data, even more so after various transformation steps. Some data holders express that such linking to the source allows consumers to verify and trust the services and products they use in which data is incorporated. We do not suggest a specific approach to be included.
- For re-use **update frequency and timeliness** are key to most use cases, otherwise its use value rapidly diminishes to zero. Real time provision of observations data that is captured every few minutes, and of radar data often is crucial. As described in the general section on benefits, another group of use cases, which depend on validated data or time series have less need for real time delivery, although timeliness of provision after the creation of such data is still very important.
- Concerning the **granularity of the datasets**, higher temporal resolution of measurements is more useful, as is providing such measurements for each measuring station, as well as ensuring the highest geographic resolution is available (for NWP e.g. a 2.5km grid if available). In short, more detail is always better.
- Finally, in terms of **key attributes**, for (validated) observations and radar data providing data as complete as it has been gathered is of importance. In the case of observations variables like for instance global solar radiation, its importance for re-use has grown, however some data holders only include a number of basic variables in the published open data, which often leaves out radiation. Like with granularity, completeness of captured data is a key driver of potential re-use value.

### 3.3.2.1 Expected benefits

The patterns of benefits emerging in MS that have already made the switch to open data are clear, even though for some countries (e.g. Ireland, which adopted an open data policy in 2016) that beneficial impact is thus far seen as less pronounced, although specific examples are then still named. These patterns are:

- Strong growth in re-use, both number of re-users and volume per re-user, sometimes by orders of magnitude

- Reduction or removal of data fees leads to non-linear jump in re-use, indicating latent demand and a price elasticity above 1.
- Many different sectors involved in re-use
- Re-use growing outside the traditional meteorological value added services
- Novel uses regularly want more real-time data (e.g. observations, radar), want previously less re-used variables (e.g. solar radiation in observations). Additionally climate data and quality controlled observations data that enter into the climate records are seeing novel re-use for practical planning and decision making (in diverse sectors from agriculture to energy production).
- Growth in economic value creation (turnover, start-ups, employment) leads to additional tax revenue easily outpacing both the costs of provision and loss of revenue from data fees.

Data holders in their responses all expect to see growth of re-use across all stakeholder groups and across a wide variety of sectors and use cases, similar to the experiences made by countries that already opened up their data. Essentially they expect the already visible benefits described in 3.3.1.2. to emerge more fully across MS.

Re-users expect an additional beneficial effect from the availability of meteorological data across the EU under similar conditions, and point to how now the diversity of conditions for data provision are making it harder, especially to new entrants, to create value-added re-use with meteorological data across countries. This notion is reinforced by various MS noting how the number of international re-users interested in their data has been growing (e.g. Germany).

Countries such as Finland (in 2013) and Denmark (in 2016 and in 2017) initiated studies to model some of these benefits as part of the business case for their open data initiatives. In both examples expected benefits of one or just a few use cases already were significantly (i.e. multiples) larger than the costs of transition and expected loss of revenue in total.

In short, there is no doubt among data holders nor re-users of the expected types of benefits for the meteorological datasets in scope, even if the quantity of those benefits is unknown to them up front. These expectations put the focus for the NWSs we interacted with on how to approach the transition and particularly the expected costs in terms of development and technological needs, as well as the potential loss of revenue and possible compensation for that.

### 3.3.2.2 Expected costs

Taking into account the scope of the datasets to be made available as HVD and building on the recommended modes of provision suggested above, the expected costs that Member States would have to face when adapting to the PSI HVD rules for meteorological data can be discussed. The table below summarises the main cost drivers for the provision of meteorological data and provides information collected for these categories.

Cost category	Description	Insights from the data collection
Infrastructural costs	Costs related to infrastructural investments such as portals, APIs, Servers (cloud), etc.	<p>Infrastructural costs are seen as the main costs driver by most MS, especially due to the development of APIs:</p> <ul style="list-style-type: none"> <li>- <b>Austria:</b> expects costs to increase, and needs additional people for API development. Currently exploring needed changes.</li> <li>- <b>Denmark:</b> about half of the expected costs of their open data project concerns API development and infrastructure (total 5.4 million Euro for the four year project. The other half concerns revenue loss mostly.)</li> <li>- <b>Finland:</b> will replace API and foresees high costs, but less expensive if internal/external API will be combined</li> <li>- <b>France:</b> expects to need 10FTE for 3-5 years for API</li> </ul>

---

development

- **Germany:** API development, switching from the current FTP facilities, will raise costs of infrastructure
- **Hungary:** Currently preparing for setting up infrastructure for open data services, with an estimated cost of 5% of annual NWS budget.
- **Greece:** expects significant infrastructure costs, beyond the ability to absorb in the existing budgets for the next few years. Costs have not been calculated yet though.
- **Ireland:** expects significant API development costs of up to 5% of annual budget and a similar amount for maintenance, as well as costs related to high data volumes (e.g. NWP data).
- **Italy:** Current infrastructure costs are about 1% of annual budget, and would rise to 1.3% of annual budget due to API development.
- **Lithuania:** expects cloud hosting to bring down infrastructure costs
- **Luxembourg:** no resources for API development available, but see this is as the main cost, looks to European collaboration on infrastructure for data provision.
- **Netherlands:** expects to double costs for infrastructure by moving from self-owned infrastructure to cloud hosted, mainly due to API costs and increased data volume to handle. Nevertheless total costs still well below 5% of annual budget as before, partly by combining open data and SLA data provision into one infrastructure (now separate).
- **Poland:** concerns about storage of data volume for e.g. radar. Planning to spend 450 000 Euro on the creation of a data sharing center that also covers earth observation and environmental data.
- **Slovakia:** expects significant but unknown costs of up to 2 million Euro, suggests international collaboration w.r.t. data provision infrastructure.
- **Slovenia:** Expects infrastructure and API development costs to be the most significant. APIs are under development, but need to be financed from within current budgets. European collaboration welcomed.
- **Romania:** currently improving data collection infrastructure through cohesion fund financing, unknown if that would ease data provision infrastructure development.
- **Sweden:** calculated the cost increase for when they adopted an open data policy for infrastructure at around 450 000 Euro.

---

Costs related to data processing including data cleaning, preparation of metadata, aggregation, anonymisation, etc.

The data transformation costs are rarely quantified, regularly seen as much less than infrastructure costs. Some MS remarked upon these costs:

- **Austria:** expects costs increase, and specifically mentions the labour intensiveness of quality controlled data (although those are not related to data provision per se)
- **Cyprus:** non automated observation data is labour intensive to validate and process
- **Finland:** expects costs of maintaining metadata and documentation to increase in response to higher quality demands
- **Hungary:** expects ongoing costs w.r.t. data formats and conversion at about 5% of annual budget for open data provision
- **Italy:** currently 0.2% of annual budget, no change expected
- **Lithuania:** expects these costs at first but to be minor longer term. A current project to improve existing databases and metadata costs around 450 000 Euro.
- **Netherlands:** expects no change
- **Slovenia:** expects these costs to amount to 30% of the costs for API development.

Data transformation costs



Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	<p>Similarly to the data transformation costs, the operational costs are rarely made explicit, though some mention the switch to cloud as a source of increased operational costs:</p> <ul style="list-style-type: none"> <li>• <b>Finland:</b> expects provision costs to rise when the API serves as the back-end for re-users</li> <li>• <b>Germany:</b> expects to handle more user interaction specifically for NWP data, unquantified</li> <li>• <b>Hungary:</b> some additional operational costs expected for open data provision, but not distinguishable from regular maintenance and operational costs at 10-15% of annual budget.</li> <li>• <b>Italy:</b> Currently 0.6% of annual budget, no change expected</li> <li>• <b>Netherlands:</b> due to the switch to cloud, operational costs will be more visible and transparent than now. This means a doubling of apparent costs, but still well below 5% of annual budget as before.</li> <li>• <b>Poland:</b> concerns about costs of replying to user requests</li> <li>• <b>Slovakia:</b> expects up to 200k Euro in additional annual operational costs.</li> <li>• <b>Slovenia:</b> expects to need 2FTE for user interaction/support (ARSO had 89 FTE in 2010, so 2% increase in comparison)</li> </ul>
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	<p>Mostly no information on other costs was collected.</p> <ul style="list-style-type: none"> <li>• <b>Spain:</b> expects significant effort to update current regulation w.r.t. to data provision charges, and reviewing existing user contracts.</li> </ul>

As the table and the data collected suggest, expected costs are rarely quantified, specifically for data transformation and operational costs which aren't always easy to separate out for NWSs. Infrastructural costs have been scoped out more readily, specifically in the context of ongoing or planned projects.

Taking a look at the suggested publishing dimensions

---

**Recommended dimension for publication      Budgetary implication (little to none, low, medium, high)**

---

License: open license/no use/no database right	Little to none: changing the license in itself (specifically when using established open licenses such as CC0) would have no significant impact on countries' costs.
--	--

---

Formats:	Low: the suggested formats are already common in current practice, and used both for data sharing with NWSs elsewhere and for data provisioning to re-users. Some suggested formats (such as HDF5 for radar, and NetCDF) are specifically designed to reduce the resulting filesize of otherwise high volume data.
----------	--

---

Modes of provision: both API and bulk download	High: the establishment of APIs and/or bulk download would drive the most impactful costs for those countries which do not have one or the other (or both), or are replacing their currently existing API or download facilities. Meteorological data is highly dynamic (except for validated climate data and historical data), and can be voluminous (e.g. radar, NWP). Several NWSs indicate that they will look at a file based API first to more easily maintain timely delivery by avoiding transformation efforts on top of real time data. This would help contain costs for API development.
--	---

---

---

**Recommended dimension for publication**      **Budgetary implication (little to none, low, medium, high)**

---

Metadata and documentation: complete and in .csv format      Low: While some NWSs assume costs will rise, in expectance of a higher quality of metadata and documentation needed for a wider variety of re-users, the cost of providing metadata in .csv and documentation can be considered low in comparison to the (correlating) costs of setting up an API and infrastructure. IT literature considers that you need around a week of an FTE to create the documentation for an API and then one off costs for updating the documentation needs to be foreseen. For the metadata, there are no reliable estimations in terms of resources needed. Some of the suggested data formats can contain their own metadata.

---

Data linking: /      None: there is no recommended measure concerning data linking to be implemented by Member States.

---

Shared vocabularies and taxonomies: /      None: there is no recommended measure concerning adopting shared vocabularies and taxonomies.

---

Traceability: /      Little to none: traceability is currently requested in the form of attribution ('where') by some countries by way of the license or terms of use. Most have existing documentation on methods used to generate data ('how' and 'why'). The traceability of the time data was created ('when') is an integral part of the data itself.

---

Update frequency and timeliness: when available (real time)      Medium: The re-use value of observations and measurements for many use cases correlate to the timeliness of provision in (near) real time. This may impact the demand made of existing processes and infrastructure for data capturing. Otherwise these costs are tied to data provision, and as such taken into account in the infrastructure costs. It needs to be mentioned that for validated data (as validation is a labour intensive step), timely availability means 'as it comes available'.

---

Granularity: highest temporal, geographic available      Little to none: the level of granularity indicated in the recommended measures is geared towards current data collection practices 'as is'. However not all MS currently share data at the highest granularity available to them, which will have an impact on the infrastructure needs for data provision.

---

Key attributes: key identifiers      Little to none: the availability of attributes indicated in the recommended measures is geared towards current data collection practices 'as is'. However not all MS currently share all data attributes available to them (e.g. from observations), which will have an impact on the infrastructure needs for data provision

---

The costs to be expected from applying the HVD publishing dimensions to the datasets in scope for the meteorological theme reinforce the mentioned expectations of NWSs: infrastructure costs will be the main cost driver.

Additionally the requirement to make data available free of charge creates a cost in the form of loss of revenue for NWSs that don't already publish open data. Several countries have in the past seen such revenue loss from the transition to open data compensated through general government funding, which the literature suggests is the preferred option<sup>407</sup> (also as economic benefits accrue with the central government in the form of additional tax revenues that over time outpace loss of revenue). Where revenues were or are already directly funnelled to central government and not accruing with the data holder the impact of revenue loss is not perceived at the level of the data holder. The table below provides an overview with charging examples from various MS.

<b>Country</b>	<b>Experience / information w.r.t. charging</b>
Austria	Charge full true costs currently, a change of funding model is needed. They expect to see a shift away from commercial services (like Denmark is currently undergoing, and Netherlands, Norway and Germany saw beforehand)
Denmark	Denmark until now charged full production costs, with 2019 revenue at about 3 million Euro, which is around 10% of the annual budget. Expects to lose up to 800 000 Euro (2-3% of annual budget) revenue annually due to the release of open data, with the disappearing revenue for quality controlled climate data and forecast data the biggest components. Will be centrally compensated for this loss of revenue.
Croatia	Revenue from data fees make up roughly 5% of the annual budget of the NWS, would require compensation from central national budget.
Cyprus	Data is free of charge, but climatological data services are charged for under ministerial mandate. The 2019 revenue was under 20 000 Euro and is seen as insignificant.
France	Timely and complete data sets are charged for, while some basic and summarized data are free of charge. Unclear how much revenue is currently involved. The full observational data for France is currently listed for a fee of 126 000 Euro annually.
Germany	Saw annual revenue reduced by 2.3 million Euro since 2017 when data was no longer charged for. No budget changes were involved as the revenue accrued with the central government beforehand.
Greece	Fees charged are well below marginal costs
Hungary	There is a government mandate to charge above marginal costs. Revenues make up some 60% of the NWS budget (with ~35% of the budget funded from central government). Currently a fundamental change of funding model is expected. Studies for a transition to a more open data policy have been done, and a change in model is currently under discussion within government. Central budget

<sup>407</sup> Funding a System of Key Registers, De Vries, 2012, for the Danish government

---

financing up to 90% of annual NWS budget is deemed to be required.

---

Italy W.r.t. Servizio Meteorologico dell'Aeronautica Militare: Most of the fees for data provision are at marginal cost. Revenues for data provision are 0.4% of annual budget, and 45% of total revenue. Revenues from commercial service agreements flow to central government.

---

Lithuania Charges marginal costs, prices set by the responsible Ministry in 2015. Yearly revenue around 300 000 Euro.

---

Netherlands Abolished data fees in 2009. At the time revenue was below 5% of annual budget, and was compensated by the central government. Service level agreements on data provision are charged at marginal costs, the revenues of which are well below 5% of annual budget.

---

Luxembourg Switched in 2014 to open data, which was not a big impact for the NWS as a range of data was already free, and more importantly revenue did not accrue with the NWS, but with central government.

---

Poland Charges commercial use cases marginal costs, which is used for maintenance and development of the observational network. Yearly revenue around 1.5 million Euro. The change which made data for non-commercial use free of charge was not compensated by the central government and led to loss of FTE's.

---

Romania Currently 80% of the annual budget is funded from central government, 20% comes from revenue for both data provision and commercial services. Revenues remain with the NWS for the improvement and development of its capabilities. Would need to be compensated from central budget to not severely impact performing its public tasks.

---

Slovakia Revenue, above marginal costs, makes up ~25% of annual budget (average over the past 5 years)

---

Slovenia Currently has an open data policy in place. Upon transition no compensation was received from the central budget. In 2009 fees were reduced 95% (POPSIS study). Loss of revenue (360k Euro in 2009) was partially compensated by revenue from new added-value services

---

Spain Charges at marginal costs.

---

Sweden Sweden calculated an expected loss of yearly revenue of around 800 000 Euro when they removed data fees.

---

For countries that currently require data provision fees the cost of providing the listed HVD at no charge may be significant, especially if the current funding model is based on full costs recovery. For the latter it

will also likely mean changes in governance models when the funding model changes (away from commercial service provision e.g.), and may require regulatory changes.

Given all of the above, the following table provides estimates concerning the costs impact of HVD requirements per MS.

Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
Austria				X		Main costs would relate to the need to revise current full cost recovery funding model, and infrastructural costs.
Belgium					X	Revenue loss will play a role
Bulgaria					X	
Croatia	X	X				Revenue loss will affect 5% of annual budget, additionally investment in infrastructure (cloud, APIs) need.
Cyprus		X				Main costs manual data transformation, and new API now under development. Negligible revenue loss.
Czech Republic					X	Revenue loss plays a role, as likely infrastructure, based on currently publicly accessible sites/data.
Denmark	X	X				Main costs concern API development (~50% of 5.4 million as total cost for 4 year transition), and loss of revenue (to be compensated, which makes up most of the other 50% of total costs). These costs are already allocated and budgeted, so no additional impact from HVD plans.

Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
Estonia		X				Main costs would concern data delivery portal, some potential loss of revenue
Finland			X			Expect high development costs for APIs as current one is outdated. Expects higher operational costs due to becoming back-end for re-users
France		X	X			Revenue loss plays a likely significant role, though no information was provided to what extent. Some API infrastructure available, dataholder estimates 10FTE for 3-5 years (30-50 person years) for API development.
Germany		X				Infrastructure costs to rise from switch to API from FPT.
Greece		X	X			Revenue loss plays a small role as current fees are well below marginal costs. Significant costs for infrastructure (APIs) and metadata expected, beyond the capabilities of current budget for the next few years. Costs have not been quantified though.
Hungary				X		Current funding model relies on above marginal cost charging, and constitutes 60% of annual budget. New funding model with up to 90%

Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
						central budget funding already under discussion.
Ireland			X			<b>Expects significant API development costs of up to 5% of annual budget and a similar amount for maintenance, as well as costs related to high data volumes (e.g. NWP data).</b>
Italy	X					The airforce meteo service sees main new costs in the creation of API's but no higher than 0.3% of annual budget (rising from 1% to 1.3%). Revenue will be impacted 45%, which amounts to 0.4% of annual budget.
Latvia					X	Revenue loss will play a role
Lithuania	X	X				Potential revenue loss 300 000 Euro annually. Currently investing 450 000 in improving databases and metadata, but low data transformation costs in longer term. See operational costs as most impactful
Luxembourg			X			API development biggest hurdle, no resources available.
Malta		X				The Maltese IT Agency is developing a shared



Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
						infrastructure for all HVD.
The Netherlands	X					Expect infrastructural and operational costs to double, from switching to cloud provision, but to remain well below 5% of budget.
Poland		X	X			Potential revenue loss for commercial use cases 1.5 million Euro per year. Expecting high data storage costs. Previously loss of revenue wasn't compensated and led to loss of FTE's. Currently investing 450 000 Euro in a data sharing center that also covers EO/Environment.
Portugal					X	Revenue loss will play a role
Romania		X	X			Revenue loss will play a role (20% of annual budget, which includes commercial services)
Slovakia			X			Revenue represents 25% of annual budget. Expected one-off costs at 2 million Euro, and up to 200k Euro annually.
Slovenia	X					Already open, APIs under development, expects to need 2FTE additionally.

<b>Country</b>	<b>Low costs</b>	<b>Medium costs</b>	<b>High costs</b>	<b>Very high</b>	<b>Unknown (based on information available)</b>	<b>Comments</b>
Spain	X	X				Revenue loss from marginal cost charging plays a role, but unknown to what extent. Expects costs for information security and improvement of current API, but not quantified.
Sweden	X					No major costs expected for meteorological data (total costs for HVD across all sectors estimated at 4 million Euro)

Of the countries information has been available for, 4 are expect to experience relatively low costs, 4 low to medium, 4 medium costs, 4 medium to high, 3 high costs, 2 very high costs (both because of fundamental changes in financing model). In general the existing level of data provision infrastructure (due to the data involved being highly dynamic and voluminous) and the significance of current revenues play a determining role. It must be assumed that for countries for which no or little information was available both those aspects also play a significant role. Although the provision of HVDs will mean a considerable investment for some countries, either in direct costs or in reduction of revenue, it is important to realise that all available evidence of the last decade and more points to the benefits outpacing those costs, certainly over time. This is however of little consolation for those NWSs that cannot count on being compensated for the transition, or lack the resources for transition. MS with smaller NWSs may also find that the domestic market for the re-use of meteorological information develops at a lower rate than elsewhere.

To illustrate that last point we've made a cost benefit analysis for both Denmark and Ireland to compare them. Denmark is currently in active transition to an open data policy from 2020-2024. In the preparation phase, 2016 and 2017 expected benefits were documented, and these are reflected in the benefits listed below. As is a common pattern, costs concentrate on infrastructure (creation of APIs etc.), and on loss of revenue. Ireland adopted an open data policy in 2016, and has seen a more muted domestic market for re-use. With technological developments meaning larger data volumes (e.g. NWP data) and a need for more timely delivery and easier sharing, investments and expertise are needed to replace or improve existing APIs and download facilities. Mapping those costs and benefits for both countries gives the following results.

*Denmark (currently in active transition, with documented expected benefits)*

Cost components	Weight	Score	Weighed score
Infrastructural costs (~>5% annual budget)	0.30	-2	-0.60
Data transformation costs	0.15	-1	-0.15
Operational costs	0.20	-1	-0.20
Lost income for data supplier (2-3% of annual budget)	0.30	-1	-0.30
Other costs	0.05	0	0
<b>Aggregated costs for Denmark</b>		-5	-1.25

Benefit components	Weight	Benefit indicators	Score	Weighed score
Economic	0.30	<ul style="list-style-type: none"> <li>• Competition</li> <li>• Consumer benefits [X]</li> <li>• Economic output [X]</li> <li>• Employment [X]</li> <li>• Product market dynamism [X]</li> <li>• Productivity[X]</li> </ul>	+3	+ 0.90
Environmental/ climate change	0.20	<ul style="list-style-type: none"> <li>• Citizen engagement [X]</li> <li>• Energy management and efficiency [X]</li> <li>• Environment management [X]</li> </ul>	+3	+ 0.60
Innovation & AI	0.04	<ul style="list-style-type: none"> <li>• Citizen innovation [X]</li> <li>• Entrepreneurialism &amp; private sector</li> </ul>	+2	+ 0.08

innovation [X]				
Public services and public administration	0.20	<ul style="list-style-type: none"> <li>Public sector revenue [X]</li> <li>Public services management [X]</li> </ul>	+2	+ 0.40
Re-use	0.25	<ul style="list-style-type: none"> <li>Demand for information [X]</li> <li>Trust and confidence in information [X]</li> <li>Volume and range of information [X]</li> </ul>	+3	+ 0.75
Social	0.01	<ul style="list-style-type: none"> <li>Mobility efficiency and planning [X]</li> </ul>	+1	+ 0.01
<b>Aggregated benefits for Denmark</b>			<b>+14</b>	<b>+2.73</b>

Benefits and costs for Denmark	Score
Aggregated benefits	+2.73
Aggregated costs of	-1.25
<b>Overall impact</b>	<b>+1.48</b>
<b>Benefit/cost ratio</b>	<b>+2.18</b>

*Ireland (already open, facing new investment, with few observed domestic benefits)*

Cost components	Weight	Score	Weighed score
Infrastructural costs (~>5% annual budget)	0.30	-2	-0.60
Data transformation costs	0.15	-1	-0.15
Operational costs (~>5% annual budget)	0.20	-2	-0.40
Lost income for data supplier	0.30	-0	-0.00
Other costs	0.05	0	0
<b>Aggregated costs for Ireland</b>		<b>-5</b>	<b>-1.15</b>

Benefit components	Weight	Benefit indicators	Score	Weighed score
Economic	0.30	<ul style="list-style-type: none"> <li>Competition</li> <li>Consumer benefits [X]</li> <li>Economic output</li> <li>Employment</li> <li>Product market dynamism [X]</li> <li>Productivity</li> </ul>	+2	+ 0.60
Environmental/ climate change	0.20	<ul style="list-style-type: none"> <li>Citizen engagement [X]</li> <li>Energy management and efficiency [X]</li> <li>Environment management</li> </ul>	+2	+ 0.40
Innovation & AI	0.04	<ul style="list-style-type: none"> <li>Citizen innovation</li> <li>Entrepreneurialism &amp; private sector innovation [X]</li> </ul>	+1	+ 0.04
Public services and public administration	0.20	<ul style="list-style-type: none"> <li>Public sector revenue</li> <li>Public services management [X]</li> </ul>	+1	+ 0.20

Re-use	0.25	<ul style="list-style-type: none"> <li>• Demand for information [X]</li> <li>• Trust and confidence in information [X]</li> <li>• Volume and range of information [X]</li> </ul>	+3	+ 0.75
Social	0.01	<ul style="list-style-type: none"> <li>• Mobility efficiency and planning</li> </ul>	+0	+ 0.00
<b>Aggregated benefits for Ireland</b>			<b>+9</b>	<b>+1.99</b>

<b>Benefits and costs for Ireland</b>		<b>Score</b>
Aggregated benefits		+1.99
Aggregated costs of		-1.15
<b>Overall impact</b>		<b>+0.84</b>
<b>Benefit/cost ratio</b>		<b>+1.73</b>

While both Denmark and Ireland show a net positive result in the CBA, the difference between benefits and costs is much narrower for Ireland than for the Danish case. The difference is more marked even, considering that of the use cases into account for the Danish, at least 4 identified use cases already have larger benefits on their own than the cost of transition. In contrast the Irish case to justify new investments is harder to make, seen at the national level at least. Sharing expertise or even solutions across MS may play a role in mitigating some of the impact of transition to HVDs. Also seeking to let existing collaborative structures between MS, such as the ongoing creation of a European meteorological infrastructure, which are now focused on international cooperation between MS, assist MS in open data provision, may be a path. This to ensure all MS have a more equal opportunity to benefit from meteorological HVD provision.

### 3.3.3 Recommended policy options

All of the above suggests that all datasets taken into scope should be considered high-value datasets under the PSI Directive. The cost-benefit perspective can be expected to always come out in favour of benefits, as the experiences of NWSs who already provide this data have experienced. Existing research and cases consistently point out that the economic and societal benefits of such a policy choice would exceed the costs of implementation for the Member States and would bring great benefits to the data economy at the EU level, over time resulting in additional tax revenue outpacing the costs of provisioning. Meteorological data sets are complicated data sets in terms of volume and frequency, therefore providing these data sets come with a considerable effort in terms of infrastructure and for some also loss of revenue attached. Several options can be contemplated to reduce or limit the costs of transition:

1) **Shorten the list of datasets in scope:** if a decision on shortening the list of datasets needs to be taken, our team recommends to prioritise above all observations data and climate data for inclusion in the Implementing Act. Observations are at the very start of the meteorological value chain, and removing current barriers and fragmentation will already mean a strong step forwards for the European re-use market. Observations and validated observations as part of the permanent climate data record together are valuable to the widest group of use cases and sectors. This is not to say that e.g. radar data and NWP data are not valuable, on the contrary, yet they are of interest to a more narrow but also more innovative set of use cases and re-users. Both radar and NWP are the data sets that due to timeliness and volume also put the most strain on infrastructure (although the suggested data formats in the publishing dimensions help mitigate volume). It is not advised to limit the attributes within a provided data set, e.g. leaving out specific observation variables. Novel uses regularly build on heretofore less used attributes /

variables (e.g. global solar radiation has increased in importance to re-users significantly), and leaving them out would preclude such uses.

2) **Lengthen the transition:** allowing more time can help even out the costs for transitioning to an open data policy for these data sets, and thus reduce the burden for NWSs and smoothen planning of the transition, including for compensating for loss of revenue (e.g. Denmark has embarked on a 4 year transition). Lengthening the transition might be done in two ways. Either by adding all datasets to the list of HVD's now, and allowing a certain longer time period for implementation, or by adding observations, climate data and weather alerts now, and additionally deciding upon a future date on which the remaining data sets will be added.

3) **Ease the transition with pragmatic technological choices:** Already the proposed publishing dimensions aim at allowing NWSs pragmatic choices, such as for data formats that allow faster delivery where data transformation steps would perhaps reduce the value coming with timeliness, or for data formats that reduce the needed bandwidth, or data formats that work better for re-users outside of the meteorological sector itself. Similarly NWSs might make different choices, based on practicality, between preferring an API or bulk download, or the type of API (e.g. file based API, or an API that does allow for subset selections) most suited for different use cases, in order to e.g. better facilitate real time data provision (where adding data to an API might introduce time delays detrimental to re-use value).

4) **Ease the transition with collaboration:** The NWSs of the MS have different capabilities and resources at their disposal, as well as different domestic markets where re-use value can emerge. This makes for uneven starting positions. Already NWSs collaborate on a European meteorological infrastructure, and for some this may provide a path to needed expertise, collaboration and shared resources, to comply with the HVD list. Currently the European meteorological infrastructure isn't developed with an eye to public data provision, but it would likely not be a difficult addition. One would need to be careful however that provision over e.g. a shared infrastructure does not become detrimental to the timely delivery of real time data, where delays mean the quick deterioration of re-use potential.

Data holders and re-users see a clear value chain for meteorological data. Such value is being created in a large variety of sectors. MS with existing open data policies generally show a non-linear increase in demand after making the change, and show benefits generally outweighing costs, often by multiples on even a few use cases. The highest frictions identified in re-use by both re-users and data holders come from mixed licensing issues and fees being charged, especially also for cross-border use cases. Given the datasets within scope (observations data, climate data, radar data, numerical weather prediction model data, and weather alerts), revenue loss for data holders that do not have an open data policy yet, will always be in play. Re-use value of meteorological data is strongly linked to timeliness, completeness and highest temporal/geographic resolution available.

The policy options presented in the next sub-sections therefore focus on variations in the data involved, and less on the publishing dimensions involved. With all interventions the suggested publishing dimensions for formats are aimed at staying close to current practice in the sector (e.g. formats already being used for data sharing between NWSs). They therefore should not be seen as an ambition that is put forward, but a minimum which is certainly not intended to limit the efforts of MS that want to move to more technologically advanced publishing dimensions (provided they are open standards based).

Data holders and re-users see a clear value chain for meteorological data. Such value is being created in a large variety of sectors. MS with existing open data policies generally show a non-linear increase in demand after making the change, and show benefits generally outweighing costs, often by multiples on even a few use cases. The highest frictions identified in re-use by both re-users and data holders come

from mixed licensing issues and fees being charged, especially also for cross-border use cases. Given the datasets within scope (observations data, climate data, radar data, numerical weather prediction model data, and weather alerts), revenue loss for data holders that do not have an open data policy yet, will always be in play. Re-use value of meteorological data is strongly linked to timeliness, completeness and highest temporal/geographic resolution available. The policy options presented therefore focus on variations in the data involved, and less on the publishing dimensions involved. With all interventions the suggested publishing dimensions for formats are aimed at staying close to current practice in the sector (e.g. formats already being used for data sharing between NWSs). They therefore should not be seen as an ambition that is put forward, but **a minimum which is certainly not intended to limit the efforts of MS that want to move to more technologically advanced publishing dimensions** (provided they are open standards based).

Box 3 – Validation workshop results: meteorological data, overall appreciation of policy intervention options

During the validation workshop organised on 28 July 2020, participants were requested to indicate whether they agreed or disagreed with the three<sup>408</sup> proposed policy options. The meteorological data options received the following appreciations (53 Respondents): Agree: 94% and Disagree: 6%.

In addition, the policy options were evaluated by participants as regards their relevance with regards to the overall environment of the thematic area and the respective needs of the participants. The three options obtained the following scores:

- Low: This option obtained the score of 4.3/10 (52 Respondents).
- Medium: This option obtained the score of 4.7/10 (52 Respondents).
- High: This option obtained the score of 6.5/10 (52 Respondents).

### 3.3.3.1 Lower intensity interventions

Observations from (automated) weather stations are the starting point of the entire meteorological data value chain. At any level of intensity interventions would need to include them, as they are fundamental. To this digitised structured historical climate data are added. This forms the permanent record of meteorological data, to which the validated observations data are added daily.

Weather alerts are at the core of the public task of meteorological offices (NWSs), preventing harm and damage from adverse weather conditions, and are already published as information. Including them as open data in the lower intensity intervention requires a minimal effort on the side of NWSs.

The lower intensity intervention proposed therefore encompasses three datasets: observations, climate data, which consists of validated observations (which may contain corrections from the original measurements or the removal of anomalies etc, and are then entered into the permanent climate record) and digitised structured historical climate data, and weather alerts.

Dealing with loss of revenue is unavoidable even at these lower intensity interventions. 13 MS currently charge for observations data.

---

<sup>408</sup> In the initial version of this Deliverable, three policy intervention options were considered per thematic area. For the final version of this Deliverable, and upon request of the Commission, the initial three policy options were merged into two policy options, a lower and higher intensity options. All elements composing the initial three options were transferred through to the final two options, and as such, the validation of the stakeholders still holds.

Timeliness, completeness and highest temporal/geographic resolution available drive the value of re-use. For observations reducing the number of variables made available (e.g. limited to temperature, humidity and precipitation) or the timeliness and/or temporal resolution (e.g. hourly averages, instead of 5-10 minute intervals), would immediately limit re-use value.

With regard to publishing dimensions the low intensity interventions stay close to what is already common for NWSs, based on how NWSs currently share such data amongst each other, or provide them as open data. As stated such publishing dimensions should be taken as minimum standard, leaving room for MS to adopt more technologically advanced, open standards based, choices in the future.

<b>Dimensions</b>		<b>Observations</b>	<b>Climate data: validated observations, and digitised structured historical data</b>	<b>Weather alerts</b>
<b>Openness-data specification</b>	License (terms of use)	Creative Commons 0 (CC0) or CC-BY 4.0; no terms of use		
	Format	BUFR, NetCDF, ascii (for RT), JSON (for hourly)	NetCDF, JSON	XML (Cap and/or RSS/Atom)
	Machine-readability	Mandatory		
	Availability of API, bulk download	Both API and bulk download		
<b>Documentation</b>	Metadata (dataset content description)	Complete (xml, .csv document available)		
	Data linking	No recommendation		
	Documentation (incl. structure and semantics)	Complete and web-available		
	Shared vocabularies	No recommendation		
	Taxonomies	No recommendation		
<b>Completeness</b>	Traceability	No recommendation	No recommendation	NWS or National open data portal
	Update frequency	Every 5-10	Validated hourly	As issued or



	and timeliness	minutes in real time for automated stations, hourly unvalidated for all stations, for the last 24hrs	observations (or better temporal resolution) published at least daily, and daily average observations data; structured historic data	hourly
	Granularity	Per weather station, full temporal resolution	Per weather station, full temporal resolution	Alerts, 24hrs or more ahead
	Key attributes	All observation variables measured	All validated measured observation variables ; daily averages per variable	

As the table suggests, some of the recommended modes of provision apply to all datasets in scope of these interventions, for some publishing dimensions no specific recommendations are given, while some (formats, update frequencies/timeliness, granularity and key attributes) are specific. The reasoning for these recommended measures is as follows:

- When it comes to **licences and terms of use**, for open data in general CC0 is preferable. However multiple MS indicate a wish to use, or are already using CC-BY licenses or a national license equivalent to either CC0 or CC-BY. As long as no further terms of use are attached to an equivalent-to-open national license (e.g. specifying how attribution should take place and be styled, or limiting commercial re-use) this need not be detrimental to ease of re-use. However when combining datasets from across MS differences in licenses quickly create high friction for re-use, by needing to satisfy a variety of licensing conditions and terms of use. The use of preferably CC0 or alternatively CC-BY provide a clear, unambiguous and globally recognised signal to re-users on the conditions of re-use.
- The mentioned **formats** are chosen to follow current common practice in the field. They should not be seen as limiting more advanced technology choices, but as a minimum level which MS as well as European or international collaborative meteorological membership organisations are free to decide to move beyond. While BUFR is an accepted WMO standard for the exchange of observations data, it takes an encoding step, making it often easier to provide real time observations data in ascii. Where timeliness is key that makes an important difference. Using json for non-real time observations data, next to e.g. BUFR, like ascii is more appreciated by re-users outside of the meteorological domain, for its ease of re-use and combination of human and machine readability. All mentioned formats are **machine-readable**.
- No recommendation is made with regard to **data linking, shared vocabularies and taxonomies** in the context of the PSI Directive. No specific recommendations emerge from the gathered information from data holders and re-users. Both INSPIRE and the existing collaborative European and

international organisations in the sector are already more pertinently positioned to provide guidance on these topics, obviating the need for making such presumptions under the PSI Directive.

- Although not always provided today, **metadata (complete and in csv or xml format) and complete documentation (web available)** are considered necessary for the reusability of datasets. The absence of metadata and full documentation specifically is a barrier to re-use by new entrants and re-users not directly involved in the meteorological field. To be able to understand what is contained in data and to explore its usefulness such metadata and documentation is key. Hence, the provision of accurate metadata and complete and web available documentation should become mandatory (from a scientific point of view NWSs often already provide documentation concerning the methods used for collecting or modelling data).
- Views on **traceability** are similar to those on the need to provide attribution based on the license, where combining data from a variety of sources and across multiple MS can quickly become a challenge to correctly and viably link back to the original source of data, even more so after various transformation steps. Some data holders express that such linking to the source allows consumers to verify and trust the services and products they use in which data is incorporated. We do not suggest a specific approach to be included, with the possible exception of Weather Alerts. As the latter are warnings of danger to prevent harm and damage, traceability of Weather Alerts to their source can help maintain trust for end-users of such information.
- For re-use **update frequency and timeliness** are key to most use cases, otherwise its use value rapidly diminishes to zero. Real time provision of observations data that is captured every few minutes often is crucial. Use cases which depend on validated data or time series have less need for real time delivery, although timeliness of provision after the creation of such data is still important.
- Similarly concerning the **granularity of the datasets**, providing the highest available temporal resolution of measurements and providing such measurements for each measuring station, even at the low intensity intervention level is important to maintain re-use value.
- Finally, in terms of **key attributes**, for (validated) observations providing data as complete as it has been gathered is of importance. Novel use cases often build on historically less used measurement variables (e.g. solar radiation). Like with granularity and timeliness, completeness of captured data is a key driver of re-use value.

### 3.3.3.2 Higher intensity interventions

The suggested higher intensity interventions build on the lower intensity interventions by extending the scope of data sets involved. Three additions are made, compared to the lower intensity interventions. First, digitised heretofore unstructured historical climate data are added, introducing potential data transformation costs. Second, radar data is added. Not all MS operate radar stations from within the public sector. Third, Numerical Weather Prediction (NWP) model data is added. Data from radar stations can become voluminous, and modelling data is very voluminous, though scientific data formats are in use that can help keep such volumes down. For NWP model data, a further significant reduction of data volume is possible by selecting parameters and vertical layers of meteorological interest (not every parameter is relevant for use at every layer. E.g. humidity at height might be less relevant than wind speeds at heights relevant for wind energy parks). As NWSs have concerns regarding their capabilities and resources for the creation of APIs and the needed storage to deal with NWP data and radar data, the European Meteorological Infrastructure (EMI) may play a role in data provision. Such public data provision through the EMI is currently not foreseen (its focus is on data sharing between NWSs and shared computing resources).

Differences and additions in comparison with the low intensity interventions are marked in blue in the table below.



Dimensions	Observations	Climate data: validated observations	Weather alerts	Radar	NWP data	model
Openness-data specification	License (terms of use)	CC0 or CC-BY 4.0; no terms of use				
	Format	BUFR, NetCDF, ascii (for RT), JSON (for hourly)	NetCDF, JSON	XML (Cap or RSS / Atom)	HDF5, JSON	GRIB (or NetCDF), JSON
	Machine-readability	Mandatory				
	Availability of API, bulk download	Both API and bulk download				
Documentation	Metadata (dataset content description)	Complete (xml, .csv document available)				
	Datalinking	No recommendations				
	Documentation (incl. structure and semantics)	Complete and web-available				
	Shared vocabularies	No recommendations				
	Taxonomies	No recommendations				
Completeness	Traceability	No recommendations				
	Update frequency and timeliness	Every 5-10 minutes in real time for automated stations, hourly unvalidated for all stations, for the last 24hrs	Daily validated hourly (and better temporal resolution) and daily average observations data; all digitised historical data	As issued or hourly	Near real time in 5 minute intervals (or available shortest interval)	Every 6hrs, or better temporal resolution, from the last 24hrs.
	Granularity	Per weather station, full temporal	Per weather station, full temporal	Alerts, 48hrs or more	Per radar station in the MS and	48 hours ahead or more in 1hr steps, national,

	resolution	resolution	ahead	national composite	at 2.5km/best available grid
Key attributes	All observation variables measured	All validated measured observation variables ; daily average per variable		Reflectivity, Backscatter, polarization. Precipitation, wind, and echotops	Deterministic and/or ensembles if available, for meteorologically relevant parameters and levels.

With regard to the publishing dimensions the same considerations apply as for the lower intensity interventions, with a few additional considerations specific to the extended data scope. For the ease of reference the **considerations for publishing dimensions are repeated below, with additions marked in blue.**

- When it comes to **licences and terms of use**, for open data in general CC0 is preferable. However multiple MS indicate a wish to use, or are already using CC-BY 4.0 licenses or a national license equivalent to either CC0 or CC-BY 4.0. As long as no further terms of use are attached to an equivalent-to-open national license (e.g. specifying how attribution should take place and be styled, or limiting commercial re-use) this need not be detrimental to ease of re-use. However when combining datasets from across MS differences in licenses quickly create high friction for re-use, by needing to satisfy a variety of licensing conditions and terms of use. The use of preferably CC0 or alternatively CC-BY 4.0 provide a clear, unambiguous and globally recognised signal to re-users on the conditions of re-use.
- The mentioned **formats** are chosen to follow current common practice in the field. They should not be seen as limiting more advanced technology choices, but as a minimum level which MS as well as European or international collaborative meteorological membership organisations are free to decide to move beyond. While BUFR is an accepted WMO standard for the exchange of observations data, it takes an encoding step, making it often easier to provide real time observations data in ascii. Where timeliness is key that makes an important difference. Using json for non-real time observations data, next to e.g. BUFR, like ascii is more appreciated by re-users outside of the meteo domain, for its ease of re-use and combination of human and machine readability. **For more voluminous types of data such as radar and NWP model data, the suggested formats are open standards already in use, which are designed to reduce the file size for large amounts of data. As otherwise these data volumes might unduly increase the need for storage and bandwidth, and thus increase marginal provision costs.** All mentioned formats are **machine-readable**.
- No recommendation is made with regard to **data linking, shared vocabularies and taxonomies** in the context of the PSI Directive. No specific recommendations emerge from the gathered information from data holders and re-users. Both INSPIRE and the existing collaborative European and international organisations in the sector are already more pertinently positioned to provide guidance on these topics, obviating the need for making such presumptions under the PSI Directive.
- Although not always provided today, **metadata (complete and in csv or xml format) and complete documentation (web available)** are considered necessary for the reusability of datasets. The absence of metadata and full documentation specifically is a barrier to re-use by new entrants and re-users not directly involved in the meteo field. To be able to understand what is contained in data and explore its usefulness such metadata and documentation is key. Hence, the provision of accurate metadata and complete and web available documentation should become mandatory (from a scientific

point of view NWSs often already provide documentation concerning the methods used for collecting or modelling data). It is noted that some of the suggested formats currently in use, such as HDF5 and NetCDF are suited for metadata to be included in the data.

- Views on **traceability** are similar to those on the need to provide attribution based on the license, where combining data from a variety sources and across multiple MS can quickly become a challenge to correctly and viably link back to the original source of data, even more so after various transformation steps. Some data holders express that such linking to the source allows consumers to verify and trust the services and products they use in which data is incorporated. We do not suggest a specific approach to be included, with the possible exception of Weather Alerts. As the latter are warnings of danger to prevent harm and damage, traceability of Weather Alerts to their source can help maintain trust for end-users of such information.
- For re-use **update frequency and timeliness** are key to most use cases, otherwise its use value rapidly diminishes to zero. (Near) Real time provision of observations data that is captured every few minutes often is crucial, as it is for radar data. Use cases which depend on validated data or time series have less need for real time delivery, although timeliness of provision after the creation of such data is still important.
- Similarly concerning the **granularity of the datasets**, providing the highest available temporal resolution of measurements and providing such measurements for each measuring station, [or radar station is important \(creating national composites and not also releasing individual radar station data may reduce timeliness of data availability\)](#), as well as [ensuring the highest geographic resolution is available \(for NWP e.g. a 2.5km grid if available\)](#).
- Finally, in terms of **key attributes**, providing data as complete as it has been gathered is of importance. Novel use cases often build on historically less used measurement variables. Like with granularity and timeliness, completeness of captured data is a key driver of re-use value.

### 3.4 Earth observation and environment

This section presents the micro-level assessment for the thematic area of earth observation and environment. It illustrates the current state of play of the provision of these datasets. Furthermore, it provides the recommended measures for publication together with the costs and benefits of including these datasets as high-value datasets under the PSI Directive. Lastly, it details the three policy options proposed for this thematic area.

#### 3.4.1 As-is situation: how Member States provide these datasets today

The data within scope of the Earth Observation and Environmental data thematic area, are all covered by INSPIRE Themes. INSPIRE sets as a baseline that existing data must be discoverable, and viewable for free to the public (although exceptions to free viewers are possible, and may be restricted to non-commercial use). Services should also exist to download or access the data directly, but there is no obligation within the INSPIRE Directive to provide such services for free. The Directive on access to environmental information is built on the same premise.

To establish how Member States provide these datasets today, the 'as-is' situation we can therefore expect that if the data exists, that it will be discoverable through the national and European INSPIRE portals. We can also expect that any differences in how data is being provided will concern differences in terms of use mostly (e.g. licensing, fees for non-environmental data/commercial usage), and that there will be less difference in technical aspects such as formats, provision modes, and the existence of metadata, due to the harmonizing effect of INSPIRE.

For the three data sets taken as an example the next tables summarises how MS provide these datasets currently.

First the table for Natura 2000 Areas. Eighteen MS use an open license (some requiring attribution), and all provide the data free of charge where charging information is available. Nine MS already provide APIs or both APIs and bulk download, twenty-one MS provide bulk download, one of which at request only. For two MS the information was not established.

Openness-data specification					Documentation					Completeness				
License (terms of use)	Free of charge	Form at	Machine - readability	Availability of API, bulk download	Metadat a (dataset content description)	Data linking	Docum entati on (incl. structure and semantics)	Shared vocabular ies	Taxonomie s	Traceability	Timeliness	Granularit y	Key attribute s	
Austria	License CC BY 3.0	Free of charge	GML/ XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 6 months ago	N/A	N/A
Belgium	Terms of use: internal, non-commercial use only,	Free of charge	GML/ XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 23/01/2020	N/A	N/A
Bulgaria	Open license	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 6 months ago	N/A	N/A
Croatia	Open license	Free of charge	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information not available	N/A	N/A
Cyprus	Open license	Free of charge	XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/12/2015	N/A	N/A
Czech Republic	License CC BY 4.0	Free of charge	GML/ XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 15/04/2020	N/A	N/A
Denmark	Open license	Free of charge	WMS/ WFS/ XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 6 months ago	N/A	N/A
Estonia	Open license	Free of charge	XML/ WMS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 23/09/2019	N/A	N/A
Finland	License CC BY 4.0	Free of charge	WMS/ WFS/ GML/ XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 28/05/2019	N/A	N/A



	Openness-data specification				Documentation				Completeness					
	License (terms of use)	Free of charge	Form at	Machine - readability	Availability of API, bulk download	Metadat a (dataset content description)	Data linking	Docum entati on (incl. structure and semantics)	Shared vocabular ies	Taxonomie s	Traceability	Timeliness	Granularit y	Key attribute s
France	Terms of use: copyright	Free of charge	SHP	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 16/06/2010	N/A	N/A
Germany	Terms of use: copyright and in certain areas licensing	Free of charge	XML	Yes	No API. Bulk download is only available on request	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 29/10/2018	N/A	N/A
Greece	Open license	Free of charge	RDF, SHP, WMS, WFS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/04/2015	N/A	N/A
Hungary	Information not available	Information not available	WFS	Yes	API available.	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 29/02/2016	N/A	N/A
Ireland	License	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 03/08/2019	N/A	N/A
Italy	Licenses	Information not available	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 07/08/2014	N/A	N/A
Latvia	Open license	Free of charge	GML/XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 13/12/2019	N/A	N/A
Lithuania	Terms of use: non-commercial and copyright	Free of charge	GML and SHP	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 15/05/2015	N/A	N/A

Openness-data specification					Documentation					Completeness				
	License (terms of use)	Free of charge	Form at	Machine - readability	Availability of API, bulk download	Metadat a (dataset content description)	Data linking	Docum entati on (incl. structure and semantics)	Shared vocabular ies	Taxonomie s	Traceability	Timeliness	Granularit y	Key attribute s
Luxembo urg	Open license	Free of charge	GML/XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 10 years ago	N/A	N/A
Malta	Some data are restricted from public access	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 02/07/2018	N/A	N/A
Netherla nds	Open license	Free of charge	WMS/WFS	Yes	API available.	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 01/06/2016	N/A	N/A
Poland	Open license	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/09/2019	N/A	N/A
Portugal	Open license	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 13/03/2013	N/A	N/A
Romania	Access is restricted	Information not available	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 08/12/2010	N/A	N/A
Slovakia	Open license	Free of charge	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 06/03/2019	N/A	N/A
Slovenia	Open license	Free of charge	GML/XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 13/11/2019	N/A	N/A
Spain	License CC BY 3.0	Free of charge	XML/SHP	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/01/2010	N/A	N/A
Sweden	Open license	Free of charge	XML/WMS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 23/03/2020	N/A	N/A

Second the table for Air quality measurements. As this concerns environmental data, free of charge is the norm. Air quality measurements would be well suited for publication through an API, yet only a minority of MS do so at the moment, judged by INSPIRE portal information and input from MS themselves. Most (20) but certainly not all use open licenses allowing access to this data.

Openness-data specification					Documentation					Completeness				
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timelines	Granularity	Key attributes
Austria	Open license	Free of charge	zip	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 27/03/2020	N/A	N/A
Belgium	Open license	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 09/12/2019	N/A	N/A
Bulgaria	Open license	Free of charge	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 08/11/2019	N/A	N/A
Croatia	Open license	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 29/04/2020	N/A	N/A
Cyprus	Information not available	Free of charge	Information not available	Yes	API available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information not available	N/A	N/A
Czech Republic	Open license	Price list available	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 09/01/2020	N/A	N/A
Denmark	Open license	Free of charge	WMS/WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 6 months ago	N/A	N/A

Openness-data specification					Documentation					Completeness				
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timelines	Granularity	Key attributes
Estonia	Open license	Free of charge	XML/WMS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information not available	N/A	N/A
Finland	Open license	Free of charge	WMS/WFS/GML/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 13/12/2019	N/A	N/A
France	Terms of use: copyright	Free of charge	SHP	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 16/06/2010	N/A	N/A
Germany	License CC BY 2.0	Free of charge	XML	Yes	No API. Bulk download is only available on request	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 19/10/2017	N/A	N/A
Greece	Open license	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 12/11/2019	N/A	N/A
Hungary	Information not available	Information not available	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information not available	N/A	N/A
Ireland	License CC BY 4.0	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 28/11/2019	N/A	N/A
Italy	Information not available	Information not available	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information not available	N/A	N/A
Latvia	Open license	Free of charge	XML/WMS	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 04/01/2016	N/A	N/A

Openness-data specification					Documentation					Completeness				
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timelines	Granularity	Key attributes
Lithuania	Terms of use: copyright	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 26/08/2019	N/A	N/A
Luxembourg	Open license	Free of charge	GMD/XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 2 months ago	N/A	N/A
Malta	Open license	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 24/12/2018	N/A	N/A
Netherlands	Open license	Free of charge	GMD/XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 09/12/2019	N/A	N/A
Poland	Open with specific attribution requirements	Free of charge	JSON	Yes	API	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Hourly	N/A	N/A
Portugal	Open license	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 29/04/2020	N/A	N/A
Romania	Open license	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 16/10/2017	N/A	N/A
Slovakia	INSPIRE basic license	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 03/05/2017	N/A	N/A
Slovenia	Open license	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/11/201	N/A	N/A

Openness-data specification					Documentation					Completeness				
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timelines	Granularity	Key attributes
												9		
Spain	Terms of use: The Ministry of Agriculture and Fisheries, Food and Environment is mentioned as the author and owner of the information	Free of charge	XML/SH	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 14/03/2018	N/A	N/A
Sweden	CC0 license <sup>409</sup>	Free of charge	XML/WMS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 23/03/2020	N/A	N/A

And last the table for land parcel data (cadastral parcels). Here the differences in data provision are more pronounced. At least eight MS charge in some way for this dataset, and there is a wider variety of use conditions and licenses applied. The manner of data provision, in terms of formats and API/Bulk download is much more standard across all MS for which information is available.

Openness-data specification				Documentation				Completeness						
	<i>License (terms of use)</i>	<i>Free of charge</i>	<i>Format</i>	<i>Machine-readability</i>	<i>Availability of API, bulk download</i>	<i>Metadata (dataset content description)</i>	<i>Data linking</i>	<i>Documentation (incl. structure and semantics)</i>	<i>Shared vocabularies</i>	<i>Taxonomies</i>	<i>Traceability</i>	<i>Timelines</i>	<i>Granularity</i>	<i>Key attributes</i>
Austria	Standard terms of use of the Federal Office of Metrology and Surveying (BEVS)	Standard fees of the Federal Office of Metrology and Surveying (BEVS)	WFS/WMS/GML/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 18/02/2020	N/A	N/A
Belgium	Terms of use: copyright	Free of charge	WMS/WFS/SHP/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 01/01/2017	N/A	N/A
Bulgaria	Information not available	Information not available	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information not available	N/A	N/A
Croatia	Applying to access and use on request	Costs are legally determined	WMS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 02/05/2019	N/A	N/A
Cyprus	Open license	Data provided by payment	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/12/2015	N/A	N/A
Czech Republic	Public access is not restricted	Information is not available	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 27/03/2020	N/A	N/A
Denmark	Open license	Free of charge	WMS/WFS/XML	Yes	Both API and bulk	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated	N/A	N/A

Openness-data specification					Documentation					Completeness				
	<i>License (terms of use)</i>	<i>Free of charge</i>	<i>Format</i>	<i>Machine-readability</i>	<i>Availability of API, bulk download</i>	<i>Metadata (dataset content description)</i>	<i>Data linking</i>	<i>Documentation (incl. structure and semantics)</i>	<i>Shared vocabularies</i>	<i>Taxonomies</i>	<i>Traceability</i>	<i>Timelines</i>	<i>Granularity</i>	<i>Key attributes</i>
					download are available							01/06/2018		
Estonia	Open license	Free of charge	XML/WMS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 23/09/2019	N/A	N/A
Finland	Terms of use: copyright	Free of charge	WMS/WFS/GML/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Elements are updated weekly	N/A	N/A
France	Information is not available	Information is not available	Information is not available	Yes	Information is not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information is not available	N/A	N/A
Germany	The terms of use of the Bavarian Surveying Administration apply.	The fee and price list (GebPL) of the Bavarian Surveying Administration apply.	WMS/WFS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 01/05/2019	N/A	N/A
Greece	Creative Commons CC BY 3.0	Free of charge	RDF, SHP, WMS, WFS	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 22/05/2020	N/A	N/A
Hungary	Terms of use: intellectual property rights	It is a commercial product	WMS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 14/05/2019	N/A	N/A



Openness-data specification					Documentation					Completeness				
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timelines	Granularity	Key attributes
Ireland	License	Information not available	GML/XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 10/04/2020	N/A	N/A
Italy	License	The data is free for Public Administrations, but private individuals are charged	WMS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 12/06/2018	N/A	N/A
Latvia	A request must be made for use	Information not available	GML/XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 13/12/2019	N/A	N/A
Lithuania	License: Not for commercial use and copyright	Costs apply	WMS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 16/03/2020	N/A	N/A
Luxembourg	Open license	Free of charge	WMS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 2 months ago	N/A	N/A
Malta	Download service not	Marginal costs are charged	WMS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 12/02/2019	N/A	N/A

Openness-data specification					Documentation					Completeness				
	<i>Licens e (terms of use)</i>	<i>Free of charge</i>	<i>Format</i>	<i>Machine-readability</i>	<i>Availabil ity of API, bulk download</i>	<i>Metadata (dataset content description )</i>	<i>Data linking</i>	<i>Documentati on (incl. structure and semantics)</i>	<i>Shared vocabularie s</i>	<i>Taxonomie s</i>	<i>Traceability</i>	<i>Timelines</i>	<i>Granularit y</i>	<i>Key attributes</i>
	officially available	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 25/02/2020	N/A	N/A
Netherlands	Creative Commons License BY 4.0	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 25/02/2020	N/A	N/A
Poland	Information not available	Information not available	Information not available	Yes	Information not available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Information not available	N/A	N/A
Portugal	Open license	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/03/2020	N/A	N/A
Romania	Intellectual property rights	Information not available	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 26/10/2012	N/A	N/A
Slovakia	Limitations on public access	Free of charge	XML	Yes	No API. Bulk download possible	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 12/12/2019	N/A	N/A
Slovenia	License CC BY 2.5 (transitioning to License CC BY 4.0)	Free of charge	WFS/XML	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 11/11/2019	N/A	N/A
Spain	License CC BY	Free of charge	WFS/XML	Yes	Both API and bulk	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated	N/A	N/A

Openness-data specification					Documentation					Completeness				
	<i>License (terms of use)</i>	<i>Free of charge</i>	<i>Format</i>	<i>Machine-readability</i>	<i>Availability of API, bulk download</i>	<i>Metadata (dataset content description)</i>	<i>Data linking</i>	<i>Documentation (incl. structure and semantics)</i>	<i>Shared vocabularies</i>	<i>Taxonomies</i>	<i>Traceability</i>	<i>Timelines</i>	<i>Granularity</i>	<i>Key attributes</i>
	4.0				download are available							05/12/2019		
Sweden	Terms of use: copyright	Free of charge	XML/WM S	Yes	Both API and bulk download are available	Available	N/A	N/A	Based on INSPIRE	N/A	N/A	Last updated 13/02/2020	N/A	N/A

As expected there are strong commonalities in data provision across all three datasets in terms of formats and standards. Differences are mostly

concentrated on re-use conditions and restrictions, and for earth observation data on charging practices (as environmental data is mainly free of charge, based on the Directive on access to environmental information)

Differences in usage restrictions and licensing as well as fees, are indeed very regularly mentioned as a barrier to effectively combine data sets across MS, as well as bundling them into services (e.g. Copernicus, w.r.t. in-situ data). These differences are perceived by stakeholders as a much higher barrier to usage, than aspects like the quality of the data, timeliness, or the area of coverage available. Sourcing such data is currently seen as high-effort and high-friction.

### 3.4.2 To be situation: recommended measures for publication

The table below summarises the recommended measures for publication for the example datasets which have been considered.

<b>Dimensions</b>		<b>Land parcels</b>	<b>Air quality measurements</b>	<b>Natura 2000 areas</b>
<i>Openness-data specification</i>	<i>License (terms of use)</i>	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right	Creative Commons 0 or BY (or equivalent open license) No terms of use No database right
	<i>Format<sup>410</sup></i>	GML over WFS	GML over WFS, JSON, XML	GML over WFS / GeoJSON
	<i>Machine-readability</i>	Available	Available	Available
	<i>Availability of API, bulk download<sup>411</sup></i>	API and/or bulk download	API and/or bulk download	API and/or bulk download
<i>Documentation</i>	<i>Metadata (dataset content description)</i>	Complete INSPIRE (XML document available)	Complete INSPIRE (XML document available)	Complete INSPIRE (XML document available)
	<i>Data linking</i>	/	/	/
	<i>Documentation (incl. structure and semantics)</i>	Complete and web-available	Complete and web-available	Complete and web-available
	<i>Shared vocabularies</i>	INSPIRE	INSPIRE	INSPIRE
	<i>Taxonomies</i>	INSPIRE	INSPIRE	INSPIRE
<i>Completeness</i>	<i>Traceability</i>	/	/	/
	<i>Update frequency and</i>	As changes occur	Real time	As changes occur

<sup>410</sup> Recommended formats are current open and common standards used within INSPIRE. Over time these standards may change, so are suggested as currently advisable.

<sup>411</sup> Some of the EO/Environment datasets are more suited for bulk download or API-based file retrieval, not for API data subset selection, pragmatic choices should be leading in that regard.

	<i>timeliness</i>			
	<i>Granularity</i>	Parcels	Per measuring station, full temporal resolution (hourly at least)	Natura 2000 areas
	<i>Key attributes</i>	Parcel boundaries, classification	All measured variables	Boundaries, description

**In general for the entire list of Environmental e-reporting priority themes and INSPIRE themes one would follow the recommendations already defined for both the e-reporting priorities and INSPIRE, while at most adding gml and (geo)JSON for formats, a more modern API standard such as OGC:API instead of the now common WFS, and adopting the DCAT-AP2 standard for metadata, provided the current differences with GeoDCAT are resolved.**

The justifications for each of these recommended measures are the following:

- Re-users and data holders we interacted with in this domain agree concerning **licenses and terms of use** that ideally open licenses must be the default and terms of use should be avoided by default, with the possible exception of demanding attribution of source. Creative Commons 0 (CC0) licenses are the preferred default, certainly from the perspective of re-users, with CC BY as a possible alternative where attribution of source is demanded. Any further restriction in terms of license or terms of use reduces the potential socio-economic value of re-use as it creates barriers to entry for new reuse of the information, as well as friction and inefficiencies for existing re-use of the information. Some countries such as the Netherlands use CC0 for all their data, whereas some such as Slovenia use CC BY. Other countries such as France apply their own license, but equivalent to CC BY. While CC0 licenses are preferred by re-users, open national licenses could still be deemed acceptable with one strong caveat: the use of earth observation and environmental data is not just local, and as such data sources from different MS are regularly combined and re-used. Even in the case of national open licenses being used, that are non-harmonised across the EU, will cross-border re-use be hindered, as it quickly becomes confusing how to e.g. provide a variety of attributions within different products and services. Therefore from the perspective of enabling re-use, CC0 should be used whenever possible.
- **The suggested formats** are close to the current state of play, and chosen for practicality. As all the datasets within scope, including the three held up as an example above, fall within INSPIRE the existing INSPIRE instructions regarding format and structure can apply and be seen as the minimum requirement (with MS able to anticipate more modern formats, proposed by professional bodies, e.g. UN-GGIM Europe, that will over time likely be part of INSPIRE recommendations). Currently INSPIRE mandates discoverability, meaning metadata is available, and accessibility for viewing to be free of charge (exceptions w.r.t. fees for viewing can be made). There is no such mandate for downloading and re-use of data. In many cases one or more MS already add the step of making that data available as open data, or for non-commercial purposes. Adding these datasets as HVD adds a requirement for free data download and re-use, meaning the existing format choices of those countries already doing that can be seen as a de-facto standard.
- While under the PSI Directive providing high value data sets through **APIs** would be mandatory, there is some concern that the HVD for the earth observation and environment thematic sector isn't always well suited for provision through APIs. At least not where such an API would mean the possibility of selecting subsets of the data provided. For instance were one to make it possible to select a specific

variable from a range of automated environmental monitoring stations this might introduce delays for the provision of real time data, as intermediate steps and/or transformations might be needed to make such selections possible. This is different for non-real time data, where timeliness is not a defining factor for re-use value. The use of APIs thus likely may sometimes mean the possibility of more easily specifying which entire files to retrieve, essentially an easy interface to bulk download, but not subsets of data in e.g. Json format. Additionally bulk downloads serve different use cases such as for analysis and as training sets. These concerns are voiced by both dataholders, from a fear of costs for development and transformation, and by re-users from fearing a reduction in use value for real-time data or not being able to download data in bulk and explore at their own leisure. Simpler APIs may also mean less likelihood of the data holder serving as the back-end of every form of re-use which would increase the provision costs for a data holder.

- **Metadata (complete and in xml format) and complete documentation (web available)** are considered necessary for the reusability of datasets. Metadata and full documentation specifically is of use to new entrants and re-users relatively new to the field. To be able to understand what is contained in data and explore its usefulness such metadata and documentation is key. As all datasets in scope are covered by INSPIRE, metadata requirements are largely already fulfilled.
- Concerning **data linking**, no specific recommendations can be made based on the data available. The availability of data across the MS is the primary key interest of re-users. This is also true for **shared vocabularies and taxonomies**, although in those cases existing INSPIRE prescriptions should continue to apply.
- Views on **traceability** are similar to those on the need to provide attribution based on the license, where combining data from a variety sources and across multiple MS can quickly become a challenge to correctly and viably link back to the original source of data, even more so after various transformation steps. Some data holders express that such linking to the source allows consumers to verify and trust the services and products they use in which data is incorporated. We do not suggest a specific approach to be included.
- There are different types of data within the scope of earth observation and environment. Some are measurements (e.g. environmental measurements, or seismic measurements) taking place on an ongoing basis. Other datasets are also measurements or observations, but get renewed at a much slower pace (e.g. every few years for LIDAR and ortho-photos). Yet other datasets represent administrative decisions and outcomes (designated zones, norms, registrations). While for planning and policy cycles it might be enough to have such data up to date at the moment they are needed in such cycles, for other uses and for re-use it is key that **update frequency and timeliness** are in line with the character of data capturing. This means delivery as changes occur, e.g. near real-time w.r.t. to environmental measurements, is the recommended aim.
- Concerning the **granularity of the datasets**, higher temporal and geographic resolution of any measurements is more useful, as is providing data at the same level as it is generated. In short, more detail is always better. Providing the data at the highest temporal and geographic granularity at which data is collected is to be required.
- Finally, in terms of **key attributes**, for any dataset that results from measurements (e.g. air quality, water quality) or observations (e.g. LIDAR, ortho-imagery), the consensus is to provide the data as completely as captured. Offering selections is making assumptions about potential re-use, and will by default preclude some forms of re-use, particularly novel ones. Like with granularity, completeness of captured data is a key driver of potential re-use value.

#### 3.4.2.1 Expected benefits

Dataholders and re-users identify different user groups for this data, that take a strong interest already. These groups are:

- Citizens

- Research
- Companies
- Non-profit organisations
- Public sector bodies

What stands out is that expectations of benefit by stakeholders seem more or less equally spread over these groups. This is also borne out by the stakeholder survey the Finnish Environment Institute conducted last year. Dataholders note a growing demand from citizens and researchers for more timely and disaggregated data, especially concerning environmental data. While to dataholders commercial re-users often are much less apparent, encountered use cases are spread over a wide variety of sectors and there are many existing services both aimed at consumers and business-to-business and business-to-government e.g. to aid in decision making. Where data has already been opened by MS, there is generally a non-linear growth in data demand by all re-user groups, both in number and volume.

Consistently both dataholders and re-users indicate that the public sector itself can expect large benefits from opening up data in the earth observation and environmental thematic area, and the public sector stands out as a clear re-user group. Benefits to the public sector will concentrate on two areas. First the reduction of friction and costs in sourcing and combining data from other public sector bodies for existing policy and planning processes. This works in two directions: where data is held at state or regional level, multiple MS state it is difficult and costly to acquire data for use at national level public sector entities, and conversely local decision making and planning benefit greatly from easily being able to include national level data. Where currently charges are levied between public sector bodies the reduction of friction will be most profound. Second, benefits are expected from new possibilities using data in public task execution (e.g. the use by local governments of digital elevation model data for water management, or the use of that same data for archeological site detection and protection), and in facilitating interaction with external stakeholders e.g. when they can more easily use environmental data for obligatory environmental impact assessments.

Another widely expect benefit is that reducing the friction of sourcing earth observation and environmental data from across all MS will find an immediate use in existing Copernicus services and lead to improvement of these services (e.g. land parcel and use data). The use value and impact of Copernicus services is not in doubt with stakeholders (ESA has e.g. commissioned a large number of case studies into the impact of Sentinel data within Copernicus<sup>412</sup>), and that value will grow with broader availability of in-situ and contextual data.

Consistency in timely availability of similar data across multiple or all MS, thus providing wider area coverage, will lower the threshold for new entrants, and allow service improvement and development in current sectors already seeing value from re-use. Multiple MS notice international interest in their data, and some indicate having an interest in re-using EO and environmental data from neighbouring MS themselves.

Measured and observation data is generally regarded as more valuable and re-usable if accompanied by contextual and administrative data. However that administrative data will be of limited re-use value if not accompanied by measurement and observation data. This puts the primary focus on data that concerns measurements and establishing current status (e.g. air quality measurements, land use, species distribution, lidar, seismology etc.) as they make the bigger difference in terms of re-use potential and novel uses, but we must at the same time be aware that improving the availability of administrative data

---

<sup>412</sup> EARSC, The Green Land, IIASA, 2017-2021 for ESA, <http://ears.org/Sebs/all-cases/>

(zones, sites, and particularly parcels, etc) increases the use value of measurement data, and/or provides crucial context to make a form of re-use have value in the first place.

The patterns of benefits emerging in MS from current re-use, are expected to strengthen. These patterns are:

- Strong growth in re-use, both number of re-users and volume per re-user, upon removal of usage restrictions, across all user groups (citizens, research, commercial sector, public sector)
- Reduction or removal of data fees leads to non-linear jump in re-use, indicating latent demand
- Many different sectors involved in re-use
- Re-use growing outside the traditional value added services
- Novel uses regularly want more dynamic data (e.g. environmental measurements) in real time.
- Growth in economic value creation (turnover, start-ups, employment) will leads to additional tax revenue, possibly outpacing both the costs of provision and loss of revenue from data fees.

Data holders in their responses almost all expect to see growth of re-use across all stakeholder groups and across a wide variety of sectors and use cases, similar to the experiences made by countries that already opened up their data. Essentially they expect the already visible benefits described in 3.4.1.2. to emerge more fully across MS.

Whereas dataholders generally have a limited notion of the economic value attached to current re-use or possible with new re-use, regardless of rising visible demand, MS that have explored potential benefits (e.g. Finland, Sweden), conclude very large potential economic contributions (in the case of Finland and Sweden both in the hundreds of millions Euro of economic value, see 3.4.1.2.). Given the nature of some of the data sets involved this is consistent with observed patterns in other thematic areas. Several datasets in scope, e.g. ortho-imagery, digital elevation models/LIDAR scans, are high effort datasets and unique national datasets. Others, such as land parcel and land use data are the only authoritative source, or the only high quality and high consistency source (e.g. environmental monitoring measurements). Such unique national coverage authoritative when opened as HVD, can be expected to create impacts along the lines of other base register data (e.g. Denmark’s Good Basic Data program, or the Dutch integrated system of base registers) in terms of uptake, economic impact and resulting additional tax revenue.

In short, there is no doubt among data holders nor re-users of the expected types of benefits for the datasets in scope, even if the quantity of those benefits is unknown to them up front. These expectations put the focus for the MS we interacted with on how to approach the transition and particularly the expected costs in terms of development and technological needs, as well as the potential loss of revenue and possible compensation for that.

### 3.4.2.2 Expected costs

Taking into account the scope of the datasets to be made available as HVD and building on the recommended modes of provision suggested above, the expected costs that Member States would have to face when adapting to the PSI HVD rules for earth observation and environmental data can be discussed. The table below summarises the main cost drivers for the provision of earth observation and environmental data and provides information collected for these categories. It must be stated though that the information that the MS were able to provide concerning costs were mostly generic and limited, also due to the wide variety of types of data within scope of the assessment.

Cost category	Description	Insights from the data collection
Infrastructural costs	Costs related to infrastructural investments such as portals,	Infrastructural costs are seen as the main costs driver by most MS, especially due to the development of APIs:



---

APIs, Servers (cloud), etc.

- **Austria:** Current data provision is mostly download based, and API development is likely a considerable cost.
- **Cyprus:** Does not expect additional costs, as investment in initial development of API, as well as hardware and bandwidth considerations have already been made earlier.
- **Czech Republic: expects no further costs for Nature Conservation Agency's data**
- **Denmark:** Expect no further costs for data w.r.t. reporting obligations and ortho-imagery, as these are already compliant.
- **Estonia:** Currently creating a platform for the provision of environmental data and meteorological data, at a cost of 400 000 Euro. Current costs for earth observation data provision, particularly for land parcel/use, ortho-imagery and LIDAR data are inseparable from the overall costs of central IT facilities. Expect no further costs, as open data is already current policy, and centrally funded.
- **Finland:** internal and open data provision share same infrastructure so costs for open data provision hard to separate. See efficiency gain from no longer maintaining contracts/user accounts etc for external uses, but recognise external data provision also means some additional costs.
- **Germany:** As many datasets are held at state or even communal level, not at federal level, there is concern about spiralling API development costs if APIs on all these different and fragmented datasets are needed, unless current INSPIRE facilities for download would satisfy requirements.
- **Lithuania:** Infrastructure and API development costs for all EO and environmental data sets may exceed 10 million Euro as part of national spatial information services, though impossible to estimate per dataset as they reside in the same information system.
- **Malta:** Expects INSPIRE infrastructure to already be mostly compliant. Newly needed API interfaces might create significant costs however. Work ongoing to set a common standard for APIs across the public sector.
- **Netherlands:** already adopted open data policy, expects no further costs.
- **Poland:** A shared infrastructure (ecudo.pl) is currently under construction at a costs of 450 000 Euro. This will serve both EO/environmental data and meteorological data. Costs center on infrastructure / API development.
- **Slovenia:** Expects no additional costs, as open data policy already in place, for earth observation. Transformation costs for environmental data do play a role (see below).

---

Costs related to data processing including data cleaning, preparation of metadata, aggregation, anonymisation, etc.

The data transformation costs are rarely quantified, regularly seen as much less than infrastructure costs. A Ffw MS remarked upon these costs:

- **Finland:** Data harmonisation and metadata creation may create significant costs if PSI Directive adds requirements different from existing INSPIRE requirements.
- **Malta:** Adapting data for publication in API interfaces may be a significant cost.
- **Poland:** The main costs of the planned ecudo.pl platform are for infrastructure and API development, but bringing currently disparate datasets together in machine readable formats is also a significant effort.
- **Slovenia:** Sees costs to bring together environmental data sets that are currently dispersed across different agencies, as well as transformation costs, as these data are sometimes in documentary formats (PDF e.g.) Work is ongoing to collate and then open these data.

Data transformation costs

Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	Similarly to the data transformation costs, the operational costs are rarely made explicit, though some mention rising demand means additional bandwidth costs and potentially user interaction (though in contrast other MS expect user interaction to be less costly, due to reduction in requests to process manually)
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	No information on other costs was collected, except for: <ul style="list-style-type: none"> <li>• <b>Lithuania:</b> expects labour hours needed for regulatory changes (e.g. removing current restrictions on commercial use from existing regulations)</li> </ul>

In terms of concerns about costs several MS remark on the fact that as all datasets within scope are covered by INSPIRE themes, costs would be more significant if the PSI Directive for HVD would add significantly different and additional requirements for data provisioning, data structure, and metadata in comparison with existing INSPIRE requirements. In the suggested publishing dimensions this has been taken into account, in the sense that it already builds upon the common practices that have been informed by INSPIRE requirements. Taking a look at the suggested publishing dimensions in terms of expected costs, results in the following table:

**Recommended dimension for publication      Budgetary implication (little to none, low, medium, high)**

License: open license/no terms of use/no database right	Little to none: changing the license in itself (specifically when using established open licenses such as CC0) would have no significant impact on countries' costs.
Formats:	Low: the suggested formats are already common in current practice, and used for INSPIRE download and API services.
Modes of provision: API and/or bulk download	High: the establishment of APIs and/or bulk download where one or both do not currently exist would drive the most impactful costs. Being able to take existing INSPIRE infrastructure as a starting point will help contain costs, specifically for download services. API development costs might be further contained by distinguishing between datasets where an API is better suited for timely delivery of data, and where download.
Metadata (xml) and documentation (web available): complete	Low: While some MS assume efforts will grow, in expectance of a higher quality of documentation being needed for a wider variety of re-users, the cost of providing metadata and documentation can be considered low in comparison to the (correlating) costs of setting up an API and infrastructure. The more so as existing INSPIRE metadata will be sufficient for the suggested publishing dimensions.
Data linking: /	None: there is no recommended measure concerning data linking to be implemented by Member States.

---

**Recommended dimension for publication      Budgetary implication (little to none, low, medium, high)**

---

Shared vocabularies and taxonomies: /      None: there is no recommended measure concerning adopting shared vocabularies and taxonomies, other than following the existing INSPIRE vocabularies and taxonomies.

---

Traceability: /      Little to none: traceability is currently requested in the form of attribution ('where') by some countries by way of the license or terms of use. Existing INSPIRE metadata is sufficient for other aspects of traceability ('how', 'why', 'when').

---

Update frequency and timeliness: as changes/occur, real time      Medium: The re-use value of measurements for many use cases correlate to the timeliness of provision in (near) real time. This may impact the demand made of existing processes and infrastructure for data capturing. Otherwise these costs are tied to data provision, and as such taken into account in the infrastructure costs. Other data that changes regularly can be provided as changes occur, whereas a large proportion of datasets will see infrequent changes, or have a predictable renewal rhythm. This is already largely reflected in how these datasets are now made available through INSPIRE infrastructure.

---

Granularity: highest temporal, geographic available      Little to medium: the level of granularity indicated in the recommended measures is geared towards current data collection practices 'as is' (i.e. requiring the highest temporal/geographic resolution being collected by a MS). However not all MS currently share data at the highest granularity available to them and/or provide e.g. daily aggregates, which will have an impact on the infrastructure needs for data provision.

---

Key attributes: key identifiers      Little to none: the availability of attributes indicated in the recommended measures is geared towards current data collection practices 'as is'. However not all MS may currently share all data attributes available to them, which will have an impact on the infrastructure needs for data provision

---

The to be expected costs from applying the HVD publishing dimensions to the datasets in scope for the earth observation environmental thematic area largely reinforce the mentioned expectations of MS: API development will be a cost driver, where those not already exist, and costs for providing more granular and timely data (both factors positively impacting re-use value) will apply where currently those data sets are dispersed at the moment, or where a limited set of aspects or aggregated form is currently made available.

In addition to the above the PSI Directive's requirement to make HVD available free of charge has an impact for those MS that currently don't provide already fully open data, and make use of the provisions within the INSPIRE Directive that allow fees to be applied for access and re-use of data, e.g. for commercial purposes. Charging plays a role more prominently with earth observation data, as environmental data is typically free of charge. Several countries have in the past seen such revenue loss

from the transition to open data compensated through general government funding, which the literature suggests is the preferred option<sup>413</sup> (also as economic benefits accrue with the central government in the form of additional tax revenues that over time outpace loss of revenue). Several MS have remarked upon current charging practices and revenue, presented in the following table:

Country	Experience / information w.r.t. charging
Austria	Land parcel data is currently charged for, also between public sector bodies. A reduction of overall costs and administrative simplification is expected of abolishing fees. Previous experience with removing fees for land cover data is that demand rose strongly, while beforehand charging was likely a net-loss due to administrative costs.
Czech Republic	Czech Geological Survey charges marginal costs for data provision (viewers are free of charge), would expect compensation from central budget for open data provision
Estonia	Data from the Estonian Land Board (e.g. elevation models, land cover, land parcel, ortho-imagery, water and river basins) is available under an open data policy since July 2018, no fees.
Finland	Open data policy in place since 2008 for the Finnish environmental agency, covering 19 out of 34 INSPIRE themes.
Germany	Land parcel and land use data are essential data for environmental uses, and held by states. These data are charged to, also between public sector bodies. Revenue loss would be at least partly compensated by efficiency gains for other public sector bodies.
Lithuania	Lithuania uses two pricing policies: A) Data holders are fully covered from the national budget and data is provided free of charge (although data may be restricted for commercial re-use still), B) Revenue is a key part of the data holder's budget. Environmental data is free of charge, land cover data, elevation models, and ortho-imagery have commercial re-use limitations. Land parcel data is charged, and perceived as expensive.
Netherlands	Open data policy in place, no fees.
Slovenia	Open data policy in place, no fees.
Sweden	The Swedish Land Bureau is required to charge, also to other public sector bodies, for ortho-images, land parcel data and land cover data, to partly fund the organisation. Revenue loss would be their only significant cost towards an open data policy. The revenue loss for the mentioned data sets would be around 5.8 million Euro, annually.

---

<sup>413</sup> Funding a System of Key Registers, De Vries, 2012, for the Danish government

Given all of the above, the following table provides estimates concerning the costs impact of HVD requirements per MS, where such an estimate is at least somewhat possible. For other MS where information is at the moment largely absent, this is marked as 'unknown'.

Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
Austria		X				Main costs would relate to API development, and loss of revenue.
Belgium					X	Revenue loss will likely play a role
Bulgaria					X	
Croatia					X	Revenue loss will likely play a role
Cyprus	X					Costs of API development hardware and bandwidth already previously allocated. For EO data revenue loss will play a role.
Czech Republic					X	Revenue loss will play a role
Denmark	X					Reporting obligations and ortho-imagery available through API, no significant other costs expected.
Estonia	X					Main costs concern data delivery portal currently being created, no other costs anticipated.
Finland	X					Open data policy in place covering 19 out of 34 INSPIRE themes, current infrastructure expected to be adequate, unless PSI Directive significantly deviates from technical specifications for INSPIRE.
France					X	Revenue loss will likely play a role.
Germany		X				Many datasets held at state or communal level, not federal level: concern about API development costs if APIs on all these different and fragmented datasets are needed, unless current INSPIRE facilities for download satisfactory.
Greece					X	
Hungary					X	Revenue loss will play a role
Ireland					X	

Country	Low costs	Medium costs	High costs	Very high	Unknown (based on information available)	Comments
Italy					X	Regional structures currently existing may mean high adaptation costs. Revenue loss will play a role
Latvia					X	
Lithuania		X	X			Infrastructure and API development costs for all EO and environmental data sets may exceed 10 million Euro as part of national spatial information services
Luxembourg					X	
Malta		X				The Maltese IT Agency is developing a shared infrastructure for all HVD.
Netherlands	X					Expects no additional efforts.
Poland		X	X			Shared infrastructure (ecudo.pl) is currently under construction at a cost of 450 000 Euro. This will serve both EO/environmental data and meteorological data. Costs center on infrastructure / API development, and bringing together disparate datasets. Some concern API development for some datasets may be more costly than benefits.
Portugal					X	
Romania					X	
Slovakia					X	
Slovenia					X	Has open data policy, unknown needed adaptations.
Spain					X	Regional structures currently existing may mean higher adaptation costs.
Sweden		X				No major costs expected (total costs for HVD across all sectors estimated at 4 million Euro), but expected loss of 5.8 million Euro in revenue

In general detailed cost-benefit analyses are hard to find or make, because of a lack of pertinent information. For a more detailed costs-benefits analysis (CBA), w.r.t. to the three example datasets we refer to the CBA in the Geospatial thematic area, that includes cadastral (land) parcels. Here a specific case comparing two countries is provided for digital elevation data. The case compares Sweden and the

Netherlands. Sweden commissioned an in-depth study into the costs and benefits for HVD<sup>414</sup>, which includes digital elevation model data, that is currently not open and charged for. The Netherlands has provided this data as open data since 2014, did a study into its impact in 2016<sup>415</sup>, and is currently improving both collection and provision of these data. Current information for the Netherlands was obtained through an interview.

LIDAR data collected for digital elevation models are not collected by all MS in full national coverage, but its collection is growing as it becomes technologically and operationally more feasible to collect. The reason for its absence, or regional but not national coverage seems mostly determined by opportunity therefore. Given its clear re-use value, including it in scope is in anticipation of wider collection and availability in coming years.

*Sweden (currently not open, with documented expected benefits)*

Cost components	Weight	Score	Weighed score
Infrastructural costs (API already exists)	0.30	-1	-0.30
Data transformation costs (<5% budget)	0.20	-1	-0.20
Operational costs (<5% budget)	0.20	-1	-0.20
Lost income for data holder	0.30	-2	-0.60
Other costs	0.00	0	0
<b>Aggregated costs for Sweden</b>		-5	-1.30

Benefit components	Weight	Benefit indicators	Score	Weighed score
Economic	0.25	<ul style="list-style-type: none"> <li>• Competition</li> <li>• Consumer benefits</li> <li>• Economic output [X]</li> <li>• Employment</li> <li>• Product market dynamism [X]</li> <li>• Productivity [X]</li> </ul>	+3	+ 0.75
Environmental/ climate change	0.20	<ul style="list-style-type: none"> <li>• Citizen engagement</li> <li>• Energy management and efficiency [X]</li> <li>• Environment management [X]</li> </ul>	+2	+ 0.40
Innovation & AI	0.09	<ul style="list-style-type: none"> <li>• Citizen innovation</li> <li>• Entrepreneurialism &amp; private sector innovation [X]</li> <li>• Public sector innovation</li> <li>• Public private innovation</li> </ul>	+1	+ 0.08
Public services and public administration	0.20	<ul style="list-style-type: none"> <li>• Public sector revenue</li> <li>• Public service performance [X]</li> <li>• Public services management [X]</li> </ul>	+2	+ 0.40
Re-use	0.25	<ul style="list-style-type: none"> <li>• Demand for information [X]</li> <li>• Trust and confidence in information</li> </ul>	+2	+ 0.50

<sup>414</sup> Damvad for Swedish Land Agency, 2020

<sup>415</sup> Wageningen University, 2016

		<ul style="list-style-type: none"> <li>• Volume and range of information [X]</li> </ul>		
Social	0.01	<ul style="list-style-type: none"> <li>• Public engagement and government transparency understanding</li> </ul>	+0	+ 0.00
<b>Aggregated benefits for Sweden</b>			<b>+10</b>	<b>+2.13</b>

Benefits and costs for Sweden		Score
Aggregated benefits		+2.13
Aggregated costs		-1.30
<b>Overall impact</b>		<b>+0.83</b>
<b>Benefit/cost ratio</b>		<b>+1.64</b>

*Netherlands (already open, observed domestic benefits, investing in better data)*

Cost components	Weight	Score	Weighed score
Infrastructural costs (<5% of budget)	0.30	-1	-0.30
Data transformation costs (<5% of budget)	0.20	-1	-0.20
Operational costs (<5% of budget)	0.20	-1	-0.20
Lost income for data supplier	0.30	-0	-0.00
Other costs	0.00	-0	-0
<b>Aggregated costs for the Netherlands</b>		<b>-3</b>	<b>-0.70</b>

Benefit components	Weight	Benefit indicators	Score	Weighed score
Economic	0.25	<ul style="list-style-type: none"> <li>• Competition [X]</li> <li>• Consumer benefits</li> <li>• Economic output [X]</li> <li>• Employment</li> <li>• Product market dynamism [X]</li> <li>• Productivity [X]</li> </ul>	+3	+ 0.75
Environmental/ climate change	0.20	<ul style="list-style-type: none"> <li>• Citizen engagement</li> <li>• Energy management and efficiency [X]</li> <li>• Environment management [X]</li> </ul>	+2	+ 0.40
Innovation & AI	0.09	<ul style="list-style-type: none"> <li>• Citizen innovation</li> <li>• Entrepreneurialism &amp; private sector innovation [X]</li> <li>• Public sector innovation [X]</li> <li>• Public private innovation</li> </ul>	+2	+ 0.18
Public services and public administration	0.20	<ul style="list-style-type: none"> <li>• Public sector revenue [X]</li> <li>• Public service performance [X]</li> <li>• Public services management</li> </ul>	+2	+ 0.40
Re-use	0.25	<ul style="list-style-type: none"> <li>• Demand for information [X]</li> <li>• Trust and confidence in information [X]</li> </ul>	+3	+ 0.75



		<ul style="list-style-type: none"> <li>• Volume and range of information [X]</li> </ul>		
Social	0.01	<ul style="list-style-type: none"> <li>• Public engagement and government transparency understanding</li> </ul>	+0	+ 0.00
<b>Aggregated benefits for the Netherlands</b>			<b>+12</b>	<b>+2.48</b>

<b>Benefits and costs for the Netherlands</b>	<b>Score</b>
Aggregated benefits	+2.48
Aggregated costs	-0.70
<b>Overall impact</b>	<b>+1.78</b>
<b>Benefit/cost ratio</b>	<b>+3.54</b>

The key difference in costs between Sweden and the Netherlands stems from the loss of revenue from data provision in Sweden, should elevation data be opened up. In the Netherlands the data publication is financed centrally, and by water authorities and provinces (the main public sector stakeholders), with contributions expected from all public entities after the current financing scheme ends at the end of 2022. On the benefits side, some effects have had a longer time to emerge in the Netherlands, but many were visible within 2 years of first availability. For example, annual tax revenue from economic activity growth from re-use, was at a comparable level as the annual cost of data provision, after 2 years. This points to a rising benefits/costs ratio over time.

### 3.4.3 Recommended policy options

Earth observation and environmental data already play an important role, but there is also much unused potential benefit still. That potential benefit does not reside in specific data sets as such, but in making the full breadth and depth of earth observation and environmental data available for re-use. We therefore suggest to **place all listed INSPIRE themes and the Environmental e-reporting priority data list within scope of the HVD requirements of the PSI Directive**. This would mean, without doing away with Article 13 of the INSPIRE Directive, that **the PSI Directive would add an open data and re-use mandate to the listed INSPIRE themes** by requiring there be a data provision service as in Article 11(1) point (c) of the INSPIRE Directive to which the HVD requirements of the PSI Directive apply (free of charge, open license, machine readable, through API or bulk download). That adds a layer of additional benefit on top of the existing value of, and existing infrastructure for INSPIRE data provision. Sourcing earth observation and environmental data is currently seen as high-friction and high-cost and thus the primary barrier to re-use value creation (e.g. sourcing EU wide land parcel and land use data for Copernicus services), which adding the HVD requirements to these INSPIRE themes would do away with.

It means effectively **leveraging the existing invested efforts in INSPIRE infrastructure**, data harmonisation and provision, and by **staying close to INSPIRE requirements limiting the costs of change**, while giving a boost to social and economic re-use of earth observation and environmental data at a time interest in such data is rising. Taking the existing INSPIRE efforts as starting point additionally means avoiding the introduction of novel requirements to MS concerning data provision, something that some MS explicitly stated was desirable.

There are a few options available to somewhat limit the effort involved, by including only subthemes for some of the mentioned INSPIRE Themes:

- Elevation (Annex II), w.r.t. land and shorelines
- Geology (Annex II), w.r.t. the geophysics subtheme
- Land use (Annex III), w.r.t. existing land use
- Natural Hazards (Annex III), w.r.t. floods and forest fires
- Sea Regions (Annex III), w.r.t. soil inventories

A minimum option would be to limit HVD to the Environmental e-reporting priority list only. While these data have re-use value, and the existence of reporting obligations of themselves create their own re-use demand, most re-use depends on combinations with other data outside the scope of the Environmental e-reporting priority data list.

The thematic area Earth observation and environment encompasses data w.r.t. observing the planet's physical, chemical and biological status over time, collected from earth based remote sensing, and in-situ data collection regarding the environment. (As stated space based earth observation is placed out of scope as only a minority of MS have such capabilities, which where they are the result of PPP the resulting data are not in scope of the Directive generally, while e.g. the Sentinel program is an EU level effort whereas the Directive pertains to MS). Environmental data concern both the status of the environment (in the physical, chemical and biological sense), as well as human activities impacting that status (either administrative or regulatory aspects, such as protective measures or administrative boundaries or allowed levels, as well as interventions in the physical environment such as waste or emissions, or flood prevention activities. This constitutes an extremely broad scope, and a broad and fragmented list of data as a result (MS suggested hundreds of different datasets). It also means a very broad scope of use cases and resulting benefits, both across user groups and sectors including the public sector itself (e.g. Copernicus services depend on in-situ data from MS). Though many usage examples exist, benefits are often hard to quantify especially generally for the thematic area as a whole. Use value is commonly based on the combinations of a variety of datasets from different sources and topics both within and outside the thematic area. This means the thematic area is mostly to be treated as a whole, without attempting a limitative ranking of specific datasets.

INSPIRE and the Environmental e-reporting priority data list cover a similarly broad area, and datasets suggested by stakeholders all fall within its scope. As INSPIRE sets the legal basis for the establishment of an EU spatial data infrastructure for environmental policies and policies which may have an impact on the environment (i.e. earth), **datasets falling under the scope of INSPIRE are of high value by definition**. This results in the finding that adding Earth observation and environmental under the High Value Data provision means adding an open data requirement to existing INSPIRE obligations for the relevant INSPIRE themes and the Environmental e-reporting priority data list. The presented levels of intensity for interventions are based on changing the scope of data sets involved, by including less or more INSPIRE themes, and on varying the publishing dimensions from staying close to existing INSPIRE recommendations or diverging from them.

Box 4 – Validation workshop results: earth observation and environment, overall appreciation of policy intervention options

During the validation workshop organised on 28 July 2020, participants were requested to indicate whether they agreed or disagreed with the three<sup>416</sup> proposed policy options. The earth observation and

<sup>416</sup> In the initial version of this Deliverable, three policy intervention options were considered per thematic area. For the final version of this Deliverable, and upon request of the Commission, the initial three policy options were merged into

environment options received the following appreciations (51 Respondents): Agree: 90% and Disagree: 10%.

In addition, the policy options were evaluated by participants as regards their relevance with regards to the overall environment of the thematic area and the respective needs of the participants. The three options obtained the following scores:

- Low: This option obtained the score of 3.6/10 (49 Respondents).
- Medium: This option obtained the score of 5.6/10 (49 Respondents).
- High: This option obtained the score of 5.8/10 (49 Respondents).

### 3.4.3.1 Lower intensity interventions

The lower intensity interventions limit the scope of data involved to datasets used for reporting by Member States under the environmental acquis (covering topics such as air quality, biodiversity, emissions, nature preservation, noise, waste and water), as documented in the EEA's reporting obligation database<sup>417</sup> and further detailed in the INSPIRE priority data set list for environmental reporting<sup>418</sup>. While these data have re-use value, and the existence of reporting obligations of themselves create their own re-use demand, much re-use depends on combinations with other data outside the scope of the Environmental e-reporting priority data list. The datasets in scope are those that underlie the reporting and monitoring of the implementation of environmental policies, and include the source environmental measurements under the INSPIRE theme Environmental Monitoring Facilities (EMF). (Covering the location of monitoring stations, the parameters measured, as well as the actual spatio-temporal observation data.) The datasets concerned are taken into account as collected (and without regard to the reporting cycles of the reporting obligations they are used for).

In terms of obligations, the lower intensity interventions add an open data obligation to the environmental reporting and observation data, which means mostly removing re-use restrictions and terms of use and adding minimum data provision measures.

<b>Dimensions</b>		<b>Environmental e-reporting priority data and Environmental Monitoring Facilities</b>
Openness-data specification	License (terms of use)	CC0 or CC-BY 4.0 (or equivalent open license); no terms of use)
	Format	No recommendations different from existing INSPIRE specifications
	Machine-readability	Mandatory
	Availability of API, bulk download	Both API and bulk download, as prescribed in e-reporting priority data list
Documentation	Metadata (dataset)	Complete (INSPIRE)

two policy options, a lower and higher intensity options. All elements composing the initial three options were transferred through to the final two options, and as such, the validation of the stakeholders still holds.

<sup>417</sup> <https://rod.eionet.europa.eu/>

<sup>418</sup> <https://inspire.ec.europa.eu/metadata-codelist/PriorityDataset/>

	content description)	
	Data linking	No recommendations different from existing INSPIRE specifications
	Documentation (incl. structure and semantics)	Complete and web-available (INSPIRE)
	Shared vocabularies	No recommendations different from existing INSPIRE specifications
	Taxonomies	No recommendations different from existing INSPIRE specifications
Completeness	Traceability	No recommendations different from existing INSPIRE specifications
	Update frequency and timeliness	As collected, for EMF highest collected temporal resolution
	Granularity	Highest collected temporal and geographic resolution
	Key attributes	All attributes mentioned in e-reporting priority data list, and EMF locations, parameters measured and complete observations data

### 3.4.3.2 Higher intensity interventions

For the higher intensity interventions the scope of the data involved is broadened by adding the additional INSPIRE themes relevant to earth observation and environment to the data and themes covered in the lower intensity interventions (Environmental e-reporting priority data list, and the INSPIRE Theme EMF). This allows the type of combinations that re-use value is generally build on within this thematic area, and that were found across the varied use cases encountered in the study. These interventions add open data requirements to INSPIRE (download) services. Where INSPIRE harmonises the data itself, these interventions harmonise data provision and re-use. As such they remove restrictive terms of use, as well as remove fees (e.g. for land parcels, ortho-imagery and elevation models), and add open licenses. In terms of formats and data provision methods it follows current INSPIRE standards and recommendations and does not deviate from them.

INSPIRE Themes placed within scope are:

- Hydrography (Annex I)
- Land parcels (Annex I), limited to geometries, parcel type and parcel code, as per the Geospatial thematic area's interventions for Cadastral parcels.
- Protected sites (Annex I)
- Elevation (Annex II)
- Geology (Annex II)
- Land cover (Annex II)
- Ortho-imagery (Annex II), excluding satellite sensor derived data
- Bio-geographical regions (Annex III)
- Environmental Monitoring (Annex III) (already part of the lower intensity interventions, mentioned for completeness)
- Habitats/Biotopes (Annex III)
- Land use (Annex III)

- Natural Hazards (Annex III)
- Oceanography (Annex III)
- Sea Regions (Annex III)
- Soil (Annex III)
- Species Distribution (Annex III)

Some topics relevant to this thematic area that were explicitly put forward by stakeholders, such as forestry and coastal vulnerability are covered within multiple INSPIRE themes included here.

Some data within scope (e.g. LIDAR scans of an entire MS) may currently not be in existence in a majority of MS, but the findings suggest that this data will be collected when it is feasible, and a current absence is more an expression of currently limited technological capabilities and resources, which are likely to change over time. They are nonetheless included as the PSI Directive does not mean an obligation to collect certain data, but regulates its availability if it is collected, while where such data is currently available valuable re-use cases clearly arise.

Differences with the table for publishing dimensions for the low intensity interventions are marked in blue.

<b>Dimensions</b>		<b>Environmental priority data</b>	<b>e-reporting</b>	<b>INSPIRE Themes</b>
Openness-data specification	License (terms of use)	Creative Commons 0 or CC-BY 4.0; No terms of use		
	Format	As prescribed in e-reporting priority data list	Existing recommendations	INSPIRE
	Machine-readability	Mandatory	Mandatory	
	Availability of API, bulk download	Both API and bulk download, as prescribed in e-reporting priority data list	INSPIRE download services (e.g. WFS or OGC:API) and bulk download	
Documentation	Metadata (dataset content description)	Complete (INSPIRE)	Conform INSPIRE, or DCAT-AP2	
	Datalinking	No recommendations from existing specifications	different from INSPIRE	Conform INSPIRE
	Documentation (incl. structure and semantics)	Complete and (INSPIRE)	web-available	Conform INSPIRE
	Shared vocabularies	No recommendations from existing specifications	different from INSPIRE	Conform INSPIRE

	Taxonomies	No recommendations from existing specifications	different INSPIRE	<a href="#">Conform INSPIRE</a>
Completeness	Traceability	No recommendations from existing specifications	different INSPIRE	<a href="#">Conform INSPIRE</a>
	Update frequency and timeliness	As prescribed in priority data list	e-reporting	<a href="#">Conform INSPIRE</a>
	Granularity	As prescribed in priority data list	e-reporting	<a href="#">Conform INSPIRE</a>
	Key attributes	All attributes mentioned in reporting priority data list	e-reporting	<a href="#">All features within INSPIRE Theme's scope</a>

Some suggested publishing dimensions are moving ahead of (but not diverging from) currently fully implemented INSPIRE recommendations, and suggested by both re-users and data holders. As such they would be a temporary acceleration if implemented, which would potentially add additional costs for API development and data transformation. So while not making these publishing dimensions mandatory, their encouragement is nonetheless of value. Subsequently as INSPIRE recommendations catch up, future changes in INSPIRE recommendations would be followed thereafter:

- Using CC0 as **license**, and removing any other re-use restriction. For open data in general CC0 is preferable, although CC-BY 4.0 and equivalent national open licenses can often be acceptable as well without being detrimental to re-use potential. Specifically however when combining datasets from across MS differences in licenses quickly create high friction for re-use, by needing to satisfy a variety of licensing conditions and terms of use. INSPIRE's main import is its harmonising effect, and the suggested Earth observation and environment HVD adds open data requirements to INSPIRE. Encouraging the use of CC0 only licenses takes the notion of harmonisation to its logical endpoint in open data. The use of CC0 provides a clear, unambiguous and globally recognised signal to re-users on the conditions of re-use.
- With regard to **API's** WFS is currently common within INSPIRE, while discussions and work are taking place to move to newer API specifications (RESTful, and specifically OGC APIs), which the higher intensity interventions encourage.
- Similarly, with regard to **metadata** the currently discussed DCAT-AP2 metadata standard for INSPIRE is suggested for adoption. This does still require the resolution of conflicting requirements between DCAT-AP2 and Geo-DCAT as currently in use within INSPIRE.

### 3.5 Statistics

This section presents the micro-level assessment for the thematic area of statistics. It illustrates the current state of play of the provision of these datasets. Furthermore, it provides the recommended measures for publication together with the costs and benefits of including these datasets as High Value Datasets under the PSI Directive. Lastly, it details the three policy options proposed for this thematic area.

### 3.5.1 As-is situation: how Member States provide these datasets today

The official statistics are derived from data produced by public bodies as part of their official function and they provide a record of the social, economic and environmental condition of the country. The main gateway for national data to the outside world is the National Statistical Institutes (NSI), which can lead or co-ordinate the system of data collection entities in various domains. The NSI also ensures that all data provided have the quality required by the statistical standards.

The openness of the datasets in scope is not a significant issue in the EU Member States. In all countries the datasets are publicly available and free of charge. When it comes to licences used, only eight Member States publishes the data explicitly under CC BY licence. However, in the rest of the Member States, the NSI only include in their licence conditions the requirement to be mentioned as source of the data - if data are used unprocessed they are the main reference; if the data have been modified the expression “based on data from NSI” should be included. There are several formats used to provide the datasets. The most frequent format used is XLS or XLSX (26 countries), CSV (22 countries), JSON and PX formats (10 countries), XML (12 countries) and SDMX (seven countries). There are some variations across different datasets when it comes to the format available, but often Member States includes more than one format. Two countries (Cyprus and Greece) are exceptions of these rules, as they provide data only in XLS format.

In only four country the availability of the machine-readable format is not clearly provided, while in two countries (Cyprus, Greece) is not available for any of the datasets in scope. In 16 out of 27 countries, the APIs tools are explicitly available, while in five countries they are not available at all. For example, Denmark<sup>419</sup> includes on the NSI websites a specific section related to APIs development and use. Ireland<sup>420</sup> and Netherlands<sup>421</sup> provide a short overview on APIs availability and include a direct link to the open data portal section containing the datasets downloadable via APIs. In Austria neither APIs nor bulk-download is allowed on the STATCube data, but this type of access is available for the datasets uploaded on the Open Data Portal of Austria. In four Member States the APIs and bulk-download availability is not clear.

To be able to use and re-use data a well-developed and user-friendly documentation is needed. All Member States include in their NSI’s websites detailed documentation, including information on structure and semantics of data. Classifications used as well as the definitions are part of these documentations. In addition to the general documentation, the metadata are another key information included with the datasets. Except for Cyprus, all the other Member States include the metadata information with the corresponding data. Often, they are web-based format available, but some countries offer them in CSV, SDMX formats or just in PDF file. When it comes to shared vocabularies (or taxonomies) and data linking, the information is more difficult to identify. However, for the datasets uploaded on the Open Data portals the availability of this type of information becomes crucial for their discoverability and further use and re-use.<sup>422</sup>

When it comes to timeliness and frequency of the data, several differences have been identified across Member States. In terms of frequency, often the most common occurrences are annual and quarterly data updates. For specific datasets, monthly updates might also be available across countries. In general, all countries provide annual data, but the latest available year differs between both datasets and providers. The desk research showed that most of the datasets are available for 2019 (demography: 17 countries;

<sup>419</sup> <https://www.dst.dk/en/Statistik/statistikbanken/api>

<sup>420</sup> <https://statbank.cso.ie/webserviceclient/>; <https://statbank.cso.ie/webserviceclient/DatasetListing.aspx>

<sup>421</sup> <https://www.cbs.nl/en-gb/our-services/open-data/statline-as-open-data>;

[https://opendata.cbs.nl/statline/portal.html?\\_la=en&\\_catalog=CBS](https://opendata.cbs.nl/statline/portal.html?_la=en&_catalog=CBS)

<sup>422</sup> Stakeholder interviews.

labour market: 24; GDP: 27, and government finances 17). Preliminary data for demography is available for 2020 in four countries (Finland, France, Luxembourg and Romania), while in other six countries the latest available year is 2018. In three countries the latest annual labour market data refers to 2018, and in 11 countries the government finances include data for 2018 as most recent. When it comes to quarterly data, this is also often available for the datasets in scope, but not all the countries include them (demography: 14 countries; labour market: 26; GDP: 27; government finances: 24). The fourth quarter of 2019 is the latest data for demographic dataset in 11 countries, while two countries provide data only for the first quarter of 2019 (Sweden and Austria), and other two countries have data at semester level (Poland and Spain). The labour market data are available for the fourth quarter of 2019 in 22 countries, for the third quarter 2019 in Malta and Romania, and for the first quarter of 2020 in Sweden and Spain. The latest quarter available for government finances dataset is fourth quarter of 2019 in 17 countries, and the third quarter in other seven countries.

In terms of coverage, all datasets are available at national level. Often, a regional (NUTS 2 disaggregation) is also included in datasets provisions, but its availability varies across datasets and countries (demography: 26 countries; labour market: 21; GDP: 22). When it comes to government finances, they are often disaggregated at institutional sector level such as central and local governments, and social security funds levels (11 countries). In some particular cases, a lower disaggregation (such NUTS 3 or municipality levels) is available for datasets. In 13 countries, demographic data provides information at NUTS 3/municipality level. Only three countries do so for the labour market dataset and for GDP.

When it comes to demographic dataset, there are several breakdowns and key variables available for all Member States. The most common breakdowns are sex, age and five-years age groups, place of residence. When it comes to key variables, these are life expectancy, live births, stillborn, deaths, fertility (total and age-related) rates, mortality rates, population density, internal and external migration. Some countries include some additional breakdowns such as citizenship, education level, ethnicity. Similarly, the labour market dataset includes often breakdowns such as sex, age (five-years groups), place of residence, education level, duration (for unemployment). Frequent key variables are labour cost, wages and salaries, job vacancies, employment and unemployment rates. Countries might also include breakdowns by occupations, economic sectors, nationality (national vs. foreigner groups). The most common variables for the GDP are the gross value-added, consumption (intermediary and final), exports and imports, taxes and subsidies, compensation of employees. The government finances provide information on revenues and expenditures, taxes, social contributions and social benefits. Most of the countries also include breakdown by government function of expenditures.

The table below provides a countries' overview of all information gathered on today's modes of provision of statistical information across Member States. The table makes the distinction between the different datasets included in the analysis. However, as the data usually have the same provider (NSI) there are significant overlapping of the information detailed below. Nevertheless, the table clearly indicates the differences and the gaps identified during the analysis.



Openness-data specification

	<b>License (terms of use)</b>	<b>Free of charge</b>	<b>Format</b>	<b>Machine-readability</b>	<b>Availability of API, bulk download</b>
<b>Austria</b>	Source acknowledgement @Statistics Austria	Partially (STATCube: limited access as guest, subscription fee for broader access)	XLSX/XLS, CSV (+database style), XML (SDMX), zip, TXD	Not available for STATcube; available on the open data portal	Automatic retrieval systems (bots or crawlers...) are not allowed for retrieving data from STATcube.
<b>Belgium</b>	@ NBB (National Bank of Belgium) for financial and fiscal datasets, CC BY 4.0 @Statistics Belgium for demography, labour market	yes	XLS/XLSX, CSV, PX, SDMX (XML)	Partially available	Few APIs are available for specific tables on be.STAT cube. Not clear for other official statistics providers.
<b>Bulgaria</b>	Source acknowledgement @Statistics Bulgaria	partially (subscription fee for higher granularity)	PDF, XLS, CSV; SDMX	available	no
<b>Croatia</b>	Source acknowledgement @NSI	yes	XLSX/XLS; PX-Axis	not clear	not clear
<b>Cyprus</b>	Source acknowledgment @NSI	yes	XLSX	no	no
<b>Czech Republic</b>	Source acknowledgement @Czech Statistical Office	yes	XLSX; XML, PDF (public database), CSV (open data)	Available on the open data portal; not clear for the NSI website	no
<b>Denmark</b>	CC BY 4.0; Source acknowledgment @Statistics Denmark.	yes	CSV, JSON, TXT, XLS/XLSX, SAS	available	Allows user selection, APIs and bulk download
<b>Estonia</b>	Source acknowledgement @Statistics Estonia	yes	CSV, JSON, XLS, PX, XML	available	APIs available (XML, JSON APIs)
<b>Finland</b>	CC BY 4.0; source accreditation to Statistics Finland	yes	XLSX, XML, JSON, JSON-stat, CSV, PX	available	yes; allows user selection, APIs and bulk download
<b>France</b>	Source acknowledgment to www.insee.fr, in clear and in full, when using data	yes (except the ones under IPR conditions)	CSV, XLS	available	APIs (SDMX); time-series bulk download available (CSV)
<b>Germany</b>	Data license Germany - attribution - version 2.0" ("dl-de / by-2-0", www.govdata.de/dl-de/by-2-0)	yes	XLSX, CSV, XML	available (CSV flat file)	APIs and batch processing available
<b>Greece</b>	Source acknowledgment @Statistics Greece (with link in clear - www.statistics.gr)	yes	XLS, PDF	no	no

Hungary	Source acknowledgement @Statistics Hungary	yes	XLS, SSQ	yes	APIs (SSQ)
Ireland	CC BY 4.0; Copyright of The Government of Ireland	yes	CSV, TXT, PX, XLSX, HTML, JSON	yes	APIs (JSON)
Italy	CC BY 3.0; source acknowledgment	yes	CSV, XLS, PX, XML (SDMX)	yes	yes (APIs SDMX based); allows user selection
Latvia	Source acknowledgement of data @Central Statistical Bureau	yes	PX, CSV, HTML, XML, XLSX, JSON	yes	not clear
Lithuania	Source acknowledgement @Statistics Lithuania (dissemination policy)	yes	XML, JSON (SDMX), CSV, TXT, XLSX/XLS, PDF, HTML	yes	APIs (SDMX - JSON based)
Luxembourg	CC0; source acknowledgement @STATEC	yes	XLSX, CSV, XML	yes	APIs - RSS apps (XML based)
Malta	Source acknowledgement ©National Statistics Office	yes	SDMX-ML, XLS, PDF, HTML	yes (SDMX)	not clear
Netherlands	CC BY 4.0 licence; Source acknowledgement to Statistics Netherlands	yes	HTML, CSV,	yes	APIs and feed (bulk download)
Poland	Source acknowledgement @Statistics Poland	yes	CSV, XML, XLSX, PDF, WORD, MHTML, TIFF	yes	APIs available <sup>423</sup>
Portugal	CC BY 4.0; Source acknowledgement: ©Instituto Nacional de Estatística	yes	XLSX, CSV	not clear	APIs available <sup>424</sup>
Romania	Licence for Open Government 2.0; Source acknowledgement @Statistics Romania	yes	XLSX, CSV, SDMX-ML	no	no
Slovakia	CC BY 4.0; Source acknowledgement © Statistics Slovakia	yes	XLSX, PDF (DATAcube); XLSX, XML, CSV, HTML, PDF (STATdat)	DATAcube: no STATdat: yes	not clear
Slovenia	Source acknowledgement @Statistics Slovenia	yes	PX, CSV, XLSX, JSON	yes	APIs available
Spain	Source acknowledgement @Statistics Spain	yes	XSLX, CSV, PX, JSON, TXT	yes	APIs available
Sweden	CC BY 4.0; source acknowledgment @Statistics Sweden	yes	PX, CSV, TXT, XLSX, JSON-STAT, HTML5, JSON	yes	APIs available

<sup>423</sup> <https://api.stat.gov.pl/Home/Index>

<sup>424</sup> [https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine\\_api&INST=322751522&ine\\_smenu.boui=357197120&ine\\_smenu.selected=357197822&xlang=en](https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_api&INST=322751522&ine_smenu.boui=357197120&ine_smenu.selected=357197822&xlang=en)

Documentation						
	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)			Shared vocabularies
Austria	Available (linked to data; overall document in PDF)	NA	Available	on	the	NA
Belgium	Available (linked to data; overall document in PDF)	NA	Available	on	the	NA
Bulgaria	Available (linked to data, in SDMX format)	NA	Available	on	the	NA
Croatia	Available (linked to data; overall document in PDF)	NA	Available	on	the	NA
Cyprus	Not available	NA	Available	on	the	NA
Czech Republic	Available (linked to data)	NA	Available	on	the	NA
Denmark	Available (linked to data; in both HTML and CSV formats)	NA	Available	on	the	NA
Estonia	Available (linked to data)	NA	Available	on	the	NA
Finland	Available (linked to data)		Available	on	the	Keywords included within the metadata
France	Available (linked to data)	NA	Available	on	the	NA
Germany	Available (linked to data)	NA	Available	on	the	NA
Greece	Not available	NA	Available	on	the	NA
Hungary	Available (linked to data)	NA	Available	on	the	NA
Ireland	Available (linked to data)	NA	Available	on	the	NA
Italy	Available (linked to data)	NA	Available	on	the	NA
Latvia	Available (linked to data)	NA	Available	on	the	NA
Lithuania	Available (linked to data)	NA	complete	and	web	NA
Luxembourg	Available (linked to data)	NA	Available	on	the	NA
Malta	Available (linked to data)	NA	Available	on	the	NA
Netherlands	Available (linked to data, also provided in CVS format)	NA	Available	on	the	NA
Poland	Available (linked to data)	NA	Available	on	the	NA
Portugal	Available (linked to data)	NA	Available	on	the	NA

Romania	Available (linked to data)	NA	Available website	on	the	NA
Slovakia	Available (linked to data)	NA	Available website	on	the	NA
Slovenia	Available (linked to data)	NA	Available website	on	the	NA
Spain	Available (linked to data)	NA	Available website	on	the	NA
Sweden	Available (linked to data)	NA	Available website	on	the	NA

Completeness				
	Traceability	Frequency and latest available data	Coverage	Available dimensions and breakdowns
Austria	Source of data is provided on the NSI website	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018), quarterly (2019q3); Labour market: annual (2019), quarterly (2019q4) Demography: annual (2019), quarterly (regional: 2019q1, national: 2019q4)	GDP: national, regional level (NUTS2 (2018), NUTS3 (2017)) Government finances: national, general, central, state, local level, social security funds Labour market: national, regional (NUTS2 level) Demography: national, regional (NUTS2, NUTS3)	GDP: output, input and final expenditure approaches provided; economic sector breakdown available for gross value added - NACE rev2 (A10) input/final expenditure - total only; regional, annual: NACE rev2 (A, [BDE], C, F, [G-J], [K-N], [O-T]), NUTS2, NUTS3 Government finances: expenditure and revenues, assets, liabilities, taxes, with breakdowns by ESA categories, regions Labour market: activity rate, type of employment, second employment, unemployment, job vacancies, economic sectors, occupations, gender, age groups, education level, household size Demography: population, births, deaths, marriages, divorces, migration; breakdown by sex, age (years and 5 and 15 years groups), nationality, country of birth;
Belgium	Source of data is provided on the NSI website and a link is provided to the original provider (e.g. Bureau fédéral du Plan, National Bank of Belgium, Federal Public Service Employment, Labour and Social Dialogue)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4) Demography: annual (2019)	GDP: national, regional level Government finances: national, public administration level (federal, regional) Labour market: national, regional (NUTS2 level) Demography: national, regional (NUTS2 level)	GDP: expenditure (final consumption - private, general government; gross fixed capital formation, exports and imports), income (compensation of employees, gross surplus and mixed income, taxes), production (gross value added, taxes and subsidies); breakdown by economic sectors Government finances: revenues (direct and indirect taxes, social contributions, transfers, other) and expenditure (compensation of employees, subsidies, social benefits, transfers, other); breakdowns include government functions (COFOG), non-financial accounts Labour market: employed, unemployed, inactive, job vacancies, wages and labour costs; breakdown by sex, age groups, place of residence, level of education, economic sectors (NBB, FPSELS - xls), Demography: population, births, deaths, migration; breakdowns by sex, age groups, place of residence, marital status, nationality (Belgian/non-Belgian)
Bulgaria	Source of data is provided on the NSI website (details included in the metadata files)	GDP: annual (2019); quarterly (2019q4) Government finances: annual (2018), quarterly (2019q3) Labour market: annual (2019 provisional), quarterly (2019q4), monthly (employees, wages and salaries - under short-term stats) Demography: annual (2019)	GDP: national, regional level Government finances: national, general, central, local level, social security funds Labour market: national, regional level Demography: national, regional, districts, municipalities, cities	GDP: output, input and final expenditure approaches provided, breakdown by economic sectors (A10, A3 for regional level) available Government finances: expenditures, transactions, taxes; breakdowns include categories, subsector, government functions (COFOG) Labour market: employment, unemployment, job vacancies, labour costs, earnings; breakdowns by sex, age, education level, economic sectors, place of residence Demography: population, life expectancy, fertility rate, mortality rate, internal and international migration; breakdowns by gender, place of residence, age (years and age groups)

Croatia	Source of data is provided on the NSI website	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018) Labour market: annual (2019), quarterly (2019q4), monthly (December 2019) Demography: annual (2018)	GDP: national, regional (NUTS 2, NUTS 3 - 2017) Government finances: national Labour market: national Demography: national, regional, municipalities/towns level	GDP: expenditure, production approaches, gross value added, general government, households; economic sectors breakdown available Government finances: investments (fixed assets, by type of fixed assets), financial source, regions, economic sectors Labour market: employment, unemployment, activity status; breakdowns by age groups, sex, economic sectors, education level Demography: population, births, deaths, fertility rates, mortality rates, migration (in-, out-), population estimates breakdowns by sex, age (years and 5-years groups), place of residence (regions, municipalities/towns),
Cyprus	Source of data is provided on the NSI website (including the original source, e.g. Central Bank of Cyprus, Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018), quarterly (2019q3) Labour market: annual (2018), quarterly (2019q4) Demography: annual (2018)	GDP: national Government finances: national Labour market: national Demography: national, regional	GDP: production, expenditure, income approaches available (government final consumption, private consumption, GFC, exports and imports); economic sectors breakdown provided, Government finances: government final consumption, households, gross capital formation, exports and imports; breakdown by government function (COFOG) available (data for 2017) Labour market: employment, unemployment, job vacancies, activity status, breakdowns by economic sectors, occupations, sex, education level, age groups Demography: population, births, deaths, marriages, divorces, migration; breakdowns by sex, age (5-years groups), place of residence (including rural/urban disaggregation)
Czech Republic	Source of data is provided on the NSI website (including the original source, e.g. open data portal, Labour Force Survey, Ministry of Labour and Social Affairs)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018), quarterly (2019q3) Labour market: annual (2019), registered 31.12.2019), quarterly (2019q4) Demography: annual (2019), quarterly (2019q4)	GDP: national, regional (2018) Government finances: national, administrative level Labour market: national, regional (NUTS2), districts, municipalities (MLSA) Demography: national, regional, districts, municipalities (2018)	GDP: output, input and final expenditure approaches available, Government finances: revenue, expenditure, assets, liabilities, transactions; breakdowns by financial source, sector level (general, central, local, social security funds), government functions (COFOG) Labour market: employment, unemployment, activity status, job vacancies, job applicants; breakdowns by sex, place of residence, age groups, occupations, economic sectors, Demography: population, births, deaths, immigrants, emigrants, marriages, divorces; breakdowns by sex, place of residence, age (years and age groups)
Denmark	Source of data is provided on the NSI website (including the original source, e.g. open data portal)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020) (varies across disaggregation levels) Demography: annual (2019), quarterly (2019q4), monthly (January 2020) (varies across	GDP: national, regional Government finances: national, regional (NUTS 2), municipalities Labour market: national, regional (NUTS 2, NUTS 3), municipalities Demography: national, regional (NUTS 2, NUTS 3), municipalities	GDP: production, expenditure, income approaches available (government final consumption, private consumption, GFC, exports and imports), economic sectors breakdown provided, Government finances: expenditures and revenues, assets and liabilities; breakdowns by financial instruments, sector accounts, public corporations, public sector, budgets Labour market: activity status, employment, unemployment, job vacancies; breakdowns by region, socio-economic status, sex, age (5-years groups), level of education, country of origin, Demography: live births, deaths, internal/external migration, net migration, natural increase; breakdowns by age (years and

		disaggregation levels)		5- and 10-years groups), sex, marital status, place of residence (incl. urban/rural), country of origin, citizenship, ancestry, place of birth
Estonia	Source of data is provided on the NSI website	GDP: annual (2019); quarterly (2019q4) Government finances: annual (2019) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020 preliminary) Demography: annual (2019), monthly (February 2020 preliminary)	GDP: national, regional Government finances: national, regional (partially) Labour market: national, regional (NUTS2) Demography: national, regional (NUTS2, NUTS 3)	GDP: income, expenditure, production approaches available (fixed capital, final consumption, household final consumptions); economic sectors breakdown provided, Government finances: revenues, taxes, budgets, expenditures, assets and liabilities; breakdowns by financial accounts, public corporations, social contributions Labour market: employment, activity status, unemployment; breakdown by age (5-years groups), sex, education level, economic sectors, place of residence, ethnic nationality, Demography: population, births, deaths, population estimates, fertility rates, mortality rates, migration (in-, out-), life expectancy, marriages, divorces; breakdowns by age (years and 5-years groups), place of residence (regions, municipalities, towns), education level
Finland	Source of data is provided on the NSI website (including the original source, e.g. Ministry of Economic Affairs and Employment)	GDP: annual (2019 prelim); quarterly (2019q4) Government finances: annual (2019 prelim); quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020; MEAE) Demography: annual (2019, 2020 provisional, recorded on 31.12.2019), quarterly (2019q4)	GDP: national, regional (2017) Government finances: national Labour market: national, regional Demography: national, regional (NUTS2, NUTS 3), municipalities	GDP: income, expenditure, production approaches available (gross value added, fixed capital, final consumption, household final consumptions); economic sectors breakdown provided Government finances: expenditures, revenues, assets and liabilities; breakdowns by financial instruments, sector accounts, financial and non-financial corporations, general government, households, budgets; general, central, local government, social security funds Labour market: activity status, employment, job vacancies, jobseekers, unemployment, training participations; breakdowns by economic sectors, place of residence, occupations, age groups, sex Demography: population, births, deaths, population change, migration, adoptions; breakdowns by sex, age (years and 5-years groups), place of residence (incl. rural/urban), citizenship, country of birth, marital status
France	Source of data is provided on the NSI website	GDP: annual (2019); quarterly (2019q4) Government finances: annual (2018); quarterly (2019q4) Labour market: annual (2019); quarterly (2019q4) Demography: annual (2020 provisional); monthly (February 2020)	GDP: national Government finances: national Labour market: national Demography: national, regional (departments, 2018)	GDP: income, production, final consumption approaches available (fixed capital, households final consumption, expenditures, imports and exports), economic sectors breakdown provided, Government finances: expenditures and revenues (compensation of employees, subsidies, intermediate consumption, property income, capital transfers, social benefits, taxes, acquisitions on non-financial assets, other transfers); breakdown by government functions and type of operations Labour market: activity rate, employment (total and rates), unemployment (total and rates), underemployment, active life expectancy; breakdowns by economic sectors, place of residence, type of contracts, sex, age groups (5 years, other), Demography: population, births, live births, deaths, population

				change and structure, average and median age of population, marriages and divorces; breakdowns by sex, age groups (5 years, other),
Germany	Source of data is provided on the NSI website (including the original source, e.g. Federal Employment Agency)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018), quarterly (2019q3) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020) Demography: annual (2019), quarterly (2019q4), monthly (March 2020)	GDP: national, regional (landers, 2018) Government finances: national, federal states Labour market: national, federal states Demography: national, regional (federal states, administrative districts)	GDP: gross value added, taxes and subsidies, gross fixed capital, exports and imports, households' disposable income; breakdown by economic sectors (A10) Government finances: expenditure, income, financial balance, transactions, investments, corporate groups, taxes, financial assets Labour market: employment, unemployment, job vacancies, activity status; breakdowns by economic sectors, sex, education level, age groups, Demography: population, births, deaths, marriages and divorces, migration, breakdowns by sex, age (years and 5-years groups), place of residence, nationality,
Greece	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018), quarterly (2019q3) Labour market: annual (2019), quarterly (2019q4), monthly (December 2019) Demography: annual (2019)	GDP: national, regional (NUTS 2, NUTS 3 - 2017) Government finances: national Labour market: national, regional (NUTS 2) Demography: national, regional (NUTS 2, NUTS 3), municipalities	GDP: production, expenditure and income approaches available [outputs, intermediate consumption, taxes and subsidies; final consumption, capital formation; households and general government, exports and imports, gross value added]; breakdown by economic sectors (A10) Government finances: government surplus/ deficit and debt levels, central, state, local government, social security funds; compensation of employees, subsidies, intermediate consumption, property income, capital transfers, social benefits, taxes, acquisitions on non-financial assets, other transfers Labour market: employment, unemployment, job vacancies, activity status; breakdown by sex, age groups (5 years), nationality, economic sector, education level, place of residence, Demography: population, births, deaths, migration, marriages, fertility index; breakdown by age groups (5 years), sex, nationality, place of residence,
Hungary	Source of data is provided on the NSI website [e.g. STADAT, Dissemination database (2018)]	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018), quarterly (2019q4) Labour market: annual (2019) Demography: annual (2019), quarterly (2019q1), monthly (January 2020)	GDP: national, regional (monthly - January 2020, production by economic branches) Government finances: national, regional (monthly - Jan 2020, production by economic branches) Labour market: national, regional Demography: national, regional	GDP: production, expenditure, income approach available (government final consumption, private consumption, GFC, exports and imports), economic sectors (A3 and A10) breakdowns provided, Government finances: expenditures and revenues, output, value added, taxes, subsidies, fixed capital; breakdowns by government functions (COFOG), institutional levels (general, central, state, local government, social security funds), financial and non-financial corporations, households Labour market: employment, unemployment, activity status, job vacancies (2018), labour costs, earnings (average and net); breakdowns by sex, age groups, economic sectors, occupations, duration (unemployment), education level Demography: population, marriages, divorces, births, deaths, fertility rates, average life expectancy, migration (internal and international); breakdowns by gender, age groups (5 years),



					place of residence, marital status,
Ireland	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020) Demography: annual (2019)	GDP: national, regional (2017) Government finances: national, general Labour market: national, regional Demography: national, regional		GDP: gross value added, taxes, subsidies, income, expenditure, capital formation, exports and imports; breakdown by economic sectors Government finances: transactions, revenues and expenditure, assets and liabilities, net worth, debt (gross/net) breakdowns by ESA2010 financial items; Labour market: employment, unemployment, average hours worked, earnings (net/averages); breakdowns by sex, age group, place of residence, education level, economic sector, nationality, marital status, occupations, type (full-time/part-time), duration Demography: population, births, deaths, marriages, migration; breakdowns by sex, age (years and 5-years groups), place of residence, nationality, education level
Italy	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4) Demography: annual (2019), quarterly (2019q4), monthly (February 2020)	GDP: national, regional (2018) Government finances: national, general Labour market: national, regional Demography: national, regional (NUTS 2, NUTS 3), municipalities		GDP: output (gross value added, taxes, subsidies), expenditure (final consumption, gross capital formation, imports exports), income (disposable income for households) approaches available; economic sectors breakdown included, Government finances: expenditure (compensation of employees, intermediate consumption, social benefits, and interest) and revenue (taxes and social contribution) Labour market: employment, unemployment, underemployed, inactivity, full- and part-time employment, wages and labour costs; breakdowns by sex, age groups (5 years), citizenship, household role (parent, daughter, son etc), education level, economic sectors, professional status Demography: population, births, deaths, marriages, divorces, migration, fertility rate, life expectancy (at birth, at 65 years old), mean age of the population; breakdowns by sex, age, marital status, citizenship, education level
Latvia	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey, State Employment Agency (registered unemployment))	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2018), quarterly (2019q3) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020) Demography: annual (2019), quarterly (2019q4), monthly (February 2020)	GDP: national, regional (2017) Government finances: national, administrative level Labour market: national, regional Demography: national, regional (county, municipality levels)		GDP: production (gross value added, taxes and subsidies), expenditure (final consumption - households, government, fixed capital formation, exports and imports), income (compensation of employees, mixed income, taxes on production and imports, subsidies) approaches available; economic sectors breakdown included Government finances: revenues, expenditure, market output, taxes, subsidies, social contributions, compensation of employees; by government function (COFOG), assets, liabilities, transactions, financial accounts Labour market: gender, age groups (5 years), labour status (employed, unemployed, inactive), type of employment, place of residence, economic sectors, occupations, education level, ethnicity, source of income, wages, labour costs, vacancies, working hours

				Demography: population, life expectancy, marriages, divorces, migration; breakdowns by sex, age groups (5 years), country of birth, citizenship, ethnicity, marital status, place of residence
Lithuania	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4) Demography: annual (2019), monthly (February 2020)	GDP: national, regional (2018) Government finances: national, administrative levels Labour market: national, regional Demography: national, regional (county levels)	GDP: production (gross value added, taxes and subsidies), expenditure (final consumption, gross capital formation, exports and imports) and income (compensation of employees) approaches available; economic sectors breakdowns included (2018/2017) Government finances: expenditure and revenues, assets, liabilities / transactions, balance, taxes and social security contributions; breakdowns include institutional level (general, central, local government, social security funds), functions of government (COFOG); Labour market: activity status, employment, unemployment, long-term unemployment, job vacancies, wages and labour costs, average age of employed persons; breakdowns by sex, age groups (5 years), place of residence (counties, urban/rural), economic sectors, education level, Demography: population, fertility, mortality, marriages, divorces, migration, median age of population; breakdowns by sex, place of residence (incl. rural/urban), ethnicity;
Luxembourg	Source of data is provided on the NSI website (including the original source, e.g. National Employment Agency (ADEM))	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020 prelim.) Demography: annual (2020/2019), quarterly (2019q4)	GDP: national Government finances: national, administrative level - general government Labour market: national, regional (cantons, municipalities - 2018) Demography: national, regional (cantons, municipalities)	GDP: production (output, intermediate consumption, gross value added, taxes), expenditure (gross capital formation, final consumption, exports, imports), income (compensations, taxes, mixed income) approaches available; economic sectors breakdown included Government finances: expenditures and revenues, taxes and social contributions, transactions, deficit/surplus Labour market: employment, unemployment, gender, citizenship, economic sector, country of residence, frontier workers, job vacancies, unemployment duration, education level (selected) Demography: population, population density, births, deaths, marriages, divorces, migration; breakdowns by sex, age (years and 5-years groups), nationality,
Malta	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2018), quarterly (2019q3) Demography: annual (2018)	GDP: national Government finances: national Labour market: national Demography: national, regional	GDP: production (output, intermediate consumption, gross value added, taxes), expenditure (gross capital formation, final consumption, exports, imports), income (compensations, taxes, mixed income) approaches available Government finances: revenues and expenditure (by categories), government debt Labour market: labour status (employment, unemployment, inactivity); breakdowns by sex, age group (10 years), employment type (employees/self-employed), economic sectors (A3), education level Demography: population, crude birth rate, crude death rate, total fertility rate; breakdowns by sex, place of residence,

Netherlands	Source of data is provided on the NSI website	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4) Demography: annual (2019), quarterly (2019q4), monthly (February 2020)	GDP: national, regional (2018) Government finances: national, administrative level (incl. regional 2018) Labour market: national, regional (province level) Demography: national, regional	GDP: expenditure (final expenditure, exports, imports, gross fixed capital formation, final consumption - general government, households), output (gross value added by economic sectors, intermediate consumption, taxes and subsidies) and income (compensation of employees, gross operating surplus / mixed income, taxes and subsidies) approaches available, economic sectors breakdown included Government finances: balance sheets (assets and liabilities - financial and non-financial), transactions, expenditures and revenues, taxes and social contributions, social benefits; breakdown includes government functions (COFOG) Labour market: labour status (employed, unemployed, inactive), working hours, wages and labour costs, job vacancies; breakdowns by sex, age groups, type of employment, education level, economic sectors Demography: population, live births, deaths, migration, marriages, divorces, average age of the population; breakdowns by sex, age groups (10 years), nationality, marital status, country of origin, place of residence
Poland	Source of data is provided on the NSI website (including the original source, e.g. Macro-Economic databank (MDB); Knowledge database (DBW - SWAID), Local databank (BDL))	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4), monthly (February 2020) Labour market: annual (2019), quarterly (2019q4) Demography: annual (2018), half-year (cumulative 2019q1 and q2),	GDP: national, regional (2017) Government finances: national, administrative levels (districts) Labour market: national, regional Demography: national, regional	GDP: production, expenditure and income approaches available (final expenditure, exports and imports, capital formation, gross value added, final consumption and domestic demand); economic sectors breakdown included, Government finances: investments, expenditure, revenues, taxes, balance, deficit; breakdowns include socio-economic area, funds from EU to finance programmes and projects, type of expenditure Labour market: activity status (employed, unemployed, inactive); registered unemployment, job vacancies, work time; breakdowns by sex, age (5- and 10-years groups), place of residence, level of education, economic sectors Demography: population, births, deaths, migration, marriages, divorces, median age, life expectancy, population density; breakdowns by sex, age (years and 5-years groups), place of residence (incl. urban/rural),
Portugal	Source of data is provided on the NSI website	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020) Demography: annual (2018)	GDP: national, regional levels Government finances: national, administrative levels Labour market: national, regional Demography: national, regional	GDP: total, gross value added (2018), apparent investment rate (2017), disposable income, direct and indirect taxes Government finances: revenues, direct and indirect taxes, compensation of employees, social security contributions; expenditure, revenues; breakdowns by type of expenditure, revenues (current/capital) Labour market: activity status (employed, unemployed, inactive), working schedule and hours, labour costs, job vacancies, wages and earnings; breakdowns by sex, age groups, place of residence, education level, economic sectors, employment type Demography: population, births, deaths, life expectancy, marriages, divorces, migration, age dependency, aging ratio,

				natural increase; renewal of active population, population density; breakdowns by sex, age (years and 5 years groups), country of birth, nationality, place of residence,
Romania	Source of data is provided on the NSI website (including the original source, e.g. Ministry of Public Finances, Labour Force Survey, Administrative sources)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2018), quarterly (2019q3), monthly (February 2020) Demography: annual (2020), quarterly (2019q4), monthly (March 2020)	GDP: national, regional level (2017) Government finances: national, administrative level (total, central, local, social security funds) Labour market: national, regional Demography: national, regional (NUTS 2, NUTS 3), municipalities	GDP: gross value added (by NACE rev 2, A10), taxes and subsidies, final consumption (by total, households, general government), gross fixed capital formation, exports and imports; compensation of employees; social security contributions Government finances: expenditure and revenues (by types and categories), expenditures (by COFOG, 2018), taxes and subventions, non-financial accounts; expenditure, subsidies and revenues available by regional level Labour market: activity status (employed, unemployed, inactive), working conditions, labour costs, job vacancies, wages and earnings; breakdowns by sex, age groups, place of residence, education level, marital status, economic sectors, occupations, employment type of contract, Demography: population, births, deaths, marriages, divorces, natural increase, life expectancy; breakdowns by sex, age (years and 5 years groups), place of residence (including rural/urban breakdown), education level
Slovakia	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020) Demography: annual (2019), quarterly (2019q4), monthly (February 2020)	GDP: national, regional (2018) Government finances: national, administrative levels (central, local, social security funds) Labour market: national, regional (NUTS 2) Demography: national, regional (NUTS 2), districts, municipalities	GDP: production (gross value added, taxes and subsidies), expenditure (final consumption - households, government, fixed capital formation, exports and imports), income (compensation of employees, mixed income, taxes on production and imports, subsidies) approaches available; economic sectors breakdown included Government finances: assets and liabilities, transactions and other flows (by financial instruments: total, financial and non-financial corporations, general government, households, other) Labour market: activity status (employed, unemployed, inactive), working schedule and hours, labour costs, wages and earnings, occupations, job vacancies; breakdowns by sex, age groups (10 years), education level, economic sectors, employment type of contract, Demography: population, population change, births, deaths, marriages, divorces, migration, life expectancy (incl. by 5 years age groups), population density; breakdowns by sex, age groups (5 years), place of residence, country of birth, nationality, marital status
Slovenia	Source of data is provided on the NSI website	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2019q4), monthly (February 2020)	GDP: national, regional (2018) Government finances: national, administrative levels (central, local, social security funds) Labour market: national,	GDP: income (compensation of employees, taxes, gross operating surplus and mixed income), expenditure (final consumption, gross capital formation, exports and imports), production (output, gross value added, intermediate consumption approaches available; economic sectors breakdown included (2018) Government finances: revenues and expenditures (by categories

		Demography: annual (2019), quarterly (2019q4), monthly (February 2020)	regions (NUTS 2) Demography: national, regions (NUTS 2, NUTS 3), municipalities	and transactions - taxes, social contributions, compensations, transfers, subsidies); government debt, financial and non-financial assets (annual, 2018) Labour market: activity (employed, unemployed, inactive), job vacancies, transitions in the labour market status, earnings and labour costs, average working hours; breakdowns by sex, age groups (5 years), country of birth, type of employment, occupations, place of residence, economic sectors, education level, Demography: population, births, deaths, marriages, divorces, natural increase, life expectancy; breakdowns by sex, age (years and 5 years groups), place of residence (including rural/urban), citizenship, marital status, education level,
Spain	Source of data is provided on the NSI website (including the original source, e.g. Ministerio de Hacienda (MH) <sup>425</sup> )	GDP: annual (2019), quarterly (2020q1) Government finances: annual (2018); data from MH: annual (2019), quarterly (2019q4), monthly (march 2020) Labour market: annual (2019), quarterly (2020q1) Demography: annual (2020 provisional, 2019), semester (2019s1)	GDP: national, regional (2018) Government finances: national, administrative levels Labour market: national, regional Demography: national, regional (NUTS 2, NUTS 3), municipalities	GDP: production (gross value added, taxes and subsidies), expenditure (final consumption, gross capital formation, exports and imports) and income (compensation of employees) approach; economic sectors Government finances: revenues and expenditures (by categories and transactions - taxes, social contributions, compensations, transfers, subsidies); government debt, financial and non-financial assets, balance sheets (2017) Labour market: labour status (employed, unemployed, inactive), working hours, wages and labour costs; breakdowns by sex, age groups (5 years), type of employment, education level, economic sectors, nationality, occupations Demography: population, births, deaths, marriages, divorces, migration; crude birth and death rates, fertility rates; breakdowns by sex, age (years and 5-years groups), place of residence, nationality, country of birth
Sweden	Source of data is provided on the NSI website (including the original source, e.g. Labour Force Survey)	GDP: annual (2019), quarterly (2019q4) Government finances: annual (2019), quarterly (2019q4) Labour market: annual (2019), quarterly (2020q1), monthly (March 2020) Demography: annual (2019), quarterly (2020q1), monthly (March 2020)	GDP: national, regional (NUTS 2, NUTS 3 - 2018) Government finances: national, administrative levels Labour market: national, regional (NUTS 2, NUTS 3), municipalities Demography: national, regional (NUTS 2, NUTS 3), municipalities	GDP: expenditure (final consumption, by households, general government, gross capital formation, imports and exports), production (gross value added by economic sectors breakdown, taxes and subsidies, government) approaches, household consumption Government finances: revenue, expenditures and transactions, financial and non-financial assets, balance sheets Labour market: activity status (employed, unemployed, inactive), job vacancies, labour costs, hours worked; breakdowns by sex, age groups (5 years), education level, citizenship, marital status, economic sectors Demography: population, births, deaths, migration, marriages, divorces, average age of the population, population density; breakdowns by sex, age (years and 5- and 10-years groups), place of residence, country of birth, citizenship

<sup>425</sup> <https://www.igae.pap.hacienda.gob.es/cigae/> - dashboard

A few considerations can be deduced from the table above and should be considered further:

- Licences and terms of use converge across Member States' NSI, and source acknowledgment is common occurrence and good practice. Official statistics are free of charge in all countries. In one country (Austria), the access to data is partially free of charge, as for data with higher granularity a subscription fee is needed to access them. Payable customised service is common practice across countries, as often offered by NSI for specific requests of data from users/re-users. However, the fees charged for these services refer to the service provided, not the data themselves.<sup>426</sup>
- For most of the countries transitioning towards an open format to provide the datasets won't be a problem. Except for Cyprus, Greece and Slovakia where only Excel format is provided, the rest of the countries include at least one open format in dataset provisions.
- Only half of the countries includes APIs for data dissemination, for the others this provision is not always clearly stated. While some of the stakeholders mentioned the possibility of development of standardised APIs, the majority consider that customisable APIs might respond better to users' needs.<sup>427</sup> For example, Statistics Estonia offer the possibility to create and use APIs in XML and JSON formats to access the information stored in the database. These APIs can be customised to retrieve all the data within a data cube or only specific dimensions, which is extremely useful for large data cubes (over 1 000 000 observations).<sup>428</sup>
- All NSIs publish metadata and the detailed documentation information. One exception of this pattern is Cyprus where the metadata information is missing.
- Update frequency and the latest available data (timeliness) are consistent across Member States. Often the differences encountered between the data timeliness (at annual or quarterly levels) in countries can be explained by the national implementation of the process regarding the production of statistical data (especially, when there are also other data producers external to the NSIs).<sup>429</sup> The coverage of datasets disseminated is influenced by both the national and international contexts. For some of the datasets, certain levels of regional disaggregation in the case of small countries such as Malta and Cyprus, might not be feasible, as it could lead to a break of the statistical confidentiality rule.<sup>430</sup>
- There are several key variables and breakdowns common to all Member States for each type of dataset analysed. Based on the national policy developments and needs, additional breakdowns are provided with these datasets. Sex, age groups and place of residence are common breakdowns for labour market and demographic datasets. Some countries will also include citizenship as to facilitate analysis of foreign workforce or population incidence. Education levels is another breakdown included in these two datasets. Economic sectors are often specific to the labour market and GDP datasets. Different types of breakdowns offer various opportunities to users to re-use statistical data and extract relevant information for their own needs.

In addition, based on the feedback and input received from Eurostat, the above-mentioned list has been updated with the datasets proposed as HVDs for statistics proposed by the members of the European Statistical System Committee (ESSC).<sup>431</sup>

---

<sup>426</sup> Stakeholder interviews.

<sup>427</sup> Focus group discussions and stakeholder interviews.

<sup>428</sup> <https://www.stat.ee/public/andmebaas/API-instructions.pdf>

<sup>429</sup> Stakeholder interviews.

<sup>430</sup> Stakeholder interviews.

<sup>431</sup> The first draft list (received in May 2020) contained an indicative list of datasets from nine areas – demography, poverty and inequality, national accounts, labour market, prices, regional statistics, government finances, business statistics and health statistics. A second input, received in June 2020, presented a more detailed version of the proposed high-value datasets for macroeconomic statistics, using the before-mentioned statistical areas. The detailed final proposal is presented in the European Commission document "high-value datasets in the statistics category" (Ares(2020)3505834).

### 3.5.2 To be situation: extending the PSI HVD rules to these datasets

This section presents the recommended measures for publication, as well as the expected costs and benefits of including these datasets under the scope of the PSI Directive as HVD.

#### 3.5.2.1 Recommended measures for publication

As the previous section shows there is a lot of consistency between the way Member States provide these datasets due to the European Statistical System (ESS). Data holders have pointed out that some of the characteristics analysed have well define guidelines for provisions, stated in the ESS guidelines (these provisions include update's frequency and level of granularity). When putting together the recommendations we also considered the current status across Member States, not only in terms of availability but also from further use perspective.

The table below summarises the recommended measures for publication for the categories of datasets which have been considered.<sup>432</sup>

---

<sup>432</sup> The list has been updated in accordance with the input received from Eurostat, found in the European Commission document "high-value datasets in the statistics category" (Ares(2020)3505834).

	<b>Description</b>	<b>Social statistics</b> (Demography, labour market, poverty and inequality, health)	<b>Macroeconomic statistics</b> (national accounts, prices, government finances)	<b>Business statistics</b> (short-term business statistics, trade, tourism flows)
Openness	<i>License and terms of use</i>	CC BY 4.0	CC BY 4.0	CC BY 4.0
	<i>Format</i>	Provide at least one type of open format (e.g. CSV, XML, JSON)	Provide at least one type of open format (e.g. CSV, XML, JSON)	Provide at least one type of open format (e.g. CSV, XML, JSON)
	<i>Machine-readability</i>	Recommended	Recommended	Recommended
	<i>Availability of API, bulk download</i>	Recommended to provide both APIs and bulk download	Recommended to provide both APIs and bulk download	Recommended to provide both APIs and bulk download
Documentation	<i>Metadata (dataset content description)</i>	Recommended (important)	Recommended (important)	Recommended (important)
	<i>Documentation (incl. structure and semantics)</i>	Recommended	Recommended	Recommended
	<i>Data linking</i>	To be considered	To be considered	To be considered
	<i>Shared vocabularies/taxonomies</i>	Recommended (including DCAT, CKAN option)	Recommended (including DCAT, CKAN option)	Recommended (including DCAT, CKAN option)
Completeness	<i>Versioning</i>	Recommended (clear indication of dataset and metadata version available)	Recommended (clear indication of dataset and metadata version available)	Recommended (clear indication of dataset and metadata version available)
	<i>Update frequency</i>	Annual Quarterly Monthly (specific datasets)	Annual Quarterly Monthly (specific datasets)	Annual Quarterly Monthly (specific datasets)
	<i>Breakdowns</i>	<b>Demography:</b> sex, five years age groups, place of residence, marital status, citizenship, country of birth, education level (specific datasets); regional breakdown - NUTS 2 and	<b>National accounts:</b> economic sectors (NACE rev. 2), types of assets and regional level (NUTS 2 for specific datasets) <b>Price statistics:</b> classification of	<b>Short-term business statistics:</b> type of activities (NACE rev. 2) <b>Trade statistics:</b> type of activities (NACE rev. 2) and products, partners and flows



Description	Social statistics (Demography, labour market, poverty and inequality, health)	Macroeconomic statistics (national accounts, prices, government finances)	Business statistics (short-term business statistics, trade, tourism flows)
<i>Key variables</i>	<p>NUTS 3, for specific sets (fertility, mortality and population).</p> <p><b>Labour market:</b> sex, five years age groups, place of residence, education level, citizenship, occupations, regional breakdown (NUTS 2 level).</p> <p><b>Poverty and inequality:</b> sex, five years age groups, place of residence, education level, citizenship, country of birth, activity status and regional breakdown (NUTS 2 level).</p> <p><b>Health statistics:</b> functions, providers, financing schemes</p>	<p>individual consumption by purpose (COICOP)</p> <p><b>Government finances:</b> categories of revenue, expenditure and government liability.</p>	<p><b>Tourism flows statistics:</b> country of origin, country of destination, duration of trip, means of transport and accommodation, booking modalities and geographic breakdown.</p>
	<p><b>Demography:</b> Population, births and deaths, life expectancy, crude rates for births and for deaths (including infant mortality), fertility rates (total, by mother's age), median age, old age dependency.</p> <p><b>Labour market:</b> employment and unemployment rate, types of employment (full-time, part-time, self-employed, temporary), duration of unemployment (including long-term unemployment), job vacancies, persons looking for jobs.</p> <p><b>Poverty and inequality:</b> people at-risk of poverty and social exclusion, severe material deprivation, material and social deprivation, households with low work intensity, income quintile, Gini coefficient of equivalised disposable income</p> <p><b>Health statistics:</b> current health expenditure</p>	<p><b>National accounts:</b> gross domestic product at market prices, gross value-added, final consumption expenditure of households, exports and imports, taxes and subsidies, gross capital formation (total and non-financial corporations, households), gross operating surplus and mixed income of non-financial corporations, financial sector assets and liabilities, compensation of employees, households sector assets and liabilities employment, disposable income and savings of households, gross national income, net lending/borrowing (total economy, financial and non-financial corporations, households).</p> <p><b>Price statistics:</b> HICP monthly index and rate of change (monthly and annually), HICP annual (rate of change).</p> <p><b>Government finances:</b> government revenue, expenditure, net</p>	<p><b>Short-term business statistics:</b> industrial production index, industrial producers price index, domestic and non-domestic producers price indexes (euro and non-euro areas), production in construction index</p> <p><b>Trade statistics:</b> retail trade volume index, statistical values, net mass and supplementary quantity of international trade.</p> <p><b>Tourism flows statistics:</b> nights spent at tourist accommodation establishments, participation in tourism, tourism trips and expenditure made by EU residents, tourism nights spent by EU residents</p>

<b>Description</b>	<b>Social statistics</b>	<b>Macroeconomic statistics</b>	<b>Business statistics</b>
	(Demography, labour market, poverty and inequality, health)	(national accounts, prices, government finances)	(short-term business statistics, trade, tourism flows)
		lending/borrowing of the general government, government gross debt	

As the table suggests, the similar recommended modes of provision apply to the four categories of datasets in scope. Differences are encountered in the completeness of information section, due to particular characteristics of datasets or specific rules of update linked to statistical provisions. The justifications for each of these recommended measures are the following:

- When it comes to licences and terms of use, both re-users and data holders showed preference to **CC BY** type of licensing. This licence lets others distribute, remix, adapt, and build upon others' work, even commercially, as long as they credit the source for the original creation. It is the most accommodating of licences offered under the Creative Commons. It is recommended for maximum dissemination and use of licensed materials. As most of the licences used by NSI have similar characteristics to the CC BY licence, the recommendation propose won't change significantly the regular modus operandi of the national statistical institutes. In addition, users and re-users consider source acknowledgment as a natural behaviour especially when it comes to data use in reports, analysis and articles.<sup>433</sup> Data holders consider the CC BY licence as appropriate too as it allows to identify the originators of data (even when the original data have been transformed to fit the users' needs) and understand the data comparability across multiple sources.
- The re-users prefer an open **format** whenever applicable, with particular preference for JSON over XML (for the light structure of the former), but CSV is also a format that is useful to them.<sup>434</sup> As the analysis of the current provisions across Member States showed, 75% of the countries offer the CSV format and 35% include JSON files. Both data holders and re-users agree that while XLSX is the most frequent type of format found, it is not a machine-readable one. However, due to its characteristics, the format allows a relatively easy transformation into open, machine-readable format.
- Re-users agree that both APIs and bulk download should be made available in order to foster the re-use of the datasets. AS the previous analysis showed, currently, only half of the countries provide APIs or bulk-download (or both). Some of the re-users consider that both APIs and bulk-download should be available, as each type of service fits for different purposes - APIs allow the provision of specific customised information, bulk download allows to carry out analysis with large amount of data. One of the downsides of the bulk download is that any changes made in the original database won't be reflected into the downloaded batch, therefore the time-stamp of downloaded data (database versioning) becomes extremely important.
- Both metadata and documentation are recommended to be provided, as they are important factors in reusability of the datasets. When it comes to the documentation, all Member States already comply with the recommendation as they publish this information on the NSI websites, even if the degree of completeness might vary across countries. A similar situation is found in the case of the metadata provisions, where only one country (Cyprus) does not have this information available.
- The use of **shared vocabulary** and **taxonomies** is highly recommended as it will increase the discoverability of datasets and data services. Using standardised models and vocabulary, such as Data Catalogue Vocabulary (DCAT) will facilitate the interoperability between data catalogues published across the web and allow consumption and aggregation of metadata from multiple catalogues. In the same time, it will significantly contribute to the improvement of data linking between datasets available on the internet.
- Discussions with both stakeholders and data holders showed that more important than traceability (associate often with identification of the data source) is the **versioning** of the datasets. It is recommended to include under the characteristic, information about the versions for both data and the corresponding metadata. This will allow clear identification of datasets and the moment of extraction and it will improve datasets re-usability.

---

<sup>433</sup> Focus group discussion and stakeholders' interviews.

<sup>434</sup> Focus group discussion.

- **Update frequency** and **breakdown availability** are two important factors for the use and re-use of statistical datasets. A better timeliness (more frequent updates) or/and a higher disaggregation level increase the relevance of these datasets for re-users. However, often there is a trade-off between data timeliness and accuracy as higher frequency might reduce the quality and accuracy of the data provided. The European Statistical System (ESS) already provides a set of guidelines for the geographic breakdown of statistical datasets and for the highest frequency update available. And these guidelines are periodically revised and updated. In ESS guidelines, the GDP and government finances are provided both on annual and quarterly basis and are available at national level. At this stage, the recommendation to provide a regional breakdown for the GDP is only on optional basis. Recent development showed that countries have started to offer a regional breakdown for the GDP, based on the information from the GDP income approach. However, there is an important time gap between this breakdown and the GDP at national level, often between one to two years delay.
- Key variables (or dimensions) are an important characteristic for the datasets, as they will allow better exploitation of the information provided. Availability of certain breakdowns for these variables could increase significantly the usability and re-use of the datasets. Sex and age breakdowns provide important insights when it comes to demography and labour market datasets. Demographic aging as well as workforce aging are important parameters when analysing economic perspectives of countries or regions. Labour market dynamic is reflected through both employment and unemployment attributes. Other important variables in this context are job vacancies and labour costs. The job vacancies correspond to the part of labour demand that is not met by labour supply and thus provide key information on the size and structure of labour market mismatches. Labour costs represent important information for entrepreneurs and the business community in their development planning.

### 3.5.2.2 Expected costs

Considering the scope of the datasets to be made available as HVD and building on the recommended modes of provision suggested above, it is possible to discuss the expected costs that Member States would have to face when adapting to the PSI HVD provisions. First it is important to understand what are the categories of costs that data holders bear today for the provision of datasets in order to have a baseline and some insights on the magnitude of present costs. However, one important aspect mentioned by the stakeholders was that since providing these datasets is part of the regular activity of the NSI, it is extremely difficult to identify specific costs for particular datasets. The following assessments refer to the additional “extra” costs, incurred to the organisations for complying with the requirements stipulated by the Implementing Act, on top of their current costs with providing the datasets under normal conditions.

The table below aims to provide a comprehensive overview of the main cost drivers for the provision of data.

<b>Cost category</b>	<b>Description</b>	<b>Insights from the data collection</b>
Infrastructural costs	Costs related to infrastructural investments such as portals, APIs, Servers (cloud), etc.	Infrastructural costs vary significantly across countries depending on the existing IT infrastructure, but also on the country’s size. In general, the infrastructure existing within the NSI supports the whole activity of the institution, not only the production of the datasets. Therefore, is often difficult to separate these costs from the general IT costs of the organisation as a whole. In this case, many of the stakeholders mentioned that precise figures cannot be provided only for particular datasets provisions.
Data transformation costs	Costs related to data processing including data cleaning,	The data transformation costs are rarely available, as data transformation is considered business as usual by NSIs and the related costs are too “hidden” in the budget. Often the

	preparation of metadata, aggregation, anonymisation, etc.	specific open data initiatives recently started, costs of the overarching initiatives might be available, but these costs cover more than just the datasets in scope of the analysis.
Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	Similarly, to the data transformation costs, the operational costs are rarely made explicit. These types of activities often are part of the current activity of the organisation and not separately counted.
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	No precise figures could be collected on the other costs.

As the data collected and the information given by different stakeholders suggest, estimating costs for providing a sub-set of datasets is not an easy task. Often open data production and dissemination activity are integral part of regular activity and internal process of the organisations. The different sources to gather the data are on many occasions common to more than one dataset and breaking down the information is not feasible. Moreover, data transformation and editing apply to all statistics at once and the results are often stored in common databases as part of an organisation integrated process. When it comes to people involvement, even then the process is not as clear-cut as needed to make a division of the estimated costs of particular datasets since often they work on more than one statistic at the same time.<sup>435</sup> Therefore, the following estimates will often refer to the broader area of the official statistics rather than to particular datasets.

To understand how this cost structure would be affected by the implementation of the Directive one then needs to look at what could be (in general) the main budgetary implications of the recommended measures for publication for the countries which are not yet close to provide them. The table below assess the possible level of impact on current budgets in comparison with the costs incurred currently by the organisations (in the current conditions of data provisions). As the impact is often difficult to quantify, the assessment uses qualitative values, such as little to none, low, medium, high.

<b>Recommended dimension for publication</b>	<b>Budgetary implication (little to none, low, medium, high)</b>
License and terms of use: CC BY	Little to none: as all Member States provide open licences with many similarities with the CC BY licence, the changing of licence would have more likely no impact on countries' costs.
Format: adding an open format such as CSV/JSON/XML	Little to none: Member States already include multiple formats when providing the data, and often the CSV open format is included (21 countries out of 27). Moreover, 35% of the countries provide data in JSON format and 40% include an XML file. Therefore, adapting the data provision to recommended formats would require a minimum level of investments for the countries that do not yet provide an open format.
Modes of provision: both	Medium to high: for some of the countries, the establishment of APIs and

<sup>435</sup> In the past, Statistics Netherlands had developed a project that attempted to build a model for cost estimations per statistic. During the process, the team found out that this type of calculations, using an activity-based price per statistic, is not possible and therefore, the project has been dropped.

API and bulk download	bulk download would be the most impactful change in terms of costs, especially when none if the option is provided. Often cost estimations for these developments are not provided separately from other costs related to the infrastructure costs.
Metadata and documentation: recommended	Little to none: as all countries already include metadata and documentation on their websites, the costs of providing metadata and documentation will be relatively low, mostly related to the update and maintenance of information already provided. For countries where metadata information is not available or needs to be re-structured, it is possible that the impact to be significantly higher than for the rest of the countries. <sup>436</sup>
Data linking: to be considered	Medium to high: as the data linking is often related to the use of shared vocabulary and uploading data on open data platforms, the current recommendation sets up the premises for future use of this characteristic. The option might have a high impact on organisations due to its complexity as significant time and resources (human and/or financial) will be needed to implement it properly. <sup>437</sup>
Shared vocabularies and taxonomies: recommended	Low: adopting shared vocabularies and taxonomies could be costly for Member States but these are only recommended and therefore there could be little to no costs in the short term.
Versioning: recommended	Low to medium: including the versioning information for both datasets and metadata within might require some efforts in complying from the data holders' side. Depending on the solutions adopted and the information availability, the impact of the recommended measure could vary from very little to medium (e.g. for the countries where metadata information is not available and needs to be developed).
Update frequency and timeliness: annual, quarterly or monthly based on the dataset's specificity	Low to medium: the recommended frequency of publication is in concordance with the current provisions from a majority of Member States. Moreover, similar provisions are included into the ESS production and dissemination guidelines. However, for those countries where adjustments of the updating process to improve datasets timeliness are still needed, it might result in possible budgetary implications (the level of this implications is difficult to estimate as it will depend of the countries and their particular situation).
Granularity: coverage (national or regional level) and breakdowns <b>(depending on the</b>	Little to none: the level of granularity indicated in the recommended measures is already provided by all NSI. Some budgetary implication might appear if additional adjustments are needed at breakdown levels to comply with the current recommendation.

<sup>436</sup> Stakeholder interviews.

<sup>437</sup> Stakeholder interviews.

## dataset)

Key variables	Little to none or low: the key variables recommendation built on the most common variables already included by the majority of Member States under current provisions. Some budgetary implication could appear for those countries that need to adjust their list of key variables to match with the recommended ones.
---------------	--

As the table clearly suggests, a number of recommended measures would have limited to no impact on the budgets of data holders due to the fact that countries already provide the datasets as it is recommended and does not need to adapt or the adjustments are rather small and does not involve high costs. The recommendations that might have budgetary implications for countries are mostly related to the metadata provision (where needed), data and metadata versioning and APIs and bulk download developments. However, as mentioned previously, often the changes and adjustments might refer to broader activity in datasets production than to only a few ones.

Cost components	Cost components description	Magnitude of costs (range)
Infrastructural costs	Establishment of the API and bulk download, adaptation of the IT infrastructure to real time provision	Initial investment (one time only) depending on the solution, in between 500 000 and 3 000 000 EUR. For maintenance and further developments, depending of the size, in between 10 000 and 20 000 EUR (yearly)
Data transformation costs	Costs related to data processing including data cleaning, preparation of metadata, aggregation, anonymisation, etc.	Between 30 000 and 50 000 EUR (yearly)
Operational costs	Costs related to data updates, replies to user requests, corrections of errors in the datasets, etc.	In between 100 000 and 200 000 EUR (yearly)
(Lost) income for data supplier	(Share of) revenue related to the provision of the HVD	Not significant as often the datasets are provided for free. However, depending on country's model, some loss of income might incur.
Other costs	Any other costs such as legal advice on GDPR, training costs, etc.	In between 20 000 and 50 000 EUR (yearly)
Negative impact on competition	The estimated impact of competition distortion vis-à-vis private organisations active in the domain.	Not applicable

Building on the data collection activities and on the above-mentioned general analysis of costs, we tried to estimate the magnitude of costs (on a scale from low costs to very high costs) for all EU countries for which the relevant information was available. The results of this assessment are provided below.

Country	Cost impacts					Comments
	Low	Medium	High	Very high	Unknown	
Austria					x	If the datasets are already provided through the Austria open data portal, the impact on

			the organisation will be low. If the datasets are currently provided only via subscription, the impact will be related to the loss of revenue for the organisation. Moreover, this could further affect the organisation's business model and result in a higher impact than initially presumed.
Belgium		x	The datasets in scope have several different official statistics producers, from federal and regional levels. In general, the datasets are available under open data provisions. One shortcoming is represented by the lack of APIs for the datasets (few available, but not on generalised level). It is difficult to estimate the impact (in terms of costs) of further developments of API for the organisations.
Bulgaria		x	The possible impact on the organisation of the implementation of recommendations proposed by the study is difficult to estimate as it might involve several actors of the system.
Croatia		x	The desk research provided limited information concerning some of the provisions of the datasets, especially for API and bulk download availability. Therefore, the possible impact on the organisation of updates or/and upgrades is difficult to assess at this stage.
Cyprus		x	The extra costs for the organisation will result from the adjustments needed for the current system to comply with the provisions for HVD recommended in the study, such as the availability of an open format for datasets, APIs development and bulk download, development of metadata, etc..
Czech Republic		x	The extra costs will relate to the adjustments the system might undergo and could vary from tens to hundreds of thousands of EUR.
Denmark		x	The extra costs will be mostly related to implementation of data linking (highest share of the costs) and adjustments of the documentation already provided.
Estonia		x	Estonia provides the datasets as recommended by this report. The system update and maintenance costs are estimated around 100 000 EUR/yearly.
Finland		x	Finland provides the datasets as recommended by this report. The costs will concern mainly the system's update and maintenance.



France	x			As France already provides the most of the datasets in scope as recommended by this report, the extra costs will refer mainly to update and maintenance.
Germany	x			The country provides currently the datasets as the report recommends through their dissemination database. The implementation of the current recommendations will not produce significant extra costs.
Greece		x	x	The current system will need to implement significant changes to comply with the provisions specified by this report. For example, the overall cost estimated based on the input from the stakeholder vary between €100 000 and €250 000 (without including costs for APIs development and cloud servers).
Hungary			x	The desk research provided limited information concerning some of the provisions available for the datasets. Therefore, the possible impact on the organisation of updates or/and upgrades is difficult to assess at this stage.
Ireland	x			Currently, Ireland provides all the datasets as recommended by this report. The possible extra costs that might appear in the future will be mainly related to maintenance and further developments of the system (the country plans to transition towards cloud storage of the databases for better accessibility).
Italy	x			Italy provides the datasets as recommended by this report. The costs will concern mainly the system's update and maintenance.
Latvia			x	The desk research provided limited information concerning some of the provisions of the datasets, especially for API and bulk download availability. Therefore, the possible impact on the organisation of updates or/and upgrades is difficult to assess at this stage.
Lithuania	x			The results of the desk research showed that the datasets are already provided in compliance with the study's recommendations. Therefore, the extra costs have being estimated to be rather low, concerning mainly improvements and updates of the current system.
Luxembourg			x	Luxembourg provides most of the datasets as recommended by this report. The extra costs will relate mainly to (possible) minor adjustments, regular maintenance and updates of the system.

Malta		x	The country is in the process of transitioning to a new system development. The changes include a better database design and the implementation of modern tools for dissemination, including APIs development.
The Netherlands	x		The Netherlands provides the datasets as recommended by this report. The extra costs might come from APIs' adjustments. (if required). Otherwise, the costs are often related to regular support and maintenance of the system.
Poland		x	The results of the desk research indicate that some extra costs are possible to incur due to further developments of the current system, in addition to its regular maintenance and update. The organisation aims to develop and implements APIs for the Knowledge Databases with a bulk download option. However, building APIs to databases without an additional source of funds (e.g. EU funds) would be very difficult to implement.
Portugal		x	The desk research showed that the provisions of the datasets are compliant with the study recommendations. However, the limited information available makes difficult to estimate the future impact on the organisation of updates and upgrades of the current system.
Romania	x		Since the country does not provide APIs and bulk download, the development and implementation of these tools will result in some extra costs for the organisations. However, the level of impact will also depend on the type of solution selected (in-house development or outsourcing to an external IT company).
Slovakia		x	The results of the desk research provide limited information on the extra costs that could incur due to update and upgrade of the current system to comply with the study's recommendations.
Slovenia	x		The country is in the process of modernising its website and dissemination policy. The new design includes also API development and availability of machine-readable formats.
Spain	x		Spain provides the datasets as recommended by this report. The costs will concern mainly the system's update and maintenance.
Sweden	x		Sweden provides the datasets as recommended by this report. The costs will concern mainly the system's update and

When it comes to the datasets in the scope of the study is difficult to quantify the direct impact on different organisations. However, when we look at the results, we can see that for two thirds of the countries that provided some information on possible extra costs, the impact from complying with the study’s recommendations will remain relatively low. For Czech Republic, Romania and Slovenia, the cost’s impact was estimated on the medium side, as some investments will be needed to fully comply with the provisions recommended by the study (especially related to APIs development and implementation). In Greece case, the extra costs will mostly be related to technical and operational adjustments rather than statistical production of datasets. According to the information given by the stakeholders, it is possible that the APIs developments to have the biggest impact on human and financial resources as the current system does not support them and they are not explicitly included in the future development plans. However, considering all these aspects, estimations for the current extra costs are considered to have a medium to high impact on the organisation.<sup>438</sup> For Austria and Bulgaria, the estimated impact is more difficult to quantify (qualify), as the information provided is less concrete. In the case of and Malta, as the country plans to undergo significant transformation of their current system, these particular costs cannot be separated from the overall planning.

We have looked at the analysis of costs for two particular countries (the Netherlands and Cyprus), which were selected for the CBA analysis.

In the case of the Netherlands, all publicly available data (website and statistical open datasets) of the CBS is stored in a cloud environment. Therefore, it is rather difficult to allocate the costs of the API and the internal and external infrastructure to a single HVD dataset without making some arbitrary choices.<sup>439</sup> Moreover, CBS focuses on more generic usage of indicators rather than costs per dataset. In this situation, considering the lack of more detailed information, it is difficult to properly develop cost-benefit model for the organisation. However, as CBS already provides all the data (all statistics produced not only the ones in the scope of the study) in open data format, through both APIs and bulk download, together with metadata and proper documentation. In this case, we can assume that overall the costs’ impact on the organisation to comply with the propose recommendations will be rather low. Translating all the previous assumptions into the cost table, we obtain the following results.

<b>Cost components for Netherlands</b>	<b>Weight</b>	<b>Score</b>	<b>Weighed score</b>
Infrastructural costs	0.3	-1	-0.3
Data transformation costs	0.2	-1	-0.2
Operational costs	0.15	-1	-0.15
Lost income for data supplier	0.25	0	0
Other costs	0.05	-1	-0.05
Negative impact on competition	0.05	0	0
<b>Aggregated costs of HVD</b>		<b>-4</b>	<b>-0.7</b>

In the case of the Cyprus, it is also true that all datasets produced are publicly available on the organisation’s website. However, in this case, several adjustments of the current system might be needed to comply with the provisions for HVD recommended by this study (inclusion of an open format for

<sup>438</sup> Stakeholder interview and feedback.

<sup>439</sup> For example, which dataset (HVD) determines the load of the cloud environment if multiple users do a bulk download?

datasets, development of APIs, bulk download availability, development of metadata, etc.). These developments will produce some extra costs for the organisation during the transformation process. As in previous case, the table below include the results of the impact for the four datasets together.

<b>Cost components for Cyprus</b>	<b>Weight</b>	<b>Score</b>	<b>Weighed score</b>
Infrastructural costs	0.3	-3	-0.9
Data transformation costs	0.2	-3	-0.6
Operational costs	0.15	-2	-0.3
Lost income for data supplier	0.25	0	0
Other costs	0.05	-1	-0.05
Negative impact on competition	0.05	0	0
<b>Aggregated costs of HVD</b>		<b>-9</b>	<b>-1.85</b>

The impact of the implementation of the changes to current systems to comply with the recommendations proposed by the study will depend also on the current developments of the systems in place. For some countries, this impact will be higher if more adjustments are needed to their systems to fit the proposed requirements.

### 3.5.2.3 Expected benefits

Reuse of existing data improves the efficiency of the whole information industry, and reduces costs and burden caused to respondents as less direct data collection is needed. Governments promote Open Data as a driver of economic growth and job creation. Studies show that fast-growing economies often base their success on rich information, which translates into knowledge and more complex and diverse products.<sup>440</sup>

The evidence-based decision making is one of the strongest motivations for producing data and information. In general, relevant statistics have one or more of the following characteristics: many users, essential in fulfilling the mandates of several organizations, facilitate trade or development, their unavailability might create inequities or asymmetric information. Statistics are produced to be used and to make an impact on society through a higher degree of openness and transparency, avoiding misuse of data, ensuring confidentiality and equal access to information as part of human rights. Therefore, a society will have more empowered people, better policies, more effective and accountable decision making, greater participation and stronger democratic mechanisms.<sup>441</sup>

Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic and social situation:

- Enable the decision makers (public sector, business sector or individuals) to function on the basis of high-quality information, thus leading to better outcomes.
- Allow citizens to hold public and other bodies accountable, enabling better understanding society by providing relevant information while respecting the rights of people, in the same time.

<sup>440</sup> César Hidalgo, *Why Information Grows. The Evolution of Order, from Atoms to Economies* (New York: Basic Books, 2015).

<sup>441</sup> UNECE, *Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics* (New York and Geneva: United Nations, 2018).

- Facilitate research and analysis to proceed on the basis of a comprehensive evidence-base, leading to innovation and improved economic and social outcomes.

There are different types of users of official statistics: media and general public, international policies and organisations, decision makers, analysts, non-governmental organisations (NGOs), the civil society, producers of statistics and the scientific community and researchers. All these users have different needs when it comes to data. For example, media and general public, international actors and decision makers, NGOs, civil society and analysts are more in favour of higher level of processed and aggregated data. On the other hand, producer of statistics and information, researchers and the scientific community would prefer to have access to more detailed and complex raw data.

### *Increase social benefits*

The use of official statistics is demonstrated every day in the newspapers, social media and websites, radio and TV. Having access to information means that people can make better decisions affecting their lives, from both short- and long-term perspective. It also allows them to identify more accurately and quickly the needs and social pressures. Good local data, for example, on population and housing makes it possible to plan and target government services better, such as schools and health care facilities, and thus avoid unnecessary spending of scarce public resources.

Offering certain level of datasets with particular key attributes might allow different users, such as campaigners, analysts and advocates, to develop ideas for the projects that could exist if only those particular data are available. In general, certain data gains value through network effects, raises some important issues for the quantification of value, and will help point towards those datasets where standardisation is particularly important.<sup>442</sup>

For example, in Denmark, where the legislation allows for researchers to utilise data for research of general relevance and importance, register-based research an important opportunity offered to the scientific community. Most of the information collected in these registers is for administrative purposes and spans over decades covering the Danish population from cradle to grave. Using this opportunity, researchers working at the National Centre for Register-based Research (Aarhus University) are able to offer their expertise on all aspects of population-based epidemiology to the public at large, covering different aspects, from legislative and ethic on handling data to data management and security and implementation of epidemiological research.<sup>443</sup>

Public finances are ultimately at the heart of government activity, constituting one of the main levers of public action through which governments shape society.<sup>444</sup> Debt, taxation and subsidies are part of the most frequent covered topics in the context of fiscal transparency<sup>445</sup>, alongside more obvious themes of budgets and expenditures. Studies have shown that increased public transparency and openness of data generate confidence in the markets. Statistics can also enhance political accountability and reduce corruption, promoting public accountability, and, most importantly, could potentially enhance the effectiveness and efficiency of public budgets and spending. Moreover, the available of government finances contributes significantly in improving citizen understanding of the state's fiscal behaviour and encouraging a greater civic participation and oversight. In 2007, the Open Knowledge Foundation's Jonathan Gray developed the idea for "Where Does My Money Go"<sup>446</sup> as a visual breakdown of the UK budget, tapping into a growing appetite for both data visualisation and open data ideas. Between 2013

<sup>442</sup> <http://www.timdavies.org.uk/2019/08/14/high-value-datasets-an-exploration/>

<sup>443</sup> <https://econ.au.dk/the-national-centre-for-register-based-research/danish-registers/>

<sup>444</sup> <https://www.stateofopendata.od4d.net/chapters/sectors/government-finance.html>

<sup>445</sup> <https://www.imf.org/external/np/fad/trans/index.htm>

<sup>446</sup> <http://app.wheredoesmymoneygo.org/about.html>

and 2017, the number of projects and platforms emerged from civil society organisations has increase significantly. Some of them have specific objectives of using public finance data for investigation in journalism or to enhance civic participation.<sup>447</sup>

#### Increase in the economic benefits

As part of the official statistics, the datasets in scope fulfil two roles in assessing progress or lack of it: providing a baseline for the phenomenon analysed or being the measure to understand consequences and outcomes of different events. The report "Coronavirus: The World Economy at Risk" from OECD uses the GDP data to develop several scenarios to illustrate the potential global economic effects that could result from the COVID-19 outbreak in China and the risks that it spreads to other economies.<sup>448</sup> The Strategic Development Goals have increased the demands for comprehensive datasets from social, economic, and environmental sectors to measure progress toward the 2030 targets.<sup>449</sup>

The level of granularity (or disaggregation) of information and increased flexibility in how users can work with this data influence significantly the benefits brought forward by this dataset. Moreover, these factors can also attract new users that previously didn't used or re-used these types of datasets. Providing the data in machine-readable format reduce the costs in the long run. Also, data published into the open can be used and reused without diminishing its value, in various contexts - for mobile phone applications, analyses, and other applications. It can contribute to create new jobs and new business opportunities.

Availability of trustworthy and timely statistics is crucial, for instance for a correct assessment of the monetary and economic situation of a country. Demographic data inform decisions to allocate resources across programmes and plan public services, such as building new hospitals, schools or roads. Macro-economic statistics influence the direction of fiscal, economic and trade policies, social welfare and environmental policy decisions, and target efforts to improve efficiency and productivity, and identify cost savings.

Every day people, companies and public institutions in the socio-economic environment are making choices, and inability to access proper data and information could have important consequences for each of them. In the case of public sector institutions, the lack of access to well-based statistics could lead to increased costs of particular decisions. For example, in New Zealand reactions to short-term population change without full consideration of the ongoing demographic transition resulted in a surplus of schools in some regions and a shortage in others. In one area, underestimation of pre-school children led to a shortfall of approximately 40 million USD in government funding during one year.<sup>450</sup> In case of businesses and entrepreneurs is also true. When they decide about stores' location (in development planning, for example) this could lead to investments loses on long-term if their choices are made in the absence of proper information. Knowing about the demography of the region as well as poverty or wealth levels can be crucial for business development and entrepreneurship. Moreover, complementing this information with the one provided by employment and unemployment, types of occupations and levels of education are increase the changes of better planning of businesses development and entrepreneurial initiatives.

For example, a random survey of Austrian tobacco owners showed that they consider beneficial the use of statistical and location-based data.<sup>451</sup> For example, one tobacco owner stated that demographic factors are

---

<sup>447</sup> <https://www.stateofopendata.od4d.net/chapters/sectors/government-finance.html>

<sup>448</sup> OECD, "Coronavirus: The World Economy at Risk," *Interim Report March 2020 in OECD Economic Outlook* (Paris: OECD Publishing, 2020); <https://doi.org/10.1787/7969896b-en>

<sup>449</sup> <https://unstats.un.org/sdgs/report/2016/leaving-no-one-behind>

<sup>450</sup> [http://icots.info/9/proceedings/pdfs/ICOTS9\\_5A1\\_FORBES.pdf](http://icots.info/9/proceedings/pdfs/ICOTS9_5A1_FORBES.pdf)

<sup>451</sup> Barbara Huber, Alexander Kurnikowski, Stephanie Müller, Stefan Pozar, "The Economic and Political Dimension of Open Government Data in Austria," Institute for Entrepreneurship & Innovation, WU Vienna University of Economics and Business, Spring 2013.

important for their range decision, in order to adapt to the existing market and to avoid failures' risks. The survey's results showed that the demographic information relevant to participants included population data (age groups, birth rates, migration data), crime statistics, unemployment rate and wage tax statistics. The population characteristics were considered valuable in relation to the adjustment of the stock of congratulations cards to the birth rate and age distribution of the respective district, as well as consideration of the respective religion. The wage tax and unemployment statistics were related to the adjustment of cigarette and magazine inventory. For a district with relatively high unemployment rate, fewer exclusive and expensive cigarettes or magazines will be offered. Additionally, demographic data are not relevant only for tobacco owners but to property managing companies too. In this case, the potential need for apartments in the future can be derived by taking demographic developments into account or demographic factors can provide information about possible or probable rent price changes.

From the perspective of the decision makers, having reliable data on unemployment, job vacancies, occupations and education level it will help them better design policy to address labour shortages by trying to correlate and complement information from different areas, but also to try to better plan for possible future developments of the workforce. While the final decisions in the labour market are often made by private actors, the role of the public sector, governments and social dialogue platforms in is also important when it comes to skills anticipation, matching and provision of labour market information system.<sup>452</sup>

Datasets serve researchers by providing them with wide, complex and easily linkable datasets in technically advanced environments. The datasets assist in studying complex problems that have multiple causes and cut across many areas of government, such as productivity, innovation, gender pay gap, income deprivation, climate change, joblessness, homelessness etc. In additions, accessible metadata together with a suite of research tools, applications and software offered for processing and analysing data enable researchers to focus on the key issues with which they are concerned, rather than on the preparation of the data itself. Also, the standard definitions, classifications and methods used across countries make international comparisons possible, and enable linking with the other datasets of statistical offices and even other new data.

#### 3.5.2.4 Cost – benefit analysis

When it comes to benefits, it is rather difficult to estimate particular benefits corresponding to one dataset, as often their value is better expressed in their combination with other datasets. Using the framework developed with the study, we tried to associate the datasets with a set of indicators from different macro-economic areas, considering also the desk research results and the inputs received from different stakeholders. For the cost-benefit analysis we will refer to the overall table below as then reference for the benefits related to the four datasets together (for both use cases considered), as relatively few inputs were provided for this.

<b>Benefit components</b>	<b>Weight</b>	<b>Benefit indicators</b>	<b>Score</b>	<b>Weighed score</b>
Economic	0.3	<ul style="list-style-type: none"> <li>· Economic output [X]</li> <li>· Economy monitoring [X]</li> <li>· Employment [X]</li> </ul>	3	0.9
Climate change	0.3	<ul style="list-style-type: none"> <li>· Citizen engagement in addressing climate change [X]</li> <li>· Energy management and efficiency [X]</li> </ul>	2	0.6
Innovation & AI	0.05	<ul style="list-style-type: none"> <li>· Entrepreneurialism and private sector innovation [X]</li> </ul>	2	0.1
Public services and	0.25	<ul style="list-style-type: none"> <li>· Public sector revenue [X]</li> </ul>	1	0.25

<sup>452</sup> Hana Řihová, *Using Labour Market Information: Guide to Anticipating and Matching Skills and Jobs*, Volume I (Luxembourg: European Union, 2016).

public administration		· Public services management [X] · Public services performance [X] · Public administration transparency, accountability & engagement [X]		
Re-use	0.05	· Trust and confidence in information [X]	2	0.1
Social	0.05	· Crime and justice [X] · Disease prediction and prevention [X] · Mobility access [X] · Mobility efficiency [X] · Mobility planning [X]	2	0.1
Aggregated benefits of HVD			<b>12</b>	<b>2.05</b>

Using the costs estimated in the previous section and the score obtained for the overall benefits of the statistical datasets (re-)use, we looked at the probable impact of providing the datasets in scope as HVDs on the two countries analysed. We looked first at the Netherlands situation, presented in the cost-benefit table below.

<b>Benefits and costs for the Netherlands</b>	<b>Score</b>
Aggregated benefits of HVD	2.05
Aggregated costs of HVD	-0.7
Overall impact	1.35
Benefit/cost ratio	2.93

In this case, the results show that the overall impact on the organisation is positive (+1.35) and for each unit spent in providing these datasets +2.93 unit of benefits are generated (the cost-benefit ratio). This is not a surprise, as the country already provides these datasets as open data and it has done it for quite a while already. However, quantifying precisely the impact of the transition remains a challenge as many of the changes within the Dutch system have been done over the time. And not in the least, these results should be considered carefully, as they refer to an overall general use of the datasets rather than to the specific ones (in both terms of costs and benefits).

The second case looked at the case of Cyprus. In this case, the country will need to implement some additional measures in order to comply with the recommendations proposed by the study. The results obtained are presented in the cost-benefit table below.

<b>Benefits and costs for Cyprus</b>	<b>Score</b>
Aggregated benefits of HVD	2.05
Aggregated costs of HVD	-1.85
Overall impact	0.2
Benefit/cost ratio	1.11

The results show that even in this case the effects remain on the positive side. For Cyprus, the positive impact of providing these datasets as HVD is significantly lower than the one for Netherland (only +0.2). And, for each unit spent in providing these datasets there will be +1.11 unit of benefits generated (much lower compared to the Netherland's case). As mentioned previously, the results refer to all datasets in scope and they should be considered carefully.



The two cases analysed indicate that opening specific datasets will be beneficial for both data holders and re-users. Moreover, the costs for doing this will most likely remain manageable as statistics datasets are often already provided as open data. While the needed adjustments will incur additional costs for some of the countries, the overall benefits are generally assessed as being higher than the costs.

### 3.5.3 Recommended policy options

Relevant, timely and usable data is essential for countries to set priorities, make informed choices and implement better policies for sustainable development. Users, including the government and international organizations, are often looking for compatibility, high quality and easier access to the required information. According to both re-users and data holders, **all official statistics should be considered as high value datasets under the PSI Directive**. The members of the European Statistical System also support this opinion when it comes to all official statistics covered by the European legislations. But, in the same time, they also consider of important the effective capacity to provide the statistics as HVDs data according to all of the four availability criteria needs, as standard practice on long-term basis. However, looking from a practical perspective, all these changes might not be feasible for all the Member States to implement in one-go. Moreover, while for some countries the costs related to the Implementation Act might be significant, looking at all the aspects mentioned before, these costs could be seen as short-term investments to reach the final goal.

While most of the countries provide already a wide range of datasets under CC BY-like type of licence, additional adjustments will be needed to incorporate APIs and bulk download options or to develop taxonomies and shared vocabularies. To reduce the burden and facilitate countries' transition where the impact could be significant, we shortened the list of datasets to consider. The three categories of datasets selected – **social statistics, macroeconomic statistics** and **business statistics** - provide relevant information of the social and economic environments within countries. In addition, majority of Member States already provide these datasets free of charge, free to use and re-use for both commercial and non-commercial purposes and in a machine-readable format, making the impact of the current recommendations remain on low-cost impact.

When it comes to the recommendations concerning the format, the open data and the machine-readable formats availability accompanied by well-defined metadata and proper documentation, will increase significantly the re-use opportunity of the datasets. Also, considering that most of these recommendations are already fulfilled by most of the NSIs, the effort to further adjust to comply will have a relatively low impact on the Member States. Complying with the development of shared vocabularies and taxonomies might require some additional effort for countries. However, these provisions combined with future data linking will improve both data discoverability and accessibility and increase the potential value and benefits of the datasets. Timeliness and the update frequency, as well as breakdowns availability are very important for the datasets in scope. In particular cases, e.g. for categories of indicators within labour market and demographic data a monthly provision of data could make a significant difference. Nevertheless, for update frequency both the annual and quarterly provisions recommendations are also in line with the ESS guidelines for production and dissemination of statistics. A similar situation applies for the breakdowns, despite some existing variations within datasets when it comes to the availability of specific breakdowns. Therefore, these recommendations will not add significant burden on datasets productions for the Member States. Also, the list of key variables incorporates most of the values already provided by the NSIs under the current provisions, and it won't create significant issues.

As mentioned before, the datasets proposed to be included in the HVD list show good potential of use and re-use in different areas – civil society and general public, policy and decision makers, businesses, journalists and researchers. Also, the implementation of the recommendations proposed will have a relative low impact in terms of effort and costs across Member States, also due to the current provisions

for data. Moreover, the benefits resulted from the use and re-use will overcome by the costs for compliance, for all the countries, including the ones where adjustment costs might occur.

The section presents the different policy option designed for the statistics thematic area. In the case of this particular thematic areas, the main parameters used to develop the options were linked with the measures for publication rather than the number of data fields and/or datasets to be included as high-value datasets. One reason for this approach was the current situation related to the statistics where most of the datasets in scope (if not all) are freely available across all EU Member States. The desk research showed that there are still some differences between the datasets across countries especially when it comes to the availability of breakdowns and/or key variables. Also, these results pointed out that it is rather difficult to estimate the costs incurred by selecting specific dataset to be included in a policy options, as this type of information is difficult to assess. Therefore, estimating the impact on the policy recommendation of the inclusion or exclusion of a datasets is not feasible. Moreover, the exercise conducted by Eurostat, together with the National Statistical Institutes, showed that business statistics, macro-economic statistics and social statistics are considered important socio-economic categories for assessing the economic development of countries and their national and international performance.<sup>453</sup>

Considering these findings as well as the inputs received from different stakeholder, the categories of datasets selected for the high-value datasets list are described in the table below. The list encompasses and extends the datasets proposed as HVDs in the European Commission proposal "high-value datasets in the statistics category" by including some additional variables (dimensions) within the sub-categories of datasets, such as births and deaths datasets within demography sub-category and job vacancies data within the labour market sub-category, as well as some additional breakdowns such as types of employment, unemployment duration, occupations within the labour market sub-category.

---

<sup>453</sup> The detailed results of the Eurostat proposal are presented in the European Commission document "high-value datasets in the statistics category", Ares(2020)3505834.

Table 40 – Statistics – Datasets overview

Category	Sub-category	Datasets, key variables and breakdowns
Social statistics	<b>Demography</b>	<p>The datasets in scope are <b>population, fertility, mortality, births</b> and <b>deaths</b>. <b>Key variables</b> are population, births and deaths, life expectancy, crude rates for births and for deaths (including infant mortality), fertility rates (total, by mother's age), median age, old age dependency.</p> <p><b>Breakdowns</b> include gender, five years age groups, place of residence, marital status, citizenship, country of birth, education level (specific datasets). The regional breakdown covers NUTS 2 and NUTS 3 for specific sets (fertility, mortality and population).</p>
	<b>Labour market</b>	<p>The datasets in scope are <b>employment, unemployment</b> and <b>potential labour force</b>.</p> <p><b>Key variables</b> are employment and unemployment rate, types of employment (full-time, part-time, self-employed, temporary), duration of unemployment (including long-term unemployment), job vacancies, persons looking for jobs.</p> <p><b>Breakdowns</b> include gender, five years age groups, place of residence, education level, citizenship, occupations, regional breakdown (NUTS 2 level).</p>
	<b>Poverty and inequality</b>	<p>The datasets in scope are <b>poverty</b> and <b>inequality</b>.</p> <p><b>Key variables</b> are people at-risk of poverty and social exclusion, severe material deprivation, material and social deprivation, households with low work intensity, income quintile, Gini coefficient of equivalised disposable income.</p> <p><b>Breakdowns</b> include gender, five years age groups, place of residence, education level, citizenship, country of birth, activity status and regional breakdown (NUTS 2 level).</p>
	<b>Health statistics</b>	<p>The dataset in scope is <b>Current health expenditure</b>.</p> <p><b>Key variable</b> is the current health expenditure.</p> <p><b>Breakdowns</b> include functions, providers, financing schemes</p>
Macro-economic	<b>National accounts</b>	<p>The datasets in scope are <b>GDP and main aggregates, key indicators on corporations</b> and <b>key indicators on households</b>.</p> <p><b>Key variables</b> include gross domestic product at market prices, gross value-added, final consumption expenditure of households, exports and imports, taxes and subsidies, gross capital formation (total and non-financial corporations, households), gross operating surplus and mixed income of non-financial corporations, financial sector assets and liabilities, compensation of employees, households sector assets and liabilities employment, disposable income and savings of households, gross national income, net lending/borrowing (total economy, financial and non-financial corporations, households).</p> <p><b>Breakdowns</b> include economic sectors (NACE rev. 2), types of assets and regional level (NUTS 2 for specific datasets)</p>
	<b>Price statistics</b>	<p>The dataset in scope is <b>Harmonised Index of Consumer Prices (HICP)</b>.</p> <p><b>Key variables</b> are HICP monthly index and rate of change (monthly and annually), HICP annual (rate of change).</p> <p><b>Breakdown</b> includes the classification of individual consumption by purpose (COICOP).</p>
	<b>Government</b>	<p>The datasets in scope are <b>government expenditure and revenue</b> and</p>

	<b>finances</b>	<p><b>consolidated government gross dept.</b></p> <p><b>Key variables</b> are general government revenue, expenditure, net lending/borrowing of the general government, government gross debt.</p> <p><b>Breakdowns</b> include categories of revenue, expenditure and government liability.</p>
Business statistics	<b>Short-term business statistics</b>	<p>The datasets in scope are <b>industrial production, industrial producer price index</b> and <b>production in construction</b>.</p> <p><b>Key variables</b> are industrial production index, industrial producers price index, domestic and non-domestic producers price indexes (euro and non-euro areas), production in construction index.</p> <p><b>Breakdown</b> includes the type of activities (NACE rev. 2).</p>
	<b>Trade statistics</b>	<p>The datasets in scope are <b>retail trade volume</b> and <b>EU international trade in goods (imports and exports)</b>.</p> <p><b>Key variables</b> are retail trade volume index, statistical values, net mass and supplementary quantity of international trade.</p> <p><b>Breakdowns</b> include the type of activities (NACE rev. 2) and products, partners and flows.</p>
	<b>Tourism statistics</b>	<p>The dataset in scope is <b>tourism flows in Europe</b>.</p> <p><b>Key variables</b> are nights spent at tourist accommodation establishments, participation in tourism, tourism trips and expenditure made by EU residents, tourism nights spent by EU residents.</p> <p><b>Breakdowns</b> include country of origin, country of destination, duration of trip, means of transport and accommodation, booking modalities and geographic breakdown.</p>

As mentioned before, the availability of the datasets described above is not a significant issue, even when the number of key variables and breakdowns available might slightly vary across countries. Also, this list of datasets is not a limitative one, as it can be later enriched with new datasets based on relevance, feasibility and countries' needs. Being part of the European Statistical System, the national statistical systems are widely harmonised. Therefore, complying with the requirements for the above-mentioned datasets won't raise the impact on the organisations' activities. In addition, the lack of information on costs related to specific datasets will result in making arbitrary assumptions on costs for different types of datasets.<sup>454</sup> In this sense, this section presents two possible options for intervention, with different levels of intensity: a lower intensity and a higher intensity, that will mostly rely on the publication's recommendations.

Box 5 – Validation workshop results: statistics, overall appreciation of policy intervention options

During the validation workshop organised on 28 July 2020, participants were requested to indicate whether they agreed or disagreed with the three<sup>455</sup> proposed policy options. The statistics options received the following appreciations (47 Respondents): Agree: 85% and Disagree: 15%.

In addition, the policy options were evaluated by participants as regards their relevance with regards to

<sup>454</sup> The options related to key attributes and/or breakdowns are also linked to the cost information. Therefore, modifying these attributes will not be feasible as lack of cost information will result in unreliable assumptions and will negatively increase the error in the modelling process.

<sup>455</sup> In the initial version of this Deliverable, three policy intervention options were considered per thematic area. For the final version of this Deliverable, and upon request of the Commission, the initial three policy options were merged into two policy options, a lower and higher intensity options. All elements composing the initial three options were transferred through to the final two options, and as such, the validation of the stakeholders still holds.

the overall environment of the thematic area and the respective needs of the participants. The three options obtained the following scores:

- Low: This option obtained the score of 3.4/10 (48 Respondents).
- Medium: This option obtained the score of 4.3/10 (48 Respondents).
- High: This option obtained the score of 5.9/10 (48 Respondents).

### 3.5.3.1 Lower intensity intervention

This first option is a lower intensity intervention, which implies that a set of minimum changes to the current publications options available will apply for datasets considered as HVDs.

This policy option requires minimum changes to the current provisions existing for publication of statistics in the EU Member States. Due to the relatively high degree of standardisation and harmonisation of datasets publication, the current provisions have low to no impact on the National Statistical Institutes from the EU Member States. This lower intensity intervention proposes a set of adjustments' measures to ensure the reusability of the statistics' datasets in scope of this option. These measures are summarised in the table, which is followed by a more detailed description of each of them.

Table 41 - Statistics - Scope of the low intensity intervention

	Description	Social statistics (demography, labour market, poverty and inequality, health), macroeconomic statistics (national accounts, prices, government finances), business statistics (short-term business statistics, trade and tourism flows)
Openness	<b>Licence and terms of use</b>	CC-BY 4.0
	<b>Format</b>	CSV, XML (SDMX), JSON
	<b>Machine-readability</b>	Mandatory
	<b>Availability of API, bulk download</b>	Simple structured APIs and partial bulk download available
Documentation	<b>Metadata (dataset content description)</b>	Simple structured file (basic information only)
	<b>Documentation (incl. structure and semantics)</b>	Complete and web available
	<b>Data linking</b>	N/A
	<b>Shared vocabularies/taxonomies</b>	N/A

As the table suggests, the same recommended modes of provision apply to all four categories of datasets in scope of the study. The justifications for each of these recommended measures are the following:

- Concerning **licences and terms of use**, the current provisions across Member States are already similar to the CC-BY 4.0 type of licence. Therefore, this recommendation will have no impact on the national organisations. At the same time, this type of licence is also preferred by re-users, as they consider as proper behaviour giving the credits to the correct source when using the data.

- When it comes to **format**, the options proposed for datasets publication are CSV, XML (SDMX) and JSON formats. The CSV format is one of the most commonly used formats currently available for the majority of the countries (only Cyprus and Greece do not provide it). However, the two countries provide the data in the Excel format, which makes the transition to CSV relatively easy and simple. The other two formats proposed cover a more structured approach for data files. Therefore, the files are more structured and better fitted for machine-learning developments. Moreover, these types of formats improve of the level of reusability of the datasets. The XML (including SDMX) format is used in more than half of the EU Member States (15 countries). Moreover, the SDMX is already a standard used by the national organisation to interact with the organisations at European and international level (Eurostat, OECD, UNESCO). In this case, the transition should not raise additional difficulties for the national data holders' organisations. The JSON format is often preferred due to its lighter structure and it is starting slowly to gain more usability across EU Member States (10 countries already provide this format). Often, data holders make available multiple formats on their websites, thus complying with these requirements will have a relatively low impact on the organisations.
- For the **APIs and bulk download provisions**, the situation is slightly more complex across countries. The provision through APIs is mandatory under the PSI Directive for HVDs in order to foster the reuse of these datasets. Currently, 15 out of 27 EU Member States include an APIs option for these datasets, and only three (Denmark, Finland and the Netherlands) have clear provisions for bulk download. The two type of provisions fit different purposes: APIs are dynamic in nature and allow the use of slices of data for a more targeted purpose and use (e.g. during the COVID-19 crisis, the number of deaths is an important dataset and updating periodically produces better results), while via bulk download, the dataset obtained is a static image of data, at one point in time, used to carry out different type of analysis (e.g. identifying long-term trends within macro-economic analysis considering different aspects and datasets – GDP, employment, unemployment, demographic development etc.). While APIs can be used to perform a bulk download, it is not their main the purpose of use.<sup>456</sup> To respond to both types of situations, the datasets need to be provided through both APIs and bulk download. Considering the current situation across countries, we can assume that the implementation of this recommendation will have a medium to high impact, especially on the countries without these provisions in place. However, this will also depend on the solution adopted by these countries. Moreover, section 3.8 provides additional information concerning APIs developments (including cost related aspects to be considered further).
- In general, statistical offices provide both **metadata and complete documentation** for their datasets (including the ones in scope of the study). They are an important prerequisite for the reusability of the datasets. All countries provide complete and web-based documentation. Therefore, this requirement is already fulfilled. However, when it comes to metadata files, the completeness of the information varies across Member States. The quality of the metadata files is an important aspect in the data reusability, and the level of detail provided ("granularity") is relevant when developing this "data about data" files.<sup>457</sup> In case of the low intensity intervention, a lower level of granularity is required for the metadata files i.e., providing a minimum level of information using an official statistics baseline standard (e.g. Single Integrated Metadata Structure – SIMS). As SIMS is a standard used by European Statistical System, the impact on the organisations remains relatively low.
- For this type of intervention, **data linking** is not a mandatory recommendation for the datasets in scope.

---

<sup>456</sup> Joshua Tauberer, *Open Government Data (The Book)*, Second edition, 2014; <https://opengovdata.io/2014/bulk-data-an-api/>;

<sup>457</sup> Metadata with a high granularity allows for deeper, more detailed, and more structured information and enables greater level of technical manipulation. A lower level of granularity means that metadata can be created for considerably lower costs but will not provide as detailed information.

- A similar approach is used for the **shared vocabularies and taxonomies**. Both data holders and re-users consider controlled vocabulary and taxonomies extremely useful in data discoverability and reuse. Moreover, while the shared vocabularies are not a novelty for the National Statistical Institutes, they might not be using them consistently yet. Therefore, developing the two characteristics require a relatively high level of effort for stakeholders and it would be too much complicated for data holders to deal with it right now. Thus, at this stage, this recommendation is not included.

### 3.5.3.2 Higher intensity intervention

The second option in statistics thematic area is the higher intensity intervention, which adds new changes, in terms of measures for publication. The intervention builds on the measures previously included in the lower intensity option. These changes target areas that might increase the burden to comply with for some of the organisation, as they will need more effort in order to fulfil the requirements.

The table below provides an overview of these measures, highlighting in blue those that are new in comparison to the lower intensity intervention.

Table 42 - Statistics - Scope of the high intensity intervention

	Description	Social statistics (demography, labour market, poverty and inequality, health), macroeconomic statistics (national accounts, prices, government finances), business statistics (short-term business statistics, trade and tourism flows)
<b>Openness</b>	<b>License and terms of use</b>	CC-BY 4.0
	<b>Format</b>	CSV, XML (SDMX), JSON
	<b>Machine-readability</b>	Mandatory
	<b>Availability of API, bulk download</b>	Complex APIs <sup>458</sup> (including customised clients) and bulk download available
<b>Documentation</b>	<b>Metadata (dataset content description)</b>	Well-developed structured file (i.e. description of the statistical data, as well as descriptions of the statistical concepts, methodologies and information on data quality)
	<b>Documentation (incl. structure and semantics)</b>	Complete and web available
	<b>Data linking</b>	Recommended, but not mandatory
	<b>Shared vocabularies/taxonomies</b>	Controlled vocabularies and taxonomies DCAT-compatible

The difference with the lower intensity intervention refers to improved APIs and metadata files, development of controlled vocabularies and taxonomies as well as the setting the premises for development of data linking for the datasets in scope:

- In this type of intervention, the **APIs** complexity increases and full **bulk download** becomes the norm. Resource intensive APIs, with complex and intricate structure, will require significantly more

<sup>458</sup> A complex APIs is a resource intensive application, composed of multiple parts with an intricate, elaborate and interconnected structure.

effort for further developments and updates compare to simple structured applications, thus increasing the intensity's level of this intervention. Moreover, section 3.8 provides additional information concerning APIs developments (including cost related aspects to be considered further).

- When it comes to **metadata** files, the degree of granularity is also directly proportional to the costs of maintaining and upgrading the information. Basic information is easy to update, but more complex files will need significant effort to keep it relevant. The current policy intervention aims going one step further when it comes to metadata completeness. For example, metadata files content can be developed starting from a statistics baseline standard (e.g. SIMS), and further providing additional layers of details about the data quality and completeness, methodologies, sources etc. However, with increasing of details of information included, the costs incurred by the national data holders' organisations to maintain and update the respective files increases proportionally.
- In the high intensity intervention, the implementation of **data linking** recommendation is envisaged, however it is not made compulsory. While data linking makes it easier for developers to connect information from different sources, increasing the level of discoverability and re-use of the datasets, the implementation process will increase significantly the burden on data holders. On the positive side, the data linking will ensure the datasets with unique identifiers and proper linking with other sources, facilitating both data update (for data holders) and data discoverability (for data users and re-users). On the less positive aspects, several stakeholders have mentioned that this option falls on the expensive side, as both time and resources (human and/or financial) are required to implement it properly. Considering both the advantages and disadvantages of implementing the data linking, at this stage, the options is recommended but not on a mandatory base as it may excessively burden the data holder organisations.
- **Shared vocabularies and taxonomies** would be mandatory in this higher intensity intervention. Although agreeing on the controlled vocabularies would require some efforts, it would ensure the interoperability across Member States. Particularly, this intervention suggests to start from using the controlled vocabularies and taxonomies (DCAT-compatible), which includes elements from statistical standards, including descriptions of the datasets, and improve the discoverability of these datasets. Also, controlled vocabularies and taxonomies facilitate a better integrability of existing statistical data portals and expand the use-area of these datasets. Discussions with stakeholders also showed that controlled vocabularies play an important role when datasets are uploaded in open data portals. However, as it is difficult to estimate the implementation level of these vocabularies by the National Statistical Institutes, we assumed that the recommendation will have a relatively medium impact across countries.

## 3.6 Mobility

This section presents the micro-level assessment for the thematic area of mobility. It illustrates the current state of play of the provision of these datasets. Furthermore, it provides the recommended measures for publication together with the costs and benefits of including these datasets as high-value datasets under the PSI Directive. Lastly, it details the three policy options proposed for this thematic area.

### 3.6.1 As-is situation: how Member States provide these datasets today

This sub-section describes the current modalities for publication of the datasets in scope.

#### 3.6.1.1 Inland waterways and river infrastructure data

As previously presented, not all EU Member States possess navigable waterways. In addition, the assessment below focuses only on those countries in which the fairways can be considered as international i.e. which have links with other countries, in order to fully exploit the EU added-value.



The responsible data holders are widely similar across all Member States consulted. Overall, inland waterway infrastructure data is mainly held by national competent authorities for (waterborne) transport vested in the ministries of transport/mobility/infrastructure, and presented free of charge on the dedicated national inland waterway and RIS website or data portal.

On the one hand, static data concerning the fairways (e.g. overall navigable routes and related fixed characteristics) and the various immovable infrastructure (e.g. ports, locks, bridges and dams) are usually catered for directly by such aforementioned authorities, who work in collaboration with (or are the same authorities than) those who collect and publish INSPIRE transport networks data, including national mapping/surveying agencies. Currently, these datasets are mostly published in PDF documents (under the form of maps that can be downloaded) or directly presented on webpages (TXT and HTML), and are updated on a yearly basis for some Member States, or 'when relevant' i.e. when changes actual changes in the fairway or its infrastructure occur.

On the other hand, the data collection process for dynamic/urgent data is supported by third party operators (including universities, hydro-meteorological institutes or research centres, infrastructure operators, etc) contracted for the reporting of data from various gauging/metering stations along the waterways kilometres, especially for forecasting. These datasets are provided in varying formats across the EU: sometimes directly presented on webpages (TXT, HTML, without possibility for download of the data), sometimes available for download in CSV or XML files, and in some cases through webservice applications. Dynamic/urgent data are usually updated on a daily basis, if not in (near) real-time.

Last but not least, the domestic electronic navigational charts (ENCs) based on certain static and dynamic data (as presented in section 0) are mostly drawn-up and maintained by the either the (waterborne) transport authority and/or the national mapping/surveying agencies. Depending on the number of dynamic features presented within these charts, they are usually updated at least daily as they take into consideration some dynamic information, however, based on the study team's desk research, it appears some Member States have not updated ENCs since a few years.

The table below provides a high level overview of all information gathered on today's modes of provision of inland waterway infrastructure data across Member States. It should be noted that data provisions were validated only for Austria, Flanders and Germany due to the unavailability of other Member States to take part to interviews. For the latter (marked with \*), the table is based on desk research and publicly available information on the responsible authorities' website and/or dataportal. The table makes the distinction static, dynamic/urgent and the ENCs whenever possible (if any of these is not mentioned, it means the study team was unable to find this information).

Finally, concerning the topics of data linking, taxonomies and traceability, the data collection did not allow to gather information. These topics seemed to be less relevant for data holders and almost never came across from the interviews and discussion with Member States experts. For these reasons, these characteristics of data provision are often marked as Not Applicable (N/A) in the table below, as well as in section 0 – Recommended measures for provision.

Table 43 – Current modalities for provision of inland waterway and river infrastructure data

	Openness-data specification					Documentation					Completeness			
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timeliness	Granularity	Key attributes
Austria	Terms of use (registration required for certain dynamic data)	Yes	PDF (static); HTML, XML (dynamic), 000 (ENC)	Yes (dynamic and ENC)	API, Webservice, Bulk download (PDFs and ENCs only)	Available	N/A	Available for ENC (Viewing 7C)	RIS Index based on ISRS; Inland ECDIS Standard 2.3 (S-57, ENC)	N/A	N/A	(Near) real-time; Daily/when relevant	National; Waterway km/metering stations	Waterway code, km marking
Belgium (FL)	License not specified (registration required for certain dynamic data)	Yes	PDF (static); HTML, XML (dynamic), 000 (ENC)	Yes (dynamic and ENC)	Webservice; Bulk download (PDFs and ENCs only)	Available	N/A	Available for ENC (Viewing 7C)	RIS Index based on ISRS; Inland ECDIS Standard 2.3 (S-57, ENC)	N/A	N/A	(Near) real-time; Daily/when relevant	Regional; Waterway km/metering stations	Waterway code, km marking
Belgium (WAL)*	Terms of use	Yes	PDF (static), 000 (ENC)	Yes (ENC)	Webview, bulk download (PDFs and ENCs only)	Available	N/A	No	Unclear	N/A	N/A	When relevant	Regional; Waterway km/metering stations	Waterway code, km marking
Bulgaria*	License not specified	Yes	PDF (static), HTML (dynamic), 000 (ENC)	Partly (dynamic); Yes (ENC)	Webview, bulk download (PDFs and ENCs only)	Not available	N/A	Available for ENC (Viewing 7C)	RIS Index based on ISRS; Inland ECDIS Standard 2.3 (S-57, ENC)	N/A	N/A	Daily/When relevant	National; Waterway km/metering stations	Waterway code, km marking
Croatia*	License not specified	Yes	PDF (static), HTML (dynamic), 000 (ENC)	Partly (dynamic); Yes (ENC)	Webview, bulk download (PDFs and ENCs only)	Available	N/A	No	RIS Index based on ISRS; Inland	N/A	N/A	When relevant	National; Waterway km/metering stations	Waterway code, km marking

Openness-data specification					Documentation				Completeness					
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timeliness	Granularity	Key attributes
									ECDIS Standard 2.3 (S-57, ENC)					
Czech Republic*	License not specified (registration required for certain dynamic data)	Yes	PDF & TXT (static), HTML (dynamic), 000 (ENC)	Partly (dynamic); Yes (ENC)	Webview, bulk download (PDFs and ENCs only)	Available	N/A	Available for ENC (Viewing 7C)	RIS Index based on ISRS; Inland ECDIS Standard 2.3 (S-57, ENC)	N/A	N/A	(Near) real-time; Daily/when relevant	National; Waterway km/metering stations	Waterway code, km marking
France*	Terms of use	Yes	000 (ENC)	Yes (ENC)	Bulk download	Available	N/A	No	RIS Index based on ISRS; Inland ECDIS Standard 2.3 (S-57, ENC)	N/A	N/A	When relevant	National ; Waterway km/metering stations	Waterway code, km marking
Germany	National open license	Yes	PDF (static); HTML, XML (dynamic), 000 (ENC)	Yes (dynamic and ENC)	API, Webview, bulk download,	Not available	N/A	No	Unclear	N/A	N/A	(Near) real-time; Daily/when relevant	Federal/State ; Waterway km/metering stations	Waterway code, km marking
Hungary*	License not specified	Yes	PDF (static), TXT & JPEG (dynamic)	No	Webview, bulk download (PDFs and JPEG only)	Not available	N/A	No	Unclear	N/A	N/A	Daily/when relevant	National ; Waterway km/metering stations	Waterway code, km marking
Netherlands*	License not specified (registration required for certain dynamic data)	Yes	PDF (static), CSV, XLS (dynamic), 000 (ENC)	Yes (dynamic and ENC)	Webservice, bulk download (PDFs, CSV and ENCs only)	Available	N/A	Available for ENC (Viewing OpenCPN)	RIS Index based on ISRS	N/A	N/A	(Near) real-time; Daily/when relevant	National ; Waterway km/metering stations	Waterway code, km marking
Poland*	License not specified	Yes	PDF (static), HTML	Partly (dynamic); Yes	Webservice, bulk	Not available	N/A	Available for ENC (Viewing	Unclear	N/A	N/A	Daily/when relevant	National ; Waterway	Waterway code,

Openness-data specification					Documentation				Completeness					
	<i>License (terms of use)</i>	<i>Free of charge</i>	<i>Format</i>	<i>Machine-readability</i>	<i>Availability of API, bulk download</i>	<i>Metadata (dataset content description)</i>	<i>Data linking</i>	<i>Documentation (incl. structure and semantics)</i>	<i>Shared vocabularies</i>	<i>Taxonomies</i>	<i>Traceability</i>	<i>Timeliness</i>	<i>Granularity</i>	<i>Key attributes</i>
			(dynamic), 000 (ENC)	(ENC)	download (PDFs and ENCs only)		7C)						km/metering stations	km marking
Romania*	License not specified	Yes	PDF (static), HTML & TXT (dynamic), 000 (ENC)	Partly (dynamic); Yes (ENC)	Webservice, bulk download (PDFs and ENCs only)	Not available	N/A	No	Unclear	N/A	N/A	Daily/when relevant	National ; Waterway km/metering stations	Waterway code, km marking
Slovakia*	Terms of use	Yes	HTML & TXT (static/dynamic), 000 (ENC)	Partly (static/dynamic); Yes (ENC)	Webview, Bulk download (ENCs only)	Not available	N/A	Available for ENC (Viewing SeeMyENC)	RIS Index based on ISRS	N/A	N/A	(Near) real-time; Daily/when relevant	National; Waterway km/metering stations	Waterway code, km marking

A few considerations can be deduced from the table above and should be taken into account when identifying options for the future:

- **There is an open access to data but terms of use are unclear.** Based on the data collection efforts carried out by the study team, it appears that overall, inland waterway and river infrastructure data is easily accessible by the public. These datasets are published on public websites/portals, with certain dynamic data being restricted via log ins for a handful of Member States only. Regarding terms of use, limited data was available to draw tangible conclusions on the extent to which these are used across countries in order to regulate the reuse of data. However, in certain countries, data reuse was restricted for commercial purposes.
- **Charging practices are inexistent.** Not a single country charges for the provision of inland waterway infrastructure data (when registration is required for accessing full datasets, registration is also free of charge), meaning that such datasets are not currently relied upon in terms of revenues.
- **Data is mostly presented in PDF documents and/or directly on webpages/portals.** As touched upon above, inland waterway and river infrastructure data is mostly published via downloadable PDF documents for static data (notably via maps highlighting the key characteristics of the fairway and infrastructure) and dynamic data is often directly consulted on webpages/portals, and is therefore available to a lesser extent in (bulk) downloadable versions or through APIs. This element was also brought up during the focus group, as depicting a 'generation gap': seasoned sailors and transport operators would tend to look up the information prospectively on PDF/webpages and establish voyage planning based on experience, whereas the younger generations would appreciate having all relevant information available in (near) real-time and through various (mobile) apps with precise routing calculations.
- **In terms of machine readability, formats used are partly machine readable.** Linked to the elements mentioned above, data is not provided extensively in 100% machine readable formats. HTML may be read by machines using web-scraping.
- **Only one country currently has an API.** This is again linked to the conscious choice related to current practices of having the information directly displayed on webpages/or portals.
- **Timeliness of data varies significantly depending on the categories of datasets at hand and on the country.** There are various philosophies in terms of frequency of update of the data: while static data appears to be updated at least once a year in all countries, and dynamic data is updated at least daily (if not every 15 minutes like in Austria for certain dynamic data), ENCs provision varies more widely. Indeed, some countries align with dynamic data provision and work in (near) real time while others plan daily, weekly or even yearly updates.
- **Some shared vocabularies are widely used.** National RIS Indices listing all the codes and acronyms used within the datasets/maps published are reusing ISRS Location Codes and the ENCs are developed on the basis of the European Inland ECDIS Standard 2.3<sup>459</sup> at minimum.
- **Key attributes such as waterway/location codes and kilometres are always used.** As mentioned above, the ISRS codes are reused along with waterway kilometre markings in order to uniquely identify all information provided.

### 3.6.1.2 Transport networks data

As per inland waterways and river infrastructure data, the responsible data holders for transport networks datasets under the INSPIRE Directive are widely similar across all Member States consulted. These datasets are mainly held by national competent authorities vested in the ministries of transport/mobility/infrastructure, and maintained in collaboration with national mapping/surveying agencies as well as relevant public undertakings in charge of (public) transport. Further, data collection activities show that a distinction should be made between: (1) the transport networks datasets under INSPIRE and included on the Geoportal as per INSPIRE specifications and (2) 'other' transport networks datasets published (or not) via dedicated websites and platforms at national level, which are briefly discussed hereafter.

Based on data collection activities, it seems that the transport networks datasets listed on the INSPIRE Geoportal do not focus that much on the reusability of the datasets: at the time of submission of this report, the INSPIRE Geoportal contained 3683 metadata, records for only 284 downloadable and 576 viewable datasets in the transport networks data theme. In other words, the INSPIRE Geoportal currently provides the means to access the data by viewing it rather than supporting the data reusability through APIs and/or download possibilities. Overall, the datasets that are available for download are mostly provided in similar, machine-readable formats across Member States, i.e. in GML or XML. In addition, beyond the datasets being provided free of charge, Member States are also rather aligned on the fact that transport networks datasets are to be provided with (nearly) no limitations to access nor use, and privilege open licenses.

The table below provides a high level overview of all information gathered on the current modalities for publication of transport networks datasets provided by Member States on the INSPIRE Geoportal. It should be noted that data provisions were validated only for 8 Member States due to the unavailability of other Member States to take part to interviews. For the latter (marked with \*), the table is based on desk research and publicly available information on the INSPIRE Geoportal. Finally, concerning the topics of taxonomies and traceability, the data collection did not allow to gather information. These topics seemed to be less

<sup>459</sup> See: <https://ris.cesni.eu/31-en.html>

relevant for data holders and almost never came across from the interviews and discussion with Member States experts. For these reasons, these characteristics of data provision are marked as Not Applicable (N/A) in the table below, as well as in section 0 – Recommended measures for provision.

Table 44 – Current modalities for provision of transport networks data (1/2)

	Openness-data specifications					Documentation				Completeness				
	License (terms of use)	Free of charge	Format	Machine-readability	Availability of API, bulk download	Metadata (dataset content description)	Data linking	Documentation (incl. structure and semantics)	Shared vocabularies	Taxonomies	Traceability	Timeliness	Granularity	Key attributes
Austria	No conditions; No limitation of public access	Yes	GML; XML	Yes	Download available (13), Webservice view (12) (except for Water network)	Provided in DE or EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Belgium (FED)*	Terms of use, no limitation of public access	Yes	GML; XML	Yes	Download available, Webservice view	Provided in EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Belgium (FL)*	No data available on INSPIRE Geoportal						N/A	N/A						
Belgium (WAL)*	Terms of use, no limitation of public access	Yes	x-gmz; x-Shapefile	Yes	Download available, Webservice view	Provided in EN or FR	N/A	N/A	INSPIRE Data specifications	N/A	No mention	No information on INSPIRE Geoportal	Regional	As per INSPIRE data specifications for transport networks
Bulgaria*	No data available on INSPIRE Geoportal						N/A	N/A						
Croatia*	No conditions of access, no limitations	Yes	N/A	N/A	No (only metadata available on INSPIRE); download available through national access point	Provided in HR	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Cyprus	No conditions; No limitation of public access	Yes (on national access point)	N/A		No (only metadata available on INSPIRE); download available through national access	Provided in EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks

point														
Czech Republic*	Terms of use (Dle Vyhlášky č. 31/1995 Sb.)	Yes	xml	Yes	Download available (5), Webservice view (7)	Provided in CZ	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Denmark	No conditions; public access via registration	Yes	x-gmz	Yes	Webservice view (4); download available only for sailing routes	Provided in DK or EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Estonia	Licence of open data by Estonian Land Board, 1.07.2018	Yes	Shapefile, TAB, DGN, DXF	N/A	Webservice view (3); download available for 2 datasets only on INSPIRE. (On national access point, no registration needed)	Provided in EE	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	Data provider and date to be mentioned	No information on INSPIRE Geoportal (Change history maintained at object level on national geoportal)	National	As per INSPIRE data specifications for transport networks
Finland	Creative Commons 4.0 BY; no limitation public access	Yes (on national access point)	N/A	N/A	No (only metadata available on INSPIRE); download available through national access point; Webservice view (123). API available on national access point.	Provided in FI (or EN)	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal (Updated continuously and automatically on a weekly/monthly basis as relevant)	National	As per INSPIRE data specifications for transport networks
France*	No conditions; No limitation of public access	Yes	N/A	N/A	Download available (85); Webservice view (163). API available for professional use on national access point.	Provided in FR	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	Data provider to be mentioned in all diffusion	No information on INSPIRE Geoportal		As per INSPIRE data specifications for transport networks



Germany	Terms of use "Data license Germany Zero" No conditions; No limitation of public access	Yes	N/A	N/A	Download (68); Webservice view (123) according to Open Geospatial Consortium standards	Provided in DE or EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A		No information on INSPIRE Geoportal	National (6) or Regional (140)	As per INSPIRE data specifications for transport networks
Greece*	No data available on INSPIRE Geoportal						N/A	N/A						
Hungary*	No conditions; No limitation of public access	Yes	N/A	N/A	No (only metadata available on INSPIRE); download available through national access point	Provided in HU	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Ireland*	No conditions; no limitation of public access	Yes	XML	Yes	No (only metadata available on INSPIRE); download available through national access point	Provided in EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Italy*	No conditions; no limitation of public access	Yes	XML	Yes	Download available (40); Webservice view (54)	Provided in IT	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National; Regional	As per INSPIRE data specifications for transport networks
Latvia*	No conditions; no limitation of public access	Yes	GML; XML	Yes	Download available (4); Webservice view (6)	Provided in LV or EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Lithuania*	Creative Commons 4.0 BY; no limitation public access	Yes	GML	Yes	Download available (3); Webservice view (3)	Provided in EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
Luxembourg*	No conditions; No limitation	Yes	GML; XML	Yes	Download available (10); Webservice	Provided in EN	N/A	N/A	INSPIRE Data specifications	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data

	of public access				view (9). API available on national access points				for transport networks				specifications for transport networks	
<b>Malta</b>	Creative Commons 4.0; No limitation of public access	Yes	CRS(?)		Download available (2); Webservice view (2)	Provided in EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
<b>Netherlands*</b>	No conditions; No limitation	Yes	xml	Yes	Download available (13); webservice view (12). API available on national access point	Provided in NL	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
<b>Poland*</b>	No conditions; No limitations	Yes	N/A	N/A	Webservice (1); download and API available through national access point	Provided in PL	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
<b>Portugal*</b>	Use restricted; no view limitations. Download service subject to intellectual property rights.	Yes	x-shapefile	Yes	Download available (2); Webservice view (1)	Provided in P	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National, regional	As per INSPIRE data specifications for transport networks
<b>Romania*</b>	No conditions; limitations on the confidentiality of the proceedings of public authorities, where such confidentiality is provided for by law and intellectual property	Yes	Xfile gdb (Linux)	Yes	Download available (2); webservice view (3)	Provided in RO	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National, regional	As per INSPIRE data specifications for transport networks
<b>Slovakia*</b>	No conditions; no limitations	Yes	N/A	N/A	No download; webservice view (8)	Provided in SK	N/A	N/A	INSPIRE Data specifications for transport	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specification

									networks					s for transport networks
<b>Slovenia*</b>	No conditions; no limitations	Yes	XML	Yes	Download available (1); webservice view (1)	Provided in SI	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
<b>Spain*</b>	No conditions; no limitations	Yes	XML	Yes	Download available (2); webservice view (2)	Provided in ES	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks
<b>Sweden</b>	No conditions; No limitation of public access	Yes			Webservice view (23) (download for mAIS only). API available on national access points.	Provided in EN	N/A	N/A	INSPIRE Data specifications for transport networks	N/A	No mention	No information on INSPIRE Geoportal	National	As per INSPIRE data specifications for transport networks

Table 45 – Current modalities for provision of transport networks data (2/2)

	<b>Subthemes</b>				
	<i>Road</i>	<i>Rail</i>	<i>Water</i>	<i>Air</i>	<i>Cable</i>
Austria	Yes	Yes	Yes	Yes	No
Belgium (FED)*	Yes	Yes	No	No	No
Belgium (FL)*	No	No	No	No	No
Belgium (WAL)*	Yes	Yes	Yes		
Bulgaria*	No	No	No	No	No
Croatia*	Yes	Yes	Yes	Yes	
Cyprus	Yes	No	Yes (ports and nautical)	Yes	No
Czech Republic*	Yes	Yes	Yes	Yes	Yes
Denmark	Yes	Yes	Yes (sailing routes)	Yes	No
Estonia	Yes	Yes	Yes	Yes	No
Finland	Yes	Yes	Yes	Yes	No
France*	Yes	Yes	Yes	Yes	Yes
Germany	Yes	Yes	Yes	Yes	Yes
Greece*	No	No	No	No	No
Hungary*	Yes	No	No	No	No
Ireland*	Yes	No	No	No	No
Italy*	Yes	Yes	Yes	Yes	No
Latvia*	Yes	No	Yes	Yes	No
Lithuania*	Yes	Yes	No	Yes	No
Luxembourg*	Yes	Yes	N/A	Yes	Yes
Malta	Yes	No	N/A	No	No
Netherlands*	Yes	Yes	Yes	Yes	Yes
Poland*	Yes	Yes	Yes	Yes	Yes
Portugal*	Yes	Yes	Yes	Yes	Yes
Romania*	Yes	Yes	Yes	Yes	Yes
Slovakia*	Yes	Yes	No	Yes	No
Slovenia*	Yes	Yes	Yes	Yes	No
Spain*	Yes	Yes	Yes	Yes	No
Sweden	Yes	Yes	Yes	Yes	No

Last but not least, a key discussion point will be the extent to which the transport networks held at national level, beyond those published on the INSPIRE Geoportal, would be subject to the HVD provisions of the PSI Directive, too.

Regarding the transport networks datasets at national level, the study team was able to gather a limited amount of information. Based on desk research and a high level screening of the national (open) data portals providing these datasets (c.f. table below), it is difficult to establish clear trends among Member States. While on all platforms consulted, the access to data appears to be free of charge, some portals require prior registration, others do not, and, some include specific terms of use, others do not. In addition, transport networks datasets are provided in various formats, including under the form of ready-to-use maps, and are sometimes available for bulk download, sometimes in web-service view only. The extent to which Member States already have APIs established also seems to vary and thus remains a key infrastructural issue to tackle in case these datasets are to be considered under the same conditions as the INSPIRE datasets.

As a matter of example, while only metadata and webservice view for some 120 datasets are available on the INSPIRE Geoportal for Finland, the Finnish Transport Open Data Platform<sup>460</sup> encompasses over 220 transport network and infrastructure related datasets (including the INSPIRE transport networks ones). In this platform, all data are provided as:

- Open data, free of charge,
- Via CC-BY 4.0 license, and,
- Through APIs with bulk download available
- Updated continuously and automatically on a weekly/monthly basis as relevant
- Compliant with INSPIRE data specifications

In 2018, the platform counted:

- 22000 unique user IP's
- 40000 requests (only data packet downloads, no API requests)
- 730 GB (only data packet downloads, no API requests)

The inclusion of such 'additional' datasets is discussed later on in the report, as part of section 0.

Table 46 – Links to national access points to geospatial data and transport networks datasets

Country	Link to national access point
Austria	<a href="http://geometatensuche.inspire.gv.at/">http://geometatensuche.inspire.gv.at/</a>
Belgium	<a href="https://www.geo.be">https://www.geo.be</a> (FED) <a href="https://www.geopunt.be">https://www.geopunt.be</a> (FL) <a href="https://geoportail.wallonie.be">https://geoportail.wallonie.be</a> (WAL) <a href="https://geobru.irisnet.be">https://geobru.irisnet.be</a> (BRU)
Bulgaria	<a href="https://inspire-catalogue.egov.bg/">https://inspire-catalogue.egov.bg/</a>
Croatia	<a href="http://www.geoportal.nipp.hr">www.geoportal.nipp.hr</a>
Cyprus	<a href="https://eservices.dls.moi.gov.cy/#/national/inspiregeoportalmapviewer">https://eservices.dls.moi.gov.cy/#/national/inspiregeoportalmapviewer</a>

<sup>460</sup> See: <https://julkinen.vayla.fi/oskari/?lang=en>

Country	Link to national access point
Czech Republic	<a href="http://www.geoportal.gov.cz">www.geoportal.gov.cz</a>
Denmark	<a href="http://www.geodata-info.dk">www.geodata-info.dk</a>
Estonia	<a href="http://www.inspire.maaamet.ee">www.inspire.maaamet.ee</a>
Finland	<a href="https://kartta.paikkatietoikkuna.fi/">https://kartta.paikkatietoikkuna.fi/</a>
France	<a href="http://www.geocatalogue.fr/">http://www.geocatalogue.fr/</a>
Germany	<a href="http://www.geoportal.de">www.geoportal.de</a>
Greece	<a href="http://geodata.gov.gr/en/dataset?groups=transportation">http://geodata.gov.gr/en/dataset?groups=transportation</a>
Hungary	<a href="http://www.inspire.gov.hu">www.inspire.gov.hu</a>
Ireland	<a href="https://inspire.geohive.ie/geoportal/catalog/main/home.page">https://inspire.geohive.ie/geoportal/catalog/main/home.page</a>
Italy	<a href="http://www.pcn.minambiente.it">www.pcn.minambiente.it</a>
Latvia	<a href="http://www.geolatvija.lv">www.geolatvija.lv</a>
Lithuania	<a href="http://www.geoportal.lt">www.geoportal.lt</a>
Luxembourg	<a href="https://geocatalog.geoportal.lu/geonetwork/">https://geocatalog.geoportal.lu/geonetwork/</a>
Malta	<a href="http://www.msdi.data.gov.mt">www.msdi.data.gov.mt</a>
Netherlands	<a href="http://www.nationaalgeoregister.nl/geonetwork/">http://www.nationaalgeoregister.nl/geonetwork/</a>
Poland	<a href="https://mapy.geoportal.gov.pl/imap/">https://mapy.geoportal.gov.pl/imap/</a>
Portugal	<a href="http://sniq.dgterritorio.gov.pt">sniq.dgterritorio.gov.pt</a>
Romania	<a href="http://geoportal.gov.ro/metadata_catalog/#searchPanel">http://geoportal.gov.ro/metadata_catalog/#searchPanel</a>
Slovakia	<a href="http://www.geoportal.gov.sk">www.geoportal.gov.sk</a>
Slovenia	<a href="http://www.geoportal.gov.si/eng/data-collections/">http://www.geoportal.gov.si/eng/data-collections/</a>
Spain	<a href="http://www.idee.es/csw-inspire-idee/">http://www.idee.es/csw-inspire-idee/</a>
Sweden	<a href="https://www.geodata.se/geodataportalen/">https://www.geodata.se/geodataportalen/</a>

### 3.6.2 To be situation: extending the PSI HVD rules to these datasets

This section presents the recommended measures for publication, as well as the expected costs and benefits of including these datasets under the scope of the PSI Directive as HVD.

#### 3.6.2.1 Recommended measures for publication

This section presents the recommendations of the study team for the publication of the datasets in scope.

##### *Inland waterways and river infrastructure data*

As presented in the previous sections, the current modalities for publication of inland waterway infrastructure data are rather similar throughout the EU. Some additional key considerations have been shared by reusers aimed at further facilitating the reuse of this data, some of which have already been piloted in/adopted by some Member States in the context of the Connecting Europe Facility (CEF) funded, multi-beneficiary "RIS COMEX" project.

In a nutshell, RIS COMEX is a pilot project launched in 2016 in which 13 European countries are working together, under the coordination of the Austrian Waterway Administration (via Donau), towards seamless and sustainable operation of cross-border RIS Services.<sup>461</sup> Harmonisation of operational exchange of underlying RIS data is one of the key objectives of the project, and fairway data is directly tackled through dedicated (sub-)activities under the project<sup>462</sup>. The project focuses on six waterway corridors – encompassing the main inland waterway transport lines across Europe:

- **Danube:** Germany – Austria – Slovakia – Hungary – Croatia – Serbia – Romania – Bulgaria
- **Rhine:** France – Germany – the Netherlands
- **Moselle:** Germany – Luxembourg – France
- **Amsterdam-Antwerp-Liège/Brussels:** Belgium – the Netherlands
- **Dunkerke-Scheldt:** France – Belgium
- **Elbe-Weser:** Czech Republic – Germany

As such, RIS COMEX encompasses all Member States discussed in section 3.6.1.1, with Poland contributing to the project as a “Cooperating Partner”. Therefore, the recommended measures for publication for static and dynamic data, as well as for the ENC’s presented in the table below have been elaborated in line with the project’s requirements, in order to ensure consistency. Most importantly, this means that in essence these recommended measures for publication are already agreed upon by Member States.

Table 47 – Recommended modalities for publication of inland waterway infrastructure data

Dimensions		Static	Dynamic/Urgent	Electronic Navigational Charts
<i>Openness-data specification</i>	<i>License (terms of use)</i>	CC-BY 4.0 No terms of use		
	<i>Format</i>	CSV, XML, (geo)JSON	CSV, XML, (geo)JSON	XML, 000, WMS
	<i>Machine-readability</i>	Mandatory		
	<i>Availability of API, bulk download</i>	Bulk download mandatory. Webservices and API recommended.	Web service, (OGC) API and bulk download mandatory.	
<i>Documentation</i>	<i>Metadata (dataset content description)</i>	Complete (*.csv document available)		
	<i>Data linking</i>	N/A		
	<i>Documentation (incl. structure and semantics)</i>	RIS Index Encoding Guide 3.0 developed by the Joint Task Force on the RIS Index	Commission Implementing Regulation (EU) 2018/2032 of 20 November 2018 for NtS	Commission Implementing Regulation (EU) 2018/1973 of 7 December 2018 for Inland ECDIS
	<i>Shared vocabularies</i>	RIS Index Encoding	NtS Standard 4.0	Inland ECDIS

<sup>461</sup> See: <https://www.riscomex.eu/>

<sup>462</sup> Under Activity 2: Corridor RIS Preparation, Sub-Activity 2.2 Level 1 Case definition aims at defining the requirements and services enabling optimal route planning based on reliable and complete fairway and infrastructure data. Under Activity 3: Corridor RIS Implementation, Sub-Activity 3.2 Level 1 Services aims to further specify and implement the services as defined in Sub-Activity 2.2. More information, see: <https://www.riscomex.eu/activities/>

Dimensions		Static	Dynamic/Urgent	Electronic Navigational Charts
		Guide 3.0		Standard 2.5
	<i>Taxonomies</i>	N/A		
Completeness	<i>Traceability</i>	N/A		
	<i>Update frequency and timeliness</i>	When necessary	Daily to (near) real time	Monthly (for shallow sections)
	<i>Granularity</i>	Individual waterway km level		National waterway network and cross-border nodes
	<i>Key attributes</i>	ISRS Location Code		

As the table suggests, the rather similar recommended modes of provision apply to the three categories of datasets in scope, with some variances data formats, documentation, shared vocabularies and frequency of updates. The justifications for each of these recommended measures are the following:

- Concerning **licenses**, the lower intensity option suggests to adopt open licenses allowing for any type of reuse (including of course commercial reuse and transmission). It appeared from the data collection regarding the 'as is' modalities for provision that in many countries licenses were currently not specified, therefore it can be expected that a formal transition to such open licenses should not raise any major concerns.
- Format-wise**, several options – already in use in the majority of Member States, have been discussed with stakeholders. Similarly to the transport networks datasets (see hereafter), it appears that the availability of multiple relevant formats does not hinder the potential for reuse of the datasets, which is the reason why multiple options are kept as recommended, bottom-line being that all of these formats are **machine-readable** and therefore in line with the PSI/Open Data Directive principles.
- This lower intensity option suggest to ensure the accessibility of the data via **webservices, APIs and bulk downloads** (except for static data where bulk download would suffice due to the nature of the – static – information prone to very few changes). The main interest here lies in having seamless access to (reuse) the information through mapping applications for routing/voyage planning while also being able to download all datasets at once in order to freely develop any other applications/use-cases. It should still be noted that web-services indeed allow to visualise the datasets but not necessarily to entirely reuse them, as such APIs make the data more reusable in the sense of the Open Data Directive.
- For APIs to be taken up however by all sort of reusers, **clear documentation** including structure of the datasets and **shared vocabularies** become particularly important. To this extent the application of the RIS Index Encoding Guides and RIS Implementing legislation and standards for NtS and ECDIS play a key harmonising role. Similarly, this low intensity option requires that complete and easily accessible **metadata** is provided in order to facilitate reuse of these datasets.
- Concerning the **data linking, taxonomy and traceability of data**, reusers provided no feedback and therefore no specific recommended measures could be developed on these aspects.
- The **timeliness or frequency of update** of the datasets is considered as key for the reusability of dynamic and urgent data, which again is directly linked to their nature. Here datasets should be provided in (near) real-time in order to guarantee the accuracy of RIS services developed on the basis thereof. On the other hand, static and Inland ECDIS require fewer updates due to their more static



nature, with at least monthly updates to ENCs (especially in shallow areas) and a needs-basis for the static infrastructure data.

- The **granularity of the datasets** is also a key aspect, with datasets provided preferably at the level of each individual waterway kilometre point for static and dynamic data, and ENCs presenting the entire national waterway network including the cross-border links and nodes. These are motivated by both safety/security aspects for the deepest level of detail as possible, as well as the multi-modal and cross-border transport ambitions known to the inland waterway sector.
- Last but not least, the ISRS Location Code is the **key attribute** to be referenced for all datasets, as these allow to disambiguate and uniquely identify all datasets. Similarly to other standards referred to in this section, the ISRS Location Codes are already widely reused in the current modes of provision of inland waterway infrastructure data by Member States.

The expected costs linked to the application of these recommended measures are discussed in the next section.

### *Transport networks data*

The recommended measures for publication concerning transport network datasets build extensively on the good practices and approaches developed for the INSPIRE Directive and there has been little debate amongst stakeholders about the main features that should characterise the provision of these datasets. The table below summarises the recommended measures which have been established by the study team for all the publication dimensions and for all transport modes (road, rail, water, air and cableways). As the table clearly shows, there is a limited need to make distinctions between the different transport modes in terms of dimensions of provision. This is because the datasets share the same key characteristics in terms of type of data (static data only) and due to the fact that reusers' requests are globally aligned across the modes.

Table 48 – Recommended modalities for publication of transport networks data

<b>Dimensions</b>		<b>Road transport</b>	<b>Rail transport</b>	<b>Water transport</b>	<b>Air transport</b>	<b>Cableways</b>
<i>Openness-data specification</i>	<i>License (terms of use)</i>	CC-BY 4.0 No terms of use				
	<i>Format</i>	GML, GeoPackage, GeoJSON				
	<i>Machine-readability</i>	Mandatory				
	<i>Availability of API, bulk download</i>	Web service, (OGC) API and bulk download				
<i>Documentation</i>	<i>Metadata (dataset content description)</i>	Complete (*.csv document available)				
	<i>Data linking</i>	Links to national INSPIRE Geoportals and datasets as relevant.				
	<i>Documentation (incl. structure and semantics)</i>	Complete and available				
	<i>Shared vocabularies</i>	INSPIRE data specifications are recommended but not mandatory.				
	<i>Taxonomies</i>	N/A				
<i>Completeness</i>	<i>Traceability</i>	N/A				

Dimensions		Road transport	Rail transport	Water transport	Air transport	Cableways
	<i>Update frequency and timeliness</i>	When necessary				
	<i>Granularity</i>	From local to national, including links with cross-border networks, if and where available				
	<i>Key attributes</i>	Any national identification code; latitude and longitude				

Further explanations on the recommended modes of provision described in the table are provided below:

- Concerning **licenses**, the lower intensity option suggests to adopt open licenses allowing for any type of reuse (including of course commercial reuse and transmission). Many countries already rely upon such open licenses and this therefore constitutes a good practice which is quite widespread. At the same time, terms of use are rather frequent, although they are not considered as acceptable by reusers. In fact, as per other thematic areas, terms of use impose restrictions (of various extents) on the possibility to use the datasets and therefore are considered by stakeholders against the “spirit” of the PSI/Open Data Directive.
- Format-wise**, the lower intensity intervention suggests to use several possible options, listed above, which are all already in use and are all valuable from a reuser perspective. Differently from other thematic areas, there was no indications of one or two formats in particular which would constitute best practices and should be adopted by all countries. In this domain, the availability of multiple relevant formats does not seem to hamper the reusability of datasets and, for this reason, the list of possible formats is voluntarily left quite open. **The machine readability** of the datasets’ format however remains the main condition for considering the format adequate under the PSI/Open Data Directive.
- This intervention would allow the possibility of accessing information **both through APIs and via bulk download**. The reason for this request lies in the different use cases that APIs and bulk download would support (i.e. map applications would rely on APIs while logistic analysis would rely on bulk download) and the greater freedom that both options would entail for reusers developing their own services. In this respect, it must be noted here that many countries today provide web-services allowing to visualise the data rather than APIs. While this is changing fast and more countries are moving/have already moved towards APIs (i.e. France, Sweden, Ireland...) and towards making these datasets more accessible for developed and non-expert users<sup>463</sup>, the reliance on web-services might be an obstacle for reusers. Most transport network web services in fact focus on visualisation and access to data rather than reuse. In this context, the development of APIs is considered as a key success factor from an Open Data perspective (see section 3.8). For APIs to be taken up however by all sort of reusers, **clear documentation** including structure of the datasets and semantics becomes particularly important and should be provided as well.

Nonetheless, data holders expressed their concerns on the deletion/prohibition of all terms of use/provision of unlimited access to datasets and especially from the perspective of the stability of the IT infrastructure: a recent example from Austria in fact demonstrated that, in the absence of clear rules for accessing APIs for instance, the IT system for the provision of the data might be overloaded

<sup>463</sup> See for instance Barbero, M., Lopez Potes, M., Vancauwenberghe, G. and Vandenbroucke, D., The role of Spatial Data Infrastructures in the Digital Government Transformation of Public Administrations , Publications Office of the European Union, Luxembourg, 2019, <https://ec.europa.eu/jrc/en/publication/role-spatial-data-infrastructures-digital-government-transformation-public-administrations>, p. 53

and this might provoke the temporary failure of the service for all reusers<sup>464</sup>. Therefore, this medium intensity intervention could introduce technical requirements in order to regulate the access to the datasets by for example limiting the number of daily calls per stakeholder. This would allow to protect the datasets, which could not be queried indefinitely through the API.

- **Metadata** should be provided in a complete and accessible manner. Fortunately, thanks to the INSPIRE Directive, metadata are already very widely provided (see the INSPIRE Geoportal for the full picture). Complete and easily accessible metadata should continue to be provided in order to facilitate reuse of these datasets.
- The aspect of **data linking** was only superficially discussed with stakeholders and solely from the perspective of facilitating reuse by clearly establishing/providing unique entry points/platforms for all datasets. While the establishment of one stop shops/centralised platforms for the provision of these datasets goes well beyond the scope of the PSI/Open Data Directive, stakeholders clearly argued that the fragmentation in the data provision (across multiple portals at the national and European level) is a major obstacle for its reuse and that one stop shop national portals should be established (or the ITS Directive National access portal should be provided with all these datasets)<sup>465</sup>. For this reason, this medium intensity intervention recommends to consider data linking to at least the national INSPIRE Geoportal as very important if not essential.
- In terms of **shared vocabularies**, this intervention suggests to follow the INSPIRE Directive, which provides sufficient clarity and harmonisation on the transport network theme. These vocabularies should be considered as a reference (although not imposed).
- Concerning the **taxonomy and traceability of data**, reusers provided limited/no feedback and therefore no specific recommended measures could be developed on these aspects.
- The **timeliness or frequency of update** of the datasets is considered of great importance by stakeholders, despite the datasets in scope are mainly static. In this context, this intervention recommends that the datasets are updated when necessary. This implies that some different modes might be updated more frequently than others, but this will be a decision to be made by each data holder.
- The **granularity of the datasets** is also considered as key by reusers who strongly advocate for as granular transport network data as possible, in terms of the scope coverage. While many countries focus on national transport networks and on the national dimension mainly, the local dimension acquires more and more importance from a multi-modal transport perspective and when looking at the uptake of automated and connected vehicles. The desired and recommended granularity hence covers datasets going from local to national and including links with cross-border networks. Nonetheless, this might be a challenge for data holders and might not be possible in all countries, as very granular data for the local level in particular are not always available and the Directive cannot impose the collection of new data. As reusers were adamant in asking, whenever and wherever available, the lowest level of granularity possible, it is recommended for the PSI/Open Data Directive to cover from the local to the national level as to include local datasets for those countries holding them and already making them available.
- Finally, some **key attributes** which are essential for the transport network datasets are elements ensuring their unique identification and disambiguation such as any national identification code or name, as well as the coordinates of the transport network feature.

The expected costs linked to the application of these recommended measures are discussed in the next section.

---

<sup>464</sup> As mentioned by one stakeholders during the online focus group, recently the Austria geoportal suffered from a temporary failure due to a user programming 1000 calls per seconds towards the API platform, which could not cope with such an overload, although limited in time.

<sup>465</sup> Online focus group

### 3.6.2.2 Expected costs

This section presents the expected costs arising from the application of the recommended measures for publication.

#### *Inland waterways and river infrastructure data*

None of the Member States consulted as part of the data collection activities were able to provide quantitative data about the current (let alone future) cost of provision of inland waterway and river infrastructure data. The reason for this was shared unanimously by respondents: it is very challenging to distinguish the costs related to the provision of these specific datasets from other internal processes. While this lack of data does not allow to conduct a proper cost-benefit analysis at micro-level as for other thematic areas, this does not restrict the assessment of costs from a more qualitative perspective, complemented with samples of quantitative data.

As already presented above, current modalities for the provision of inland waterways and river infrastructure data are already rather similar across the EU, and most importantly, they are largely aligned with the recommended *to be* situation. In addition, an EU funded initiative encompassing all relevant Member States is already bridging all data provision practices together, and in line with the recommended practices, through the RIS COMEX pilot project.

On the one hand, this means that the costs at individual Member State level can be expected to be rather low, as current practices are already rather close to the recommended ones. Most notably, charging practices are inexistent in this field, and thus there is no impact on the business model of the data holders. On the other hand, a significant share (if not all) of the costs will be tackled under the RIS COMEX pilot project, in particular the "VisuRIS COMEX" platform in which all European fairway and infrastructure data will be made available, in line with the recommended measures for publication discussed in this report.

In order to illustrate these costs, the table below provides an approximation of the budget allocated under the RIS COMEX project for the development and implementation of the VisuRIS COMEX platform, which shall be brought into operation by early 2021. These costs are challenging to estimate as the platform is a cloned and extended version of the already existing VisuRIS system in Belgium (Flanders) and because the RIS COMEX budget does not solely focus on the costs of provision of the data, but also includes staff costs related to the development of (technical) specifications of the platform, the discussion of the various data formats and interfaces, etc. Nonetheless, via Donau has provided a **rough estimation of approximately €6 million** being invested in order to realise the Fairway- and Infrastructure related data provision under VisuRIS COMEX, aside future operational/maintenance costs of approximately **€800 thousand on a yearly basis** as presented in the table below. It should be noted that the figures below account for all 13 participating countries as a whole.

Table 49 – Estimation of costs related to fairway and infrastructure data provision under VisuRIS COMEX

Cost category	Estimated costs
Infrastructural costs, including adaptation of national infrastructure and interfaces to provide related data to common system	€3 million
Data transformation costs	€0 <sup>466</sup>
Operational costs, including hosting, maintenance and user support	€800 thousand/year

<sup>466</sup> There are no explicit costs for data transformation: the data is already provided in the correct format by the national data sources to the common VisuRIS COMEX system. Any potentially necessary data transformation in terms of data format and e.g. xml message structure is done automatically by the related interfaces and backend modules.

Cost category	Estimated costs
Lost income for data suppliers	€0 <sup>467</sup>
Other costs including legal advice on GDPR, training costs, etc.	€100 thousand
Project partner staff costs related to this specific topic	€3 million

All in all, capital expenditure presented in the table above is undertaken in the framework of the CEF funding. It is expected that in the future, operational costs (estimated in the table above) of the VisuRIS COMEX would be borne by Member States themselves. The financial impact of these costs on Member States involved has been indicatively illustrated below.

Based on information available through desk research<sup>468</sup>, operating budgets of authorities in charge of inland waterway transport (or alternatively though less accurate, relevant ministries' budgets) have been compared to the potential operational costs to be borne at Member State level. Considering a simple and even division across the 13 participants (which would likely not be the case as the most plausible division would possibly take into account the countries' exposure to inland waterway traffic on their respective waterways), this would represent approximately €62 thousand on a yearly basis. In comparison with the countries' operational budget for inland waterway (transport) authorities, these costs would then indicatively range from 0.0015% to 1.90% of their annual budget, which can indeed be considered as fairly low.

Table 50 – Indicative comparison of impact of VisuRIS COMEX operational costs on Member States' annual budgets

Country	Authority	Year	Budget (€)	% Budget
Austria	Ministry of Transport, Innovation, and Technology (Verkehr, Innovation und Technologie)	2019	4 008 000 000 <sup>469</sup>	0.0015%
Belgium	Vlaamse Waterweg	2017	400 000 000 <sup>470</sup>	0.0154%
Germany	BundeswasserstraBe + WasserstraBen- und Schifffahrtsverwaltung des Bund	2020	2 034 014 000 <sup>471</sup>	0.0030%
France	Voies navigables de France	2019	1 255 924 745 <sup>472</sup>	0.0049%
Luxembourg	Navigation et transports fluviaux	2019	3 285 831 <sup>473</sup>	1.8728%
Netherlands	Rijkswaterstaat	2018	4 700 000 000 <sup>474</sup>	0.0013%
Slovakia	Ministry of Transport, Construction, and Regional Development	2019	1 153 830 897 <sup>475</sup>	0.0053%

<sup>467</sup> Charging practices are inexistent for these datasets across all Member States.

<sup>468</sup> The study team was able to find data for seven countries only via desk research.

<sup>469</sup> See: [https://service.bmf.gv.at/BUDGET/Budgets/2018\\_2019/bfq2019/Bundesfinanzgesetz\\_2019.pdf](https://service.bmf.gv.at/BUDGET/Budgets/2018_2019/bfq2019/Bundesfinanzgesetz_2019.pdf)

<sup>470</sup> See : [https://www.vlaamsewaterweg.be/sites/default/files/de\\_vlaamse\\_waterweg\\_corporate\\_brochure.pdf](https://www.vlaamsewaterweg.be/sites/default/files/de_vlaamse_waterweg_corporate_brochure.pdf)

<sup>471</sup> See : <https://www.bundeshaushalt.de/#/2020/soll/ausgaben/einzelplan/12.html>

<sup>472</sup> See: <https://www.vnf.fr/vnf/app/uploads/2019/10/BO66.pdf>

<sup>473</sup> See: <https://budget.public.lu/dam-assets/lb/budget2020/links-dokumenter/budget-2020-volume1.pdf>

<sup>474</sup> See: <http://publicaties.minienm.nl/download-bijlage/109214/annual-report-rijkswaterstaat-2018.pdf>

<sup>475</sup> See: <https://www.mindop.sk/transparentnost/rozpocet/schvaleny-rozpocet-vydavkov-kapitoly-mdvrr-sr-mdpt-sr-223>

### Transport networks data

To provide some conclusions on the costs of implementing the recommended measures for publication described in the previous section, the team started by analysing a) the current costs borne by Member States for the provision of transport network datasets and b) which would be the most expensive changes that would occur to the current state of play based on the recommendations adopted.

The analysis of the current costs of provision is made extremely difficult by the lack of primary data and by the difficulty of stakeholders to differentiate between the costs of provision for these specific datasets and the total costs of provision of INSPIRE or national geoportals for instance. **Only two of the stakeholders consulted** during the data collection were able to provide some sort of quantitative inputs on costs of provision today, and none were able to extrapolate on the possible costs of provision of tomorrow (which would be entailed by our recommended measures for publication), as presented in the box below. This is hardly surprising considering that lack of availability and precision concerning data on costs has also been encountered for the annual reporting on INSPIRE since the very first exercise<sup>476</sup>.

German stakeholders were able to provide cost indications for the provision of rail transport networks and related infrastructure. DB Netz AG moved from charging for the railway transport network data to providing it as open data in 2015 to fulfill the INSPIRE requirements. DB Netz AG receives no compensation for the loss of earnings. INSPIRE data is available via standardised geospatial web services in machine readable formats (GML) and free of charge and, thus, mainly meets the most important requirements of PSI HVD. In general estimations of current costs are difficult, because of interdependencies of data provision processes with other processes, and it was not possible to estimate possible future costs.

Table 51 – Current costs borne by DB Netz AG for the provision of rail transport networks data

Cost category	Current costs
Infrastructural costs	~ €55 thousand per year
Data transformation costs	~ €20 thousand per preparation process
Operational costs (replying to users' requests)	~ €10 thousand per year
Lost income for data supplier	~ €40 thousand for each request

In Finland, the DigiRoad service<sup>477</sup> in particular takes account of road transport networks (including all national road highways, bridges, etc), and these datasets are used by hundreds of businesses in the private sector, including mostly mapping and location services providers.

Regarding the costs of provision of this data in their current state, the overall investment in infrastructure and data transformation was estimated to a magnitude of approximately €1 million overall since 2013, with approximately €10s of thousands on a yearly basis necessary for the maintenance of the platform and data. Today, the DigiRoad service budget is €1.4 million per annum, which also includes the entire development of the platform and all operations/maintenance tasks. It should be noted that beyond the INSPIRE and ITS Directives, national law makes it compulsory to collect data

<sup>476</sup> As an example, see what the Irish INSPIRE country report 2016 writes about costs: "the total costs associated with implementing the Directive are not systematically being measured across all stakeholders", [http://cdr.eionet.europa.eu/ie/eu/inspire/reporting/envvwy5eza/20170706\\_INSPIRE\\_Triennial\\_Report\\_2016.docx/management\\_document](http://cdr.eionet.europa.eu/ie/eu/inspire/reporting/envvwy5eza/20170706_INSPIRE_Triennial_Report_2016.docx/management_document) or what the Spanish country report states: "No existen estudios recientes que evalúen el coste de la aplicación de la Directiva INSPIRE, ni tampoco estudios más generales evaluando el impacto coste/beneficio del desarrollo de las IDE a nivel nacional", [http://cdr.eionet.europa.eu/es/eu/inspire/reporting/envvzxdeg/2016\\_Reporte\\_INSPIRE\\_2013\\_2015\\_ES.pdf](http://cdr.eionet.europa.eu/es/eu/inspire/reporting/envvzxdeg/2016_Reporte_INSPIRE_2013_2015_ES.pdf)

<sup>477</sup> See: <https://vayla.fi/web/en/open-data/digiroad>

regarding transport networks in view of networks maintenance and mapping, and thus, this data would be collected anyhow.

Furthermore, the most recent INSPIRE Country reports<sup>478</sup> can be used to derive some additional limited indications on the magnitude of different costs categories which are also applicable for the transport network data theme (at least under INSPIRE). For instance, the 2016 Inspire Country Report for Italy<sup>479</sup> lists a number of cost categories for the set up and maintenance of the data sharing infrastructure and metadata (see table below).

Cost categories for Italy (linked to the implementation of the INSPIRE Directive, including transport networks data theme – 2016)<sup>480</sup>

- Set up of the infrastructure: €100 000 + 6 FTEs (= €345 000<sup>481</sup>), €445 000 euro in total
- Maintenance of the infrastructure: 12 FTEs (= €690 000<sup>482</sup>) annually
- Set up costs of the metadata in total: €5000 + 12 FTEs (= €690 000<sup>483</sup>)
  - o Creation of metadata for discovery and for use: 2 FTEs + 1 FTE (= €172 800<sup>484</sup>)
- Maintenance costs of the metadata: 6 FTEs (= €345 000<sup>485</sup>)

Additional evidence on costs of INSPIRE implementation comes from the Spanish country report which states that the costs of IT maintenance of the geoportal at the national level was of €300 000 annually for the period 2013-2015<sup>486</sup>. Finally, the Luxembourgish country report establishes that the costs of maintaining and running the geoportal for the same period (including all costs related to human resources) was of around €1.1 million per year<sup>487</sup>. These figures should be considered with extreme caution, first of all because they are out of date and secondly because they relate to the provision of the geoportals in general and not specifically of the transport networks datasets. Nonetheless, they all converge towards estimating annual costs for maintaining the infrastructure of in between €300 000 (for the IT part only) and €1.1 million per year (everything included). This can give an indicative idea of the baseline against which the recommended measure for publication should be assessed in terms of cost implications.

Further to analysing the as-is cost situation, it is necessary to make an assessment of the recommended measures for provision in order to identify those that would have the highest impact on Member States in terms of costs of implementation. It is worth reminding here that, for this specific dataset which falls

<sup>478</sup> Unfortunately, the most recent country reports for INSPIRE monitoring date back to 2016 (or even 2015) for certain countries. See: <https://inspire.ec.europa.eu/INSPIRE-in-your-Country>

<sup>479</sup> Italy Country Report 2016, INSPIRE, [http://cdr.eionet.europa.eu/it/eu/inspire/reporting/envvzzgq/Italy\\_Country\\_Report\\_2016\\_18052016.docx/manage\\_document](http://cdr.eionet.europa.eu/it/eu/inspire/reporting/envvzzgq/Italy_Country_Report_2016_18052016.docx/manage_document)

<sup>480</sup> Italy Country Report 2016, INSPIRE, [http://cdr.eionet.europa.eu/it/eu/inspire/reporting/envvzzgq/Italy\\_Country\\_Report\\_2016\\_18052016.docx/manage\\_document](http://cdr.eionet.europa.eu/it/eu/inspire/reporting/envvzzgq/Italy_Country_Report_2016_18052016.docx/manage_document)

<sup>481</sup> Based on Italian cost of labour taken from Eurostat (2019 statistics) and a number of working days of around 250 per year. See: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages\\_and\\_labour\\_costs](https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages_and_labour_costs)

<sup>482</sup> Based on Italian cost of labour taken from Eurostat (2019 statistics) and a number of working days of around 250 per year. See: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages\\_and\\_labour\\_costs](https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages_and_labour_costs)

<sup>483</sup> Based on Italian cost of labour taken from Eurostat (2019 statistics) and a number of working days of around 250 per year. See: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages\\_and\\_labour\\_costs](https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages_and_labour_costs)

<sup>484</sup> Based on Italian cost of labour taken from Eurostat (2019 statistics) and a number of working days of around 250 per year. See: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages\\_and\\_labour\\_costs](https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages_and_labour_costs)

<sup>485</sup> Based on Italian cost of labour taken from Eurostat (2019 statistics) and a number of working days of around 250 per year. See: [https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages\\_and\\_labour\\_costs](https://ec.europa.eu/eurostat/statistics-explained/index.php/Wages_and_labour_costs)

<sup>486</sup> Spain Inspire Country report 2016, [http://cdr.eionet.europa.eu/es/eu/inspire/reporting/envvzxdeg/2016\\_Reporte\\_INSPIRE\\_2013\\_2015\\_ES.pdf](http://cdr.eionet.europa.eu/es/eu/inspire/reporting/envvzxdeg/2016_Reporte_INSPIRE_2013_2015_ES.pdf)

<sup>487</sup> Luxembourg INSPIRE Country report 2016, [http://cdr.eionet.europa.eu/lu/eu/inspire/reporting/envvzleq/2016\\_INSPIRE\\_Country\\_Report\\_2016\\_Luxemburg\\_final.pdf/manage\\_document](http://cdr.eionet.europa.eu/lu/eu/inspire/reporting/envvzleq/2016_INSPIRE_Country_Report_2016_Luxemburg_final.pdf/manage_document)



under the INSPIRE Directive's rules, one can expect costs of implementation to be lower than in other cases, as there is already a general availability of these datasets and they are provided for free and in an harmonised manner (at least for the INSPIRE datasets) across all countries. Furthermore, as mentioned in the previous paragraphs, large scale IT infrastructure for providing these datasets already exist (national or INSPIRE geoportals), although not always in line with our recommended measures.

Based on the analysis of the state of play of provision, there seem to be two main recommended measures which would entail considerable or at least some investments for Member States:

- 1. The establishment of API infrastructures:** the analysis of the as-is situation showed that APIs are not yet a very widespread practice for the provision of transport network datasets across countries, with a few exceptions of course (i.e. France, Ireland). This entails that a majority of countries would have to set up APIs and add this feature to their national/INSPIRE geoportals. As discussed in other chapters of this report, the costs of setting up APIs vary considerably depending on the technical characteristics desired for this infrastructure and on the Member States<sup>488</sup>. The previous impact assessment for the PSI Directive (from 2018) estimates that setting up and maintaining APIs could cost in between €30 000 to €2 million per year per Member State and that a medium size API set up (which should be sufficient for providing the static transport network datasets) costs in average €50 000 for its development<sup>489</sup>. Considering that minimum 20 countries would have to establish such infrastructure, this would lead to in between €600 000 and max €40 million spent at the EU level or a more realistic figure of around 1million euro of investments (based on the average costs of €50 000 per API).
- 2. The provision on timeliness/frequency of update:** also considering that there is limited evidence available on the frequency of update of the datasets (for both the national and INSPIRE datasets), it seems possible that the recommended measures for timeliness/frequency of update, which are mostly based on users' needs, do not correspond yet to the standard practices of Member States in this respect. Updating the datasets more often would therefore trigger some costs for the data holders. This is one of the reason why we suggest leaving to countries the choice between complying to the daily/weekly/monthly update rules or using a "real time" approach and update the datasets whenever there are changes. The magnitude of these costs however remains unclear/unquantifiable, also considering that no country was able to share their data handling costs for today for this specific data theme.

Considering the elements discussed above and especially the costs of setting up APIs (which remains the main quantifiable aspect), it can be argued that the recommended measure would cost maximum a few millions euro at the European level in terms of one-off costs. These costs would of course be close to 0 for those countries (i.e. France, Ireland) already disposing of an API. For the vast majority of countries on the other hand the establishment of the API would be the main cost driver and the size of this investment would clearly depend on the type of infrastructure set up. From this perspective, the recommended measures on licenses and terms of use also have a certain importance: in fact, if no terms of use are allowed (including capping the number of APIs calls per second or per hour for each user) this means that more solid IT infrastructures (having greater hosting capacity and server bandwidth) need to be established and costs for Member States will be higher. The interaction between the recommended measures for the APIs and the licensing aspects can be taken into account for the development of policy options and in case the costs imposed on Member States would be considered as too high by the legislator.

---

<sup>488</sup> Acquisition of IT infrastructure has been found cheaper in some countries (i.e. Baltic countries) than in others

<sup>489</sup> Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, <https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en>



To conclude, the cost implications of including the transport network in the high-value datasets list would be rather short term (one off costs for the establishment of the APIs) and remain quite limited, especially when compared to the current funding levels of the INSPIRE infrastructures and geoportals.

### 3.6.2.3 Expected benefits

This section considers the expected benefits of the inclusion of inland waterway and river infrastructure data, and transport networks data, as HVDs under the PSI/Open Data Directive. The discussion is presented for both topics as one, as a similar reasoning applies to both.

As widely presented in previous sections tackling the current value for reuse and benefits of these datasets, it appears they have a number of use cases, which contribute to the delivery of not only economic benefits, but also environmental and social benefits:

- **Economic:** notably in terms of reduction of costs for economic operators in the (public/waterborne) transport and logistics industry following efficiency gains, and provision of new added-value services by mapping and location as well as RIS service providers, etc;
- **Environmental:** notably in terms of reduction of CO2 emissions following the optimal use of routes and multimodal mobility, etc;
- **Social:** notably in terms of increased safety and security in the (public/waterborne) transport and logistics industry and for citizens, and creation of new employment possibilities in the services mapping and location services as well as RIS industry.

It has been made clear that many of the use cases presented earlier rely on both, traffic and travel related datasets along with the datasets in scope of the current study. The former are already covered by the ITS Directive and are thus to be made available for re-use by data holders through National Access Points, in a machine readable format but not necessarily through open licenses and free of charge. While inland waterway and river infrastructure data as well as transport networks are also already covered by existing legislation (i.e. RIS Directive and INSPIRE), the inclusion of these datasets as HVDs will complement existing efforts in promoting the reuse of such data, through the formalisation of the obligation of providing these for free, in harmonised and machine readable formats and both through bulk download and APIs.

Overall, it can legitimately be expected that this will further ease the access to and reuse of data by reusers identified, notably for the provision of seamless, EU-wide information services underpinned by harmonised data provision modalities by data holders. This is further supported by the fact that Member States consulted declare having received requests from reusers to further facilitate the access and reuse of these datasets throughout the EU, notably in order to leverage these in various mobile applications and softwares for the provision of "mobility" services.

Regarding inland waterway data in particular, stakeholders have also mentioned some potential value related to artificial intelligence developments, e.g. in the context of ETA calculations or water-level predictions based on the fairway and infrastructure data as well as some traffic/travel data. The harmonisation of data provision according to the recommended measures, notably the availability of data in given machine readable formats and through APIs in all Member States, would be a key enabler in this context.

Regarding transport networks specifically, the expected benefits could be further widened should all transport networks related datasets i.e. both the INSPIRE datasets and the 'additional/other' ones provided in national platforms, be considered as HVDs. Indeed, as presented earlier for Finland, Member States possess and publish many more datasets relating to transport networks, which could equally gain

value from being considered as HVDs, and thus being provided with open licences, free of charge, in harmonised machine readable formats through APIs and bulk download as per our recommended modalities for transport networks data. To illustrate this thought, the table below provides a view on the number of view and download of INSPIRE and 'other' transport datasets from the Finnish Transport Open Data Portal. These highlight the striking difference in favour of the latter (i.e. the 'other' transport datasets), and therefore testify of the already existing value for reuse of these 'other' datasets.

Table 52 – WMS requests and downloads of transport networks and infrastructure datasets from Finnish transport open data portal (2018)

Datasets & feature	Requests	Download (MB)
TransportNetworks:TN.AirTransportNetwork.AerodromeArea	72280	29
TransportNetworks:TN.AirTransportNetwork.AerodromeNode	233261	750
TransportNetworks:TN.AirTransportNetwork.AirRouteLink	1583	44
TransportNetworks:TN.AirTransportNetwork.AirRouteLinkSequence	1170	19
TransportNetworks:TN.AirTransportNetwork.AirSpaceArea	1840	49
TransportNetworks:TN.AirTransportNetwork.ControlTower	859	21
TransportNetworks:TN.AirTransportNetwork.DesignatedPoint	923	23
TransportNetworks:TN.AirTransportNetwork.InstrumentApproachProcedure	743	19
TransportNetworks:TN.AirTransportNetwork.Navaid	1461	21
TransportNetworks:TN.AirTransportNetwork.RunwayArea	1341	27
TransportNetworks:TN.AirTransportNetwork.RunwayCentrelinePoint	1536	21
TransportNetworks:TN.AirTransportNetwork.StandardInstrumentArrival	768	27
TransportNetworks:TN.AirTransportNetwork.StandardInstrumentDeparture	744	26
TransportNetworks:TN.RailTransportNetwork.RailwayLink	3635	31
TransportNetworks:TN.RailTransportNetwork.RailwayNode	3328	30
TransportNetworks:TN.RoadTransportNetwork.RoadLink	2829	60
TransportNetworks:TN.WaterTransportNetwork.Beacon	896	30
TransportNetworks:TN.WaterTransportNetwork.FairwayArea	5859	42
TransportNetworks:TN.WaterTransportNetwork.PortNode	2348	20
TransportNetworks:TN.WaterTransportNetwork.TrafficSeparationSchemeCrossing	2544	20
TransportNetworks:TN.WaterTransportNetwork.TrafficSeparationSchemeLane	2920	21
TransportNetworks:TN.WaterTransportNetwork.TrafficSeparationSchemeRoundabout	2822	20
TransportNetworks:TN.WaterTransportNetwork.TrafficSeparationSchemeSeparator	5329	26
TransportNetworks:TN.WaterTransportNetwork.WaterwayLink	3743	27

TransportNetworks:TN.WaterTransportNetwork.WaterwayNode	4044	40
<b>Total for INSPIRE Transport Networks datasets</b>	<b>358 806</b>	<b>1443</b>
<b>Total for other national transport networks datasets</b>	<b>48 612 756</b>	<b>298250</b>

In essence, the bottom-line is that the inclusion of inland waterway infrastructure and transport networks datasets as HVDs will not have a drastic impact on the current landscape for their provision in the EU. We have indeed seen above that they are already largely provided in line with the recommended measures for publication – which is also the reason why costs related to these recommendations are considered as rather low. The main expected benefit lies rather in the fact that widely acknowledged best practices will be consolidated and set into stone as part of the robust legal framework provided by the PSI/Open Data Directive. In turn, this can only multiply and strengthen the already existing re-use of these datasets, thereby having a multiplier effect on the “socio-economic” benefits summarised above.

### 3.6.3 Recommended policy options

The assessment of the costs and benefits presented in the previous sections suggest that considering inland waterway and river infrastructure data (including static data, dynamic/urgent data and ENC)s and transport networks data (including features for road, rail, waterborne, air and cable transport as per INSPIRE) as HVDs under the PSI Directive, and providing these as recommended in section 3.6.2.1, can be considered as having a positive impact on the state of the EU data economy.

Overall, such policy development would contribute to reaping a number of economic, environmental and social benefits throughout the EU, while imposing a bearable amount of costs on Member States. In particular, regarding inland waterway data, all relevant data points are already provided by Member States through the effects of the RIS Directive. In addition, investments required for the alignment to the recommended measures for publication would mainly be borne through the EU funded CEF project, having therefore minimal incidence on Member States individually.

As regards transport networks data, costs would be concentrated on those Member States which do not (yet) provide these datasets via APIs. In addition, depending on the application of the HVD provision onto the ‘other’ transport networks datasets discussed above, should these be currently provided in other portals than the INSPIRE ones, then the infrastructural costs would be likely to increase proportionally.

In these regards, in order to limit the burden on Member States, and thereby facilitate their transition, several options can be envisaged:

- Launch the application of the recommended provisions for the “mobility” thematic area with those related to **inland waterway and river infrastructure data**, as it was demonstrated before as being a so-called “low-hanging fruit” for both data holders and reusers, having a wide range of benefits for the wider society;
- Then move on to the application of the suggested provisions for **transport networks datasets, starting with the INSPIRE data(only) sets** in order to restrict the initial efforts for Member States, who would in essence “only” have to cater for infrastructural investments related to the APIs (when not already available), as these datasets are already provided in compliance with INSPIRE data specifications which have been leveraged as part of the recommended measures for publication;
- In a third phase, **include all ‘other’ national transport networks datasets**, in light of their value for reusers and capacity to boost the expected benefits. Here again, in order to reduce the burden notably related to any data transformation (these datasets may not necessarily follow the INSPIRE

specifications, for instance) and infrastructural investments, it could be considered to adopt a phased approach, tackling e.g. certain transport modes at a time. For instance,

- The process could start with land transport (road and rail) as having seemingly more use cases that can be considered as “quick wins” and with a wide(r) number of beneficiaries than other transport modes;
- Then waterborne transport, as it was seen that maritime and inland waterway transport are key modes as part of the transport and logistics value-chain, and enablers for efficient multimodal mobility (in addition, parts of these data will also be made available through the provisions proposed for inland waterway data);
- And finally, aviation and cableways transport networks.

The aforementioned phased-approach is a mere suggestion of the study team informed by the research conducted in light of findings presented in this report. The ultimate decision (and related application planning) of the extent to which recommended measures for publication for both types of datasets in scope of the “mobility” thematic area would be actually adopted, remains in the hands of the Commission.

Last but not least, this section presents the different policy options designed for the mobility thematic area. As the current modalities for publication of the datasets considered in this field are already rather harmonised and aligned with PSI/Open data principles throughout the EU (notably thanks to the INSPIRE and RIS Directives) one single parameter was used to determine policy options of lower and higher intensity: the number or scope of data fields and/or datasets to be included as high-value datasets.

Box 6 – Validation workshop results: mobility, overall appreciation of policy intervention options

During the validation workshop organised on 28 July 2020, participants were requested to indicate whether they agreed or disagreed with the three<sup>490</sup> proposed policy options. The mobility options received the following appreciations (42 Respondents): Agree: 81% and Disagree: 19%.

In addition, the policy options were evaluated by participants as regards their relevance with regards to the overall environment of the thematic area and the respective needs of the participants. The three options obtained the following scores:

- Low: This option obtained the score of 2.8/10 (44 Respondents).
- Medium: This option obtained the score of 4.6/10 (44 Respondents).
- High: This option obtained the score of 7.5/10 (44 Respondents).

### 3.6.3.1 Lower intensity intervention

The lower intensity intervention would consist of including as HVDs only a few datasets in scope of the mobility thematic area. These would be the ‘low-hanging’ fruits, i.e. the datasets already provided widely and freely available across Member States for which alignment with the recommended measures would involve minimal effort by Member States.

---

<sup>490</sup> In the initial version of this Deliverable, three policy intervention options were considered per thematic area. For the final version of this Deliverable, and upon request of the Commission, the initial three policy options were merged into two policy options, a lower and higher intensity options. All elements composing the initial three options were transferred through to the final two options, and as such, the validation of the stakeholders still holds.

In practice, this translates into including on the list of HVDs, those datasets in scope of this study that governed by well-established EU legislation aimed at harmonising and enhancing the sharing and re-use of mobility-related data. The main added-value, as presented in Deliverable 2&3, is the formalisation of an 'Open Data' principle in addition to the existing regulatory *aquis*.

On the one hand, the lower intensity option would involve including **inland waterway and river infrastructure data** presented in the table below (along with their respective reference in the TENtec Open Method of Coordination (OMC) Glossary<sup>491</sup>) as HVDs. Indeed, as presented in Deliverable 2&3 all relevant data points are already provided by Member States through the effects of the RIS Directive. In addition, the EU funded initiative "RIS COMEX", encompassing all 13 relevant European countries (those interconnected by inland waterways across Europe) is already bridging all data provision practices together, in line with the recommended practices. This means that investments required for the alignment to the recommended measures for publication would mainly be borne through the EU funded CEF project, having therefore minimal incidence on Member States individually.

Table 53 – high-value datasets in the mobility thematic area, inland waterways

Data type	Datasets
Static data	<ul style="list-style-type: none"> <li>• Fairway characteristics (02 ILW/01-21 &amp; 33)</li> <li>• Long-time obstructions in the fairway and reliability (02 ILW/22-27)</li> <li>• Rates of waterway infrastructure charges (02 ILW/34)</li> <li>• Other physical limitations on waterways (03 LO/1; 04 LC/1-5 &amp; 8-9; 05 BR/01-06)</li> <li>• Regular lock and bridge operating times (03 LO/2 &amp; 4-9; 04 LC/6-7; 05 BR/6)</li> <li>• Location and characteristics of ports and transshipment sites (PO/3-65; 07 PT/1-2)</li> <li>• List of navigation aids and traffic signs</li> <li>• Navigation rules and recommendations</li> </ul>
<b>Dynamic and urgent data</b>	<ul style="list-style-type: none"> <li>• Water depths contours in the navigation channel (02 ILW/14)</li> <li>• Temporary obstructions in the fairway (02 ILW/22-25)</li> <li>• Present and future water levels at gauges (02 ILW/15-16)</li> <li>• State of the rivers, canals, locks and bridges (02 ILW; 03 LO; 04 LC; 05 BR)</li> <li>• Restrictions caused by flood and ice (02 ILW/24-25; 03 LO/6-7)</li> <li>• Meteorological data (incl. wind direction)Short term changes of lock and bridge operating times (03 LO/2; 05 BR/6)</li> <li>• Short term changes of aids to navigation</li> </ul>
<b>Inland electronic navigational charts (ECDIS)</b>	<ul style="list-style-type: none"> <li>• Waterway axis with kilometres indication (02 ILW/01-07)</li> <li>• Links to the external xml-files with operation times of restricting structures (03 LO/2; 05 BR/6)</li> <li>• Location of ports and transshipment sites (06 PO/58-59; 07 PT/1-2)</li> <li>• Reference data for water level gauges relevant to navigation (02 ILW/15-16)</li> <li>• Bank of waterway at mean water level</li> <li>• Shoreline construction (02 ILW/9-12; 04 LC/1-4; 05 BR/3-6)</li> <li>• Contours of locks and dams (02 ILW/15-16)</li> <li>• Boundaries of the fairway/navigation channel (02 ILW)</li> <li>• Isolated dangers in the fairway/navigation channel under and above water (02 ILW/22-25)</li> <li>• Official aids-to-navigation (e.g. buoys, beacons, lights, notice marks)</li> </ul>

<sup>491</sup> See TENtec OMC Glossary (2017), [https://ec.europa.eu/transport/sites/transport/files/tentec\\_omc\\_glossary.pdf](https://ec.europa.eu/transport/sites/transport/files/tentec_omc_glossary.pdf)

As a reminder, the RIS-COMEX project is a pilot project launched in 2016 in which 13 European countries are working together, under the coordination of the Austrian Waterway Administration (via Donau), towards seamless and sustainable operation of cross-border RIS Services.<sup>492</sup> Harmonisation of operational exchange of underlying RIS data is one of the key objectives of the project, and fairway data is directly tackled through dedicated (sub-)activities under the project<sup>493</sup>. Therefore, the recommended measures for publication for static and dynamic data, as well as for the ENC's presented in the table below have been elaborated in line with the project's requirements, in order to ensure consistency. Most importantly, this means that in essence these recommended measures for publication are already agreed upon by Member States. In other words, considering inland waterway infrastructure data as HVDs with the measures for publication displayed below would require little effort from Member States.

Table 54 – Recommended modalities for publication of inland waterway infrastructure data – Lower intensity intervention

<b>Dimensions</b>		<b>Static</b>	<b>Dynamic/Urgent</b>	<b>Electronic Navigational Charts</b>
<i>Openness-data specification</i>	<i>License (terms of use)</i>	CC-BY 4.0 No terms of use		
	<i>Format</i>	CSV, XML, (geo)JSON	CSV, XML, (geo)JSON	XML, 000, WMS
	<i>Machine-readability</i>	Mandatory		
	<i>Availability of API, bulk download</i>	Bulk download mandatory. Webservices and API recommended.	Web service, (OGC) API and bulk download mandatory.	
<i>Documentation</i>	<i>Metadata (dataset content description)</i>	Complete (*.csv document available)		
	<i>Data linking</i>	N/A		
	<i>Documentation (incl. structure and semantics)</i>	RIS Index Encoding Guide 3.0 developed by the Joint Task Force on the RIS Index	Commission Implementing Regulation (EU) 2018/2032 of 20 November 2018 for NtS	Commission Implementing Regulation (EU) 2018/1973 of 7 December 2018 for Inland ECDIS
	<i>Shared vocabularies</i>	RIS Index Encoding Guide 3.0	NtS Standard 4.0	Inland ECDIS Standard 2.5
	<i>Taxonomies</i>	N/A		
<i>Completeness</i>	<i>Traceability</i>	N/A		
	<i>Update frequency and timeliness</i>	When necessary	Daily to (near) real time	Monthly (for shallow sections)
	<i>Granularity</i>	Individual waterway km level		National waterway network and cross-

<sup>492</sup> See: <https://www.riscomex.eu/>

<sup>493</sup> Under Activity 2: Corridor RIS Preparation, Sub-Activity 2.2 Level 1 Case definition aims at defining the requirements and services enabling optimal route planning based on reliable and complete fairway and infrastructure data. Under Activity 3: Corridor RIS Implementation, Sub-Activity 3.2 Level 1 Services aims to further specify and implement the services as defined in Sub-Activity 2.2. More information, see: <https://www.riscomex.eu/activities/>

Dimensions		Static	Dynamic/Urgent	Electronic Navigational Charts
				border nodes
	<i>Key attributes</i>	ISRS Location Code		

As the table suggests, the rather similar recommended modes of provision apply to the three categories of datasets in scope, with some variances data formats, documentation, shared vocabularies and frequency of updates. The justifications for each of these recommended measures are the following:

- Concerning **licenses**, the lower intensity option suggests to adopt open licenses allowing for any type of reuse (including of course commercial reuse and transmission). It appeared from the data collection regarding the 'as is' modalities for provision that in many countries licenses were currently not specified, therefore it can be expected that a formal transition to such open licenses should not raise any major concerns.
- Format-wise**, several options – already in use in the majority of Member States, have been discussed with stakeholders. Similarly to the transport networks datasets (see hereafter), it appears that the availability of multiple relevant formats does not hinder the potential for reuse of the datasets, which is the reason why multiple options are kept as recommended, bottom-line being that all of these formats are **machine-readable** and therefore in line with the PSI/Open Data Directive principles.
- This lower intensity option suggest to ensure the accessibility of the data via **webservices, APIs and bulk downloads** (except for static data where bulk download would suffice due to the nature of the – static – information prone to very few changes). The main interest here lies in having seamless access to (reuse) the information through mapping applications for routing/voyage planning while also being able to download all datasets at once in order to freely develop any other applications/use-cases. It should still be noted that web-services indeed allow to visualise the datasets but not necessarily to entirely reuse them, as such APIs make the data more reusable in the sense of the Open Data Directive.
- For APIs to be taken up however by all sort of reusers, **clear documentation** including structure of the datasets and **shared vocabularies** become particularly important. To this extent the application of the RIS Index Encoding Guides and RIS Implementing legislation and standards for NtS and ECDIS play a key harmonising role. Similarly, this low intensity option requires that complete and easily accessible **metadata** is provided in order to facilitate reuse of these datasets.
- Concerning the **data linking, taxonomy and traceability of data**, reusers provided no feedback and therefore no specific recommended measures could be developed on these aspects.
- The **timeliness or frequency of update** of the datasets is considered as key for the reusability of dynamic and urgent data, which again is directly linked to their nature. Here datasets should be provided in (near) real-time in order to guarantee the accuracy of RIS services developed on the basis thereof. On the other hand, static and Inland ECDIS require fewer updates due to their more static nature, with at least monthly updates to ENCs (especially in shallow areas) and a needs-basis for the static infrastructure data.
- The **granularity of the datasets** is also a key aspect, with datasets provided preferably at the level of each individual waterway kilometre point for static and dynamic data, and ENCs presenting the entire national waterway network including the cross-border links and nodes. These are motivated by both safety/security aspects for the deepest level of detail as possible, as well as the multi-modal and cross-border transport ambitions known to the inland waterway sector.

- Last but not least, the ISRS Location Code is the **key attribute** to be referenced for all datasets, as these allow to disambiguate and uniquely identify all datasets. Similarly to other standards referred to in this section, the ISRS Location Codes are already widely reused in the current modes of provision of inland waterway infrastructure data by Member States.

On the other hand, the lower intensity intervention would involve including also **transport network datasets** as per the **INSPIRE datasets** Directive as HVDs. As explained in Deliverable 2&3, the Transport Networks data theme is defined within the INSPIRE Directive as: "*Road, rail, air and water transport networks and related infrastructure. Includes links between different networks. Also includes the trans-European transport network*" <sup>494</sup>, and comprises the six following sub-themes: common elements (or multi-modal datasets relevant to all modes of transport) road transport, rail transport, water transport, air transport and cableways. For each sub-theme's network schemes, the table on the next page provides a list of specific features to be included as HVDs – this list is aligned with the INSPIRE specifications for Transport Networks.

---

<sup>494</sup> OJ L 228, 9.9.1996, p. 1. Decision as last amended by Council Regulation (EC) No 1791/2006 (OJ L 363, 20.12.2006, p. 1



Table 55 – high-value datasets in the mobility thematic area, INSPIRE transport network features

Common	Road	Rail	Water		Air		Cableways
<ul style="list-style-type: none"> <li>• Access Restriction</li> <li>• Condition Of Facility</li> <li>• Maintenance Authority</li> <li>• Marker Post</li> <li>• Owner Authority</li> <li>• Restriction For Vehicles</li> <li>• Traffic Flow Direction</li> <li>• Transport Area</li> <li>• Transport Link</li> <li>• Transport Link Sequence</li> <li>• Transport Link Set</li> <li>• Transport Network</li> <li>• Transport Node</li> <li>• Transport Object</li> <li>• Transport Point</li> <li>• Transport Property</li> <li>• Vertical Position</li> </ul>	<ul style="list-style-type: none"> <li>• E-Road</li> <li>• Form Of Way</li> <li>• Functional Road Class</li> <li>• Number Of Lanes</li> <li>• Road</li> <li>• Road Area</li> <li>• Road Link</li> <li>• Road Link Sequence</li> <li>• Road Name</li> <li>• Road Node</li> <li>• Road Service Area</li> <li>• Road Service Type</li> <li>• Road Surface Category</li> <li>• Road Width</li> <li>• Speed Limit</li> <li>• Vehicle Traffic Area</li> </ul>	<ul style="list-style-type: none"> <li>• Design Speed</li> <li>• Nominal Track Gauge</li> <li>• Number Of Tracks</li> <li>• Railway Area</li> <li>• Railway</li> <li>• Electrification</li> <li>• Railway Line</li> <li>• Railway Link</li> <li>• Railway Link Sequence</li> <li>• Railway Node</li> <li>• Railway Station Area</li> <li>• Railway Station Code</li> <li>• Railway Station Node</li> <li>• Railway Type</li> <li>• Railway Use</li> <li>• Railway Yard Area</li> <li>• Railway Yard Node</li> </ul>	<ul style="list-style-type: none"> <li>• Beacon</li> <li>• Buoy</li> <li>• CEMT Class</li> <li>• Condition Of Water Facility</li> <li>• Fairway Area</li> <li>• Ferry Crossing</li> <li>• Ferry Use</li> <li>• Inland Waterway</li> <li>• Marine Waterway</li> <li>• Port Area</li> <li>• Port Node</li> <li>• Restriction For Water Vehicles</li> <li>• Traffic Separation Scheme</li> <li>• Traffic Separation Scheme Area</li> <li>• Traffic Separation Scheme Crossing</li> <li>• Traffic Separation Scheme Lane</li> </ul>	<ul style="list-style-type: none"> <li>• Traffic Separation Scheme Roundabout</li> <li>• Traffic Separation Scheme Separator</li> <li>• Water Link Sequence</li> <li>• Water Node</li> <li>• Water Traffic Flow Direction</li> <li>• Waterway</li> <li>• Waterway Link</li> <li>• Waterway Node</li> </ul>	<ul style="list-style-type: none"> <li>• Aerodrome Area</li> <li>• Aerodrome Category</li> <li>• Aerodrome Node</li> <li>• Aerodrome Type</li> <li>• Air Link</li> <li>• Air Link Sequence</li> <li>• Air Node</li> <li>• Air Route</li> <li>• Air Route Link</li> <li>• Airspace Area</li> <li>• Apron Area</li> <li>• Condition Of Air Facility</li> <li>• Designated Point</li> <li>• Element Length</li> <li>• Element Width</li> <li>• Field Elevation</li> <li>• Instrument Approach Procedure</li> <li>• Lower Altitude Limit</li> <li>• Navaid</li> <li>• Procedure Link</li> </ul>	<ul style="list-style-type: none"> <li>• Runway Area</li> <li>• Runway Centreline Point</li> <li>• Standard Instrument Arrival</li> <li>• Standard Instrument Departure</li> <li>• Surface Composition</li> <li>• Taxiway Area</li> <li>• Touch Down Lift Off Area</li> <li>• Upper Altitude Limit</li> <li>• Use Restriction</li> </ul>	<ul style="list-style-type: none"> <li>• Cableway Link</li> <li>• Cableway Link Sequence</li> <li>• Cableway Link Set</li> <li>• Cableway Node</li> </ul>

Starting with the INSPIRE datasets would allow to restrict the initial efforts for Member States, who would in essence “only” have to cater for infrastructural investments related to the APIs (when not already available), as these datasets are already provided in compliance with INSPIRE data specifications which have been leveraged as part of the recommended measures for publication.

The recommended measures for publication concerning transport network datasets build extensively on the good practices and approaches developed for the INSPIRE Directive. The table below summarises the recommended measures suggested as part of the lower intensity intervention for all the publication dimensions and for all transport modes (road, rail, water, air and cableways). As the table clearly shows, there is a limited need to make distinctions between the different transport modes in terms of dimensions of provision. This is because the datasets share the same key characteristics in terms of type of data (static data only) and due to the fact that reusers’ requests are globally aligned across the modes.

Table 56 – Recommended modalities for publication of INSPIRE transport networks data – Lower intensity intervention

Dimensions		Road transport	Rail transport	Water transport	Air transport	Cableways
Openness-data specification	License (terms of use)	CC-BY 4.0 No terms of use				
	Format	GML, GeoPackage, GeoJSON				
	Machine-readability	Mandatory				
	Availability of API, bulk download	Web service, (OGC) API and bulk download				
Documentation	Metadata (dataset content description)	Complete (*.csv document available)				
	Data linking	Links to national INSPIRE Geoportals and datasets as relevant.				
	Documentation (incl. structure and semantics)	Complete and available				
	Shared vocabularies	INSPIRE data specifications are recommended but not mandatory.				
	Taxonomies	N/A				
Completeness	Traceability	N/A				
	Update frequency and timeliness	When necessary				
	Granularity	From local to national, including links with cross-border networks, if and where available				
	Key attributes	Any national identification code; latitude and longitude				

Further explanations on the recommended modes of provision described in the table are provided below:

- Concerning **licenses**, the lower intensity option suggests to adopt open licenses allowing for any type of reuse (including of course commercial reuse and transmission). Many countries already rely upon such open licenses and this therefore constitutes a good practice which is quite widespread. At the

same time, terms of use are rather frequent, although they are not considered as acceptable by reusers. In fact, as per other thematic areas, terms of use impose restrictions (of various extents) on the possibility to use the datasets and therefore are considered by stakeholders against the “spirit” of the PSI/Open Data Directive.

- **Format-wise**, the lower intensity intervention suggests to use several possible options, listed above, which are all already in use and are all valuable from a reuser perspective. Differently from other thematic areas, there was no indications of one or two formats in particular which would constitute best practices and should be adopted by all countries. In this domain, the availability of multiple relevant formats does not seem to hamper the reusability of datasets and, for this reason, the list of possible formats is voluntarily left quite open. **The machine readability** of the datasets’ format however remains the main condition for considering the format adequate under the PSI/Open Data Directive.
- This intervention would allow the possibility of accessing information **both through APIs and via bulk download**. The reason for this request lies in the different use cases that APIs and bulk download would support (i.e. map applications would rely on APIs while logistic analysis would rely on bulk download) and the greater freedom that both options would entail for reusers developing their own services. In this respect, it must be noted here that many countries today provide web-services allowing to visualise the data rather than APIs. While this is changing fast and more countries are moving/have already moved towards APIs (i.e. France, Sweden, Ireland...) and towards making these datasets more accessible for developed and non-expert users<sup>495</sup>, the reliance on web-services might be an obstacle for reusers. Most transport network web services in fact focus on visualisation and access to data rather than reuse. In this context, the development of APIs is considered as a key success factor from an Open Data perspective (see section 3.8). For APIs to be taken up however by all sort of reusers, **clear documentation** including structure of the datasets and semantics becomes particularly important and should be provided as well.

Nonetheless, data holders expressed their concerns on the deletion/prohibition of all terms of use/provision of unlimited access to datasets and especially from the perspective of the stability of the IT infrastructure: a recent example from Austria in fact demonstrated that, in the absence of clear rules for accessing APIs for instance, the IT system for the provision of the data might be overloaded and this might provoke the temporary failure of the service for all reusers<sup>496</sup>. Therefore, this medium intensity intervention could introduce technical requirements in order to regulate the access to the datasets by for example limiting the number of daily calls per stakeholder. This would allow to protect the datasets, which could not be queried indefinitely through the API.

- **Metadata** should be provided in a complete and accessible manner. Fortunately, thanks to the INSPIRE Directive, metadata are already very widely provided (see the INSPIRE Geoportal for the full picture). Complete and easily accessible metadata should continue to be provided in order to facilitate reuse of these datasets.
- The aspect of **data linking** was only superficially discussed with stakeholders and solely from the perspective of facilitating reuse by clearly establishing/providing unique entry points/platforms for all datasets. While the establishment of one stop shops/centralised platforms for the provision of these datasets goes well beyond the scope of the PSI/Open Data Directive, stakeholders clearly argued that the fragmentation in the data provision (across multiple portals at the national and European level) is a major obstacle for its reuse and that one stop shop national portals should be established (or the ITS

---

<sup>495</sup> See for instance Barbero, M., Lopez Potes, M., Vancauwenberghe, G. and Vandenbroucke, D., The role of Spatial Data Infrastructures in the Digital Government Transformation of Public Administrations , Publications Office of the European Union, Luxembourg, 2019, <https://ec.europa.eu/jrc/en/publication/role-spatial-data-infrastructures-digital-government-transformation-public-administrations>, p. 53

<sup>496</sup> As mentioned by one stakeholders during the online focus group, recently the Austria geoportal suffered from a temporary failure due to a user programming 1000 calls per seconds towards the API platform, which could not cope with such an overload, although limited in time.

Directive National access portal should be provided with all these datasets)<sup>497</sup>. For this reason, this medium intensity intervention recommends to consider data linking to at least the national INSPIRE Geoportal as very important if not essential.

- In terms of **shared vocabularies**, this intervention suggests to follow the INSPIRE Directive, which provides sufficient clarity and harmonisation on the transport network theme. These vocabularies should be considered as a reference (although not imposed).
- Concerning the **taxonomy and traceability of data**, reusers provided limited/no feedback and therefore no specific recommended measures could be developed on these aspects.
- The **timeliness or frequency of update** of the datasets is considered of great importance by stakeholders, despite the datasets in scope are mainly static. In this context, this intervention recommends that the datasets are updated when necessary. This implies that some different modes might be updated more frequently than others, but this will be a decision to be made by each data holder.
- The **granularity of the datasets** is also considered as key by reusers who strongly advocate for as granular transport network data as possible, in terms of the scope coverage. While many countries focus on national transport networks and on the national dimension mainly, the local dimension acquires more and more importance from a multi-modal transport perspective and when looking at the uptake of automated and connected vehicles. The desired and recommended granularity hence covers datasets going from local to national and including links with cross-border networks. Nonetheless, this might be a challenge for data holders and might not be possible in all countries, as very granular data for the local level in particular are not always available and the Directive cannot impose the collection of new data. As reusers were adamant in asking, whenever and wherever available, the lowest level of granularity possible, it is recommended for the PSI/Open Data Directive to cover from the local to the national level as to include local datasets for those countries holding them and already making them available.
- Finally, some **key attributes** which are essential for the transport network datasets are elements ensuring their unique identification and disambiguation such as any national identification code or name, as well as the coordinates of the transport network feature.

### 3.6.3.2 Higher intensity intervention

The higher intensity intervention would build on the lower intensity intervention as presented above. This implies that in this intervention, the datasets concerning inland waterway infrastructure data and INSPIRE transport network would be in scope, and, that in addition to these two datasets, the higher intensity intervention would include *all* national transport network datasets i.e. not only the INSPIRE ones presented in the table above, in light of their value for reusers and capacity to boost the benefits their reuse could entail. Indeed, it appears most Member States possess and publish many more datasets relating to transport networks aside from the INSPIRE-tagged ones, which could equally gain value from being considered as HVDs, and thus being provided with open licences, free of charge, in harmonised machine readable formats through APIs and bulk download as per the recommended modalities for transport networks data.

The recommended measures for publication of these additional, transport networks datasets held at national level are based on the HVD principles, i.e. they will need to be available free of charge, in machine readable format, and through bulk-download and APIs. All other publication modalities will be recommended only, i.e. not mandatory in line with the recommendations of the lower intensity option for transport networks under INSPIRE – these are typed in grey in the table below.

---

<sup>497</sup> Online focus group

Table 57 – Recommended modalities for publication of other national transport networks data – High intensity intervention

Dimensions		Road transport	Rail transport	Water transport	Air transport	Cableways
Openness-data specification	<i>License (terms of use)</i>	CC-BY 4.0 No terms of use				
	<i>Format</i>	Formats as recommended in PSI HVD measures for transport networks under INSPIRE (see lower intensity option) but not mandatory.				
	<i>Machine-readability</i>	Mandatory				
	<i>Availability of API, bulk download</i>	Webservice, (OGC) API and bulk download				
Documentation	<i>Metadata (dataset content description)</i>	Metadata as recommended in PSI HVD measures for transport networks under INSPIRE (see lower intensity option) but not mandatory.				
	<i>Data linking</i>	Links to national INSPIRE Geoportals and datasets as recommended in PSI HVD measures for transport networks under INSPIRE (see lower intensity option) but not mandatory.				
	<i>Documentation (incl. structure and semantics)</i>	Documentation as recommended in PSI HVD measures for transport networks under INSPIRE (see lower intensity option) but not mandatory.				
	<i>Shared vocabularies</i>	Shared vocabularies as recommended in PSI HVD measures for transport networks under INSPIRE (see lower intensity option) but not mandatory.				
	<i>Taxonomies</i>	N/A				
Completeness	<i>Traceability</i>	N/A				
	<i>Update frequency and timeliness</i>	When necessary				
	<i>Granularity</i>	Granularity as recommended in PSI HVD measures for transport networks under INSPIRE (see lower intensity option) but not mandatory.				
	<i>Key attributes</i>	Key attributes as recommended in PSI HVD measures for transport networks under INSPIRE (see lower intensity option) but not mandatory.				

In order to reduce the burden on Member States related to the inclusion of these additional national datasets as HVDs (e.g. infrastructural investments), it could be considered to adopt a phased approach, tackling e.g. certain transport modes at a time. For instance:

- The process could start with land transport (road and rail) as having seemingly more use cases that can be considered as “quick wins” and with a wide(r) number of beneficiaries than other transport modes;
- Then waterborne transport, as it was seen that maritime and inland waterway transport are key modes as part of the transport and logistics value-chain, and enablers for efficient multimodal mobility (in addition, parts of these data will also be made available through the provisions proposed for inland waterway data);
- And finally, aviation and cableways transport networks.

Should this option be adopted by DG CNECT, the feasibility and possible timing of such approach should be determined in collaboration with DG MOVE and relevant stakeholders at a subsequent stage and is out of scope of the current impact assessment study.

### 3.7 Horizontal considerations for the micro-level analysis

This section provides additional considerations regarding APIs and the impacts of open data on small- and medium-sized enterprises (SMEs), which are valid across all thematic areas.

#### 3.7.1 APIs

This section aims to provide some horizontal considerations related to the use of APIs for the provision of HVDs. As initially agreed during the scope definition of this study (i.e. during the information webinar held on 12 February 2020), the provision of API standards falls outside the realm of the study. Besides the issue related to scope of the study, it should be noted that not enough data has been collected in order to provide a detailed analysis and clear recommendations in this regard. Therefore, the aim of this section is only to provide some general and horizontal considerations on this topic.

Based on the data collected during the interviews, it was found that the use of APIs for the provision of data in the thematic areas is not widespread across Member States. Only a few of them have incurred costs, and made the necessary investments in order to put APIs in place. In some cases, not having an API is a conscious choice in order to avoid the reuse of data in some countries. In others, bulk download is possible in a number of countries, where APIs are not provided on top (and vice versa).

When it was found that these APIs were in place, the data holders highlighted the technical challenges faced, and the measures adopted in order to cope with such issues. Clear rules for accessing APIs are essential in order to ensure their correct functioning. Without these rules, the IT system providing the data might be overloaded and this might provoke the temporary failure of the service for all reusers. In this sense, Member States have decided to define and implement such rules. In some Member States for example, the number of calls is limited per day, implying that each reuser can only request data a limited number of times. In others, the biggest reusers are allocated different timeslots in order to access the dataset, and thus avoid overloading and crash the system.

Besides the technical issues that might be faced, the infrastructural investments necessary to put in place and maintain APIs should also be highlighted. The costs would vary from one country to another, and would also depend on the volume of data to be handled. Recent studies indicate that costs of APIs establishment range in between 30 000 euro and 2.5 million euro depending on the type of infrastructure and technical characteristics<sup>498</sup> and that in average an API set up costs 50 000 euro<sup>499</sup>. Member States would thus need to face considerable high costs. However, the provision of these datasets under the PSI Directive would require the use of APIs. Moreover, the provision of these dataset though bulk download should also be made possible, in order to allow reusers to download and carry out analysis. The combination of APIs and bulk download should then be the norm, and would be key in order to unleash all the potential benefits of reusing these datasets.

---

<sup>498</sup> Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, <https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en>

<sup>499</sup> Study to support the review of Directive 2003/98/EC on the re-use of public sector information, 2018, page 409, <https://op.europa.eu/en/publication-detail/-/publication/45328d2e-4834-11e8-be1d-01aa75ed71a1/language-en>

The JRC is currently conducting a study further assessing the technical considerations of APIs necessary for the provision of HVDs. This study would help to shed some light on the matter.

### **3.7.2 Impact of open data on SMEs**

Another consideration to bear in mind concerning open data, and specifically HVDs, is the impact on the European market, particularly SMEs. As pointed out by the European Commission services, there is a fear that HVDs might reinforce the dominant position of (US) tech giants (mainly GAFA), undermining European SMEs.

Big tech have great capacity and skills to work with (open) data. They access data, analyse it, draw conclusions, and get value out of it, growing their business and presence world-wide. If data is not provided openly, big tech have the capacity to acquire it easily. On the other hand, small organisations, including SMEs, do suffer from the lack of open data. As their resources are limited, these small organisations have to decide whether to invest to access the data in order to reuse it.

Therefore, in order to create a level playing field, data should be provided openly. While this will create great benefits for big tech, it will also unlock a huge value for small organisations, incl. SMEs, as it will give them the opportunity to develop their businesses, innovate, and grow.

# 4 Macro-level Impacts of Policy Options for High Value Datasets

## 4.1 Comparison of policy options

This section offers a comparison of the policy options proposed – a lower and a higher intensity intervention – for each thematic area by means of a Multi-Criteria Analysis.

### 4.1.1 Methodological framework for the comparison of policy options

The comparison of the policy options against the assessment criteria is based on a **Multi Criteria Analysis** (MCA). The MCA is a largely qualitative analysis of the policy options, based on ratings and rankings with quantitative data supporting the assessment. The MCA has been performed in line with the European Commission's *Better Regulation Guidelines*<sup>500</sup> and its toolbox<sup>501</sup>, most importantly tool 63<sup>502</sup>. The assessment builds on the prior analysis of each individual option.

The MCA was used in this study for two reasons: First, it is an alternative to the Cost-Benefit Analyses (CBA) performed in the previous chapters.<sup>503</sup> Secondly, it is particularly relevant at the following stages of our impact modelling: (1) at the stage of assessing the economic, social, and environmental dimensions of each of the two intervention options, including possible trade-offs between these dimensions, and (2) at the stage of comparing the policy options against the main criteria effectiveness, efficiency and coherence as well as the criteria proportionality and (legal and political) feasibility.<sup>504</sup>

It has been concluded in the previous chapter that for none of the six areas under investigation, the baseline will be able to achieve the desired results and resolve the identified problems. Hence, a policy intervention is concluded to be necessary. Therefore, a lower intensity and a higher intensity intervention option have been proposed per thematic area. The aim of the MCA is, consequently, to assess which of the policy options in each thematic area is the most adequate.

Multi-Criteria-Analyses are a widely-used tool in policy evaluations, impact assessment studies, and feasibility studies that aim at drawing a conclusion on the comparative rating of potential policy solutions.

Deloitte has carried out such analysis in previous studies on behalf of various Directorate-Generals of the European Commission. As part of this Study, Deloitte has carried out the MCA in full detail, as exemplified in the Better Regulation Guidelines in order to take full account of the complexity of the subject matter and the level of granularity of the previous analyses

<sup>500</sup> [http://ec.europa.eu/smart-regulation/guidelines/toc\\_guide\\_en.htm](http://ec.europa.eu/smart-regulation/guidelines/toc_guide_en.htm)

<sup>501</sup> [http://ec.europa.eu/smart-regulation/guidelines/toc\\_tool\\_en.htm](http://ec.europa.eu/smart-regulation/guidelines/toc_tool_en.htm)

<sup>502</sup> [https://ec.europa.eu/info/sites/info/files/file\\_import/better-regulation-toolbox-63\\_en\\_0.pdf](https://ec.europa.eu/info/sites/info/files/file_import/better-regulation-toolbox-63_en_0.pdf)

<sup>503</sup> A cost/benefit-analysis has been conducted in the previous chapters for each of the six thematic areas. Due to a lack of data, costs and benefits have been quantified only to the extent possible. Costs and benefits have been assessed in a qualitative manner whenever quantitative data is lacking.

<sup>504</sup> [http://ec.europa.eu/smart-regulation/guidelines/toc\\_guide\\_en.htm](http://ec.europa.eu/smart-regulation/guidelines/toc_guide_en.htm)



carried out.

The MCA was carried out in the following three steps in line with the Better Regulation Guidelines:<sup>505</sup>

- **Step 1:** Establish indicators or assessment criteria against which the baseline scenario and Policy Options are assessed and compared. This includes establishing the performance of a Policy Option (i.e. the magnitude of its impact), the weight of the criteria in relation to each other, as well as the direction of the impact (negative / positive);
- **Step 2:** Build an outranking matrix in which the scores for the baseline scenario and the Policy Options and criteria are provided in order to summarise how they compare towards each other in relation to the established criteria; and
- **Step 3:** Prepare a so-called permutation matrix that enables the selection of a final ranking of all the possible Policy Options towards each other. This means that not only a preferred Policy Option can be selected but also a ranking of all other options towards each other is possible.

#### 4.1.1.1 Step 1: Assessment criteria and rating

For the MCA, the main criteria Effectiveness, Efficiency and Coherence with other Policy options and additionally the criteria proportionality and (legal and political) feasibility were considered to be relevant to rate the two policy options (lower intensity and higher intensity intervention).

According to the Better Regulation Guidelines, **effectiveness** as a central criterion in an MCA looks at the extent to which different options would achieve the objectives of the intervention, looking for evidence of why, whether and how these changes are linked to the intervention.

**Efficiency** is defined as the costs and benefits of a policy option as they accrue to different stakeholders. In the context of the six thematic areas of the PSI Directive, this criterion is also of particular relevance as costs could only be obtained to a certain extent through the research so far.

The **coherence** of each policy option focuses on how well the respective option is interlinked with overarching objectives of EU policies and other EU policy measures (external dimension) and also with the PSI Directive itself (internal dimension).

EU action should be relevant and necessary and respect for principles of subsidiarity and **proportionality** – They should offer value beyond what Member State action alone can deliver and not go further than is necessary to resolve the problem or meet the policy objective. Respect for the principle of proportionality is about ensuring that the policy approach and its intensity match the identified problem/objective.

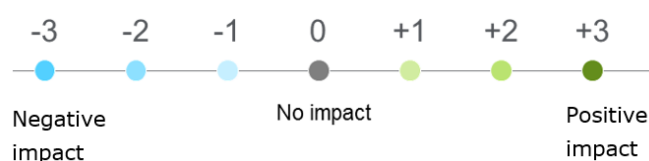
With regard to (legal and political) **feasibility** the options should respect the principle of conferral (legal feasibility). They should also respect any obligation arising from the EU Treaties (and relevant international agreements) and ensure respect of fundamental rights. Legal obligations incorporated in existing primary or secondary EU legislation may also rule out certain options. Options that would clearly fail to garner the necessary political support for legislative adoption and/or implementation could also be discarded (political feasibility).

In our input matrix, we have given a scale for the performance value from -3 to +3, hence, the scale reaches from a very strong negative impact (-3) over no impact at all (0) to a very strong positive impact (+3).

---

<sup>505</sup> [https://ec.europa.eu/info/sites/info/files/better-regulation-toolbox\\_1.pdf](https://ec.europa.eu/info/sites/info/files/better-regulation-toolbox_1.pdf)

Figure 3 – Scale for the Rating of the Criteria of the MCA



Source: Deloitte

Having defined the scale of the performance values the input matrix can be deduced, as indicated in the table below:

**Table 58 – Input Matrix: Weight, direction and performance value of the assessment criteria**

Assessment criterion	Weight	Direction	Performance value
Effectiveness	0.30	1	Qualitative: -3 / +3
Efficiency	0.20	1	Qualitative: -3 / +3
Coherence	0.10	1	Qualitative: -3 / +3
Proportionality	0.10	1	Qualitative: -3 / +3
Feasibility	0.30	1	Qualitative: -3 / +3

Source: Deloitte

Applying the performance scores from -3 to +3 to the five assessment criteria, this means the following:

For effectiveness, a score of 0 means that the intervention option is as effective as the baseline scenario, i.e. has no impact at all. A score of -3 indicates that the intervention option is highly ineffective as compared to the baseline scenario i.e. that with the intervention option the respective aim cannot be achieved. A score of -2 means a slight improvement in effectiveness, but is still ineffective overall. Asserting a score of -1 indicates that the aims of the respective intervention option can be reached still better with the baseline scenario. Positive values of effectiveness indicate that the respective intervention option is more effective than the baseline scenario. With a maximum score of +3, the objectives of the respective intervention option can fully be achieved without any compromises. With a score of +2, the intervention option is still highly effective, but its objectives cannot be reached completely. A score of +1 signifies that the respective intervention option has a slightly higher effectiveness than the baseline scenario.

Regarding efficiency, a score of 0 implies that the respective intervention option has the same efficiency as the baseline scenario. The positive and negative values, however, do not indicate the specific cost-benefit-ratio but rather the level of deterioration (negative scoring values) or improvement (positive scoring values) in the overall efficiency of the respective option. Negative scorings mean that the costs outweigh the benefits, from only a slight outweigh (-1) to a substantial gap between costs and benefits (-3). For positive values, the opposite holds, i.e. for +3 the benefits-cost-ratio is expected to be very high and for +1 there is only a modest improvement of the benefit-cost-ratio as compared to the baseline.

The coherence of the intervention options with the overall EU policy and internally can be measured in terms of very well defined and adequate (high positive values) until not given at all and even imposing strong negative effects to the overall EU policy framework (high negative values). The scores in between

indicate slight changes towards the score 0 which signifies that the intervention option at stake is as coherent with the EU policy and as internally coherent as the baseline scenario.

Respect for the principle of proportionality is about ensuring that the policy approach and its intensity match the identified problem/objective. The content and form of Union action must not go beyond what is necessary to meet the objectives of the Treaties. The following questions should help in assessing whether a measure adheres to the principle of proportionality:

- Does the initiative go beyond what is necessary to achieve the problem/objective satisfactorily?
- Is the scope of the initiative limited to those aspects that Member States cannot achieve satisfactorily on their own, and where the Union can do better? (boundary test)
- Is the form of Union action (choice of instrument) as simple as possible, and coherent with satisfactory achievement of the objective and effective enforcement?
- Does the initiative create a financial or administrative cost for the Union, national governments, regional or local authorities, economic operators or citizens? If yes, is this cost minimised and commensurate with the objective to be achieved?
- Does the Union action leave as much scope for national decision as possible while achieving satisfactorily the objectives set?
- Is there a solid justification for the choice of instrument - regulation, (framework) directive, or alternative regulatory methods?
- While respecting Union law, are well-established national arrangements and special circumstances applying in individual Member States respected?

With regard to the feasibility criterion especially the legal and the political but also the technical feasibility can be assessed. In terms of legal feasibility options should respect the principle of conferral. They should also respect any obligation arising from the EU Treaties (and relevant international agreements) and ensure respect of fundamental rights. Legal obligations incorporated in existing primary or secondary EU legislation may also rule out certain options. In terms of political feasibility options that would clearly fail to garner the necessary political support for legislative adoption and/or implementation could also be discarded. With regard to technical feasibility technological and technical constraints may not allow for the implementation, monitoring and/or enforcement of theoretical options.

The five criteria can be assigned different weights, thereby altering – as a natural consequence – the results of the policy rankings and leading to a different outcome regarding the different policy packages. There are various options to assign the weights to the criteria:

- First, it is possible to give all five criteria equal weights, i.e. assign all criteria with 20%.
- A second approach would be to value one or more criteria higher than others, giving them more relative weight.

The decision on the weightings themselves is, obviously, a political matter which depends on the preference regarding the relative importance of each criterion. Based on the study findings and discussions so far we have applied different weights for the five criteria depending on their relative importance.

#### ***Performance values of the six thematic areas***

Based on the research, the expert interviews and the comments of the validation workshop, the study teams asserted performance values for the criteria effectiveness, efficiency, coherence proportionality and feasibility for the two policy options. The following tables present the basis of the performance values for each of the policy options in the six thematic areas.

Table 59 – Explanation of the performance values: Company and company ownership

<b>Assessment criteria</b>	<b>PO1 (lower intensity intervention)</b> Reasoning/justification of the score	<b>PO2 (higher intensity intervention)</b> Reasoning/justification of the score
<b>Effectiveness</b>	<p>The lower intensity option the company thematic area includes as HVDs only some of the datasets in scope (i.e. basic information, company documents and accounts, company ownership – non-personal data). Nevertheless, a reduced number of (personal) data points under each dataset are left out. Including these datasets as HVDs in the PSI Directive would ensure and increase their reusability overall. However, excluding personal data from the scope decreases its effectiveness.</p>	<p>The higher intensity option is slightly broader in terms of scope than the lower one. In addition to the datasets in scope in the lower intensity intervention, this option includes the remaining ones: company ownership (personal), and company insolvency status. This option therefore provides a comprehensive scope, ensuring that all datasets in scope under the company thematic area are included as HVDs.</p>
<b>Efficiency (B/C-ratio)</b>	<p>The lower intensity option would bring positive benefits compared to the baseline. The formal inclusion of company data as HVDs would unleash significant benefits due to its reusability. This implies that economic benefits, innovation, social benefits, and environmental and climate change benefits can be expected. On the other hand, this option is expected to generate considerable costs, particularly related to infrastructural investment (i.e. setting up APIs), and the update frequency and timeliness (data should be provided in real time).</p>	<p>The costs under the higher intensity intervention are expected to be higher, as all the datasets under the company thematic area are in scope. This is especially relevant for the company insolvency data, as insolvency registers are very rarely compatible with measures of provision recommended, and would thus need to be adjusted. Nevertheless, the benefits would also be higher. The openness of all datasets in this thematic area would increase the reuse of data. Similarly to the lower intensity option, the higher option would achieve the same type of benefits, although increased as its scope is more comprehensive. Therefore, the benefits of this option are expected to strongly outweigh the costs</p>
<b>Coherence</b>	<p>The lower intensity option is fully coherent with the current legal framework (i.e. the Company Law Directive (2017/1132/EU) ; the Accounting Directive (2013/34/EU); the (Fifth) Anti-Money Laundering Directive (2018/843/EU); the Transparency Directive (2004/109/EC) ; the Central</p>	<p>The higher intensity option is also fully coherent with the legal framework in place. Its coherence is, however, slightly higher than the lower intensity intervention as it includes all the datasets in scope as HVDs. At national level, it should be noted that insolvency company data is often handle by</p>

	Commercial and Companies registers Interconnection System (BRIS) Directive (2012/17/EU); Regulation (EU) 2015/848 on insolvency proceedings ; Directive (EU) 2019/1151 on the use of digital tools and processes in company law). The option includes the data points required to be made public (although not necessarily in an open data format) by such framework.	different entities in charge of company basic information and documents. Therefore, some minor incoherence issues might arise.
<b>Proportionality</b>	The low intensity intervention is fully proportional with its objectives. It includes as HVDs the most necessary datasets under the company thematic area to ensure their reuse and achieve significant objectives in this field.	The higher intensity intervention is considered as slightly less proportional than the lower intensity option as it includes all the datasets in scope in the company thematic area.
<b>Feasibility (legal and political)</b>	The policy option is regarded as feasible to a great extent, as it encompasses the minimum data points which are overall already provided across Member States. Besides this, this option does not include personal data in its scope which is a sensitive point for Member States, due to the diverging approaches and cultures to the topic.	Compared to the lower intensity option, the higher intensity option is considered as politically more sensitive. The scope of this option would be broader, and would include personal data, which is a sensitive issue as explained. Therefore, the higher intensity option is considered as feasible to some extent.

Source: Deloitte compilation

Table 60 – Explanation of the performance values: Geospatial

<b>Assessment criteria</b>	<b>Weight which should be applied acc. to own research + expert interviews</b>	<b>PO1 (lower intensity intervention) Reasoning/justification of the score</b>	<b>PO2 (higher intensity intervention) Reasoning/justification of the score</b>
<b>Effectiveness</b>	The INSPIRE directive has a significant impact on the management of geospatial data. If the resources are available, the transition to open data is almost a natural action as, very	At this policy level, each Member State is required to make the cadastral parcels available in read-only mode. The availability of this data, even if it is read-only (and therefore not strictly machine-readable), still allows an advantage to the land registry itself	This level of policy provides for the total opening of the identified data with a high level of granularity and a qualitative leap towards linked data. Here the main problems lie in the adaptation to some government laws (such as the management of the land

---

often, it is sufficient to apply a license that allows their reuse.

Many of the identified datasets are already available in open data. In other cases, the obstacle to opening is linked either to the lack of resources (= incomplete data) or to the management of some sources that are subject to national laws (e.g. only the land registry can release certified data against payment) or sustainability processes for which failure to sell data could lead to a lack of government revenues.

because it speeds up consultation before making the official request. Therefore it improves situations where the sale of cadastral data is an important source of income for the government.

The difficulty, for some Member States, is to remain within some national restrictions (e.g. policies to access the cadaster).

However, it's more difficult, for this intervention, to have a global coverage for all datasets (especially Addresses and Buildings) and a high level of granularity.

registry or the choice of the cc0 license), the potential lack of income for the datasets that generate economic sustainability with the sale, the level of granularity even more detailed.

---

**Efficiency (B/C-ratio)**

The INSPIRE directive helps a lot in this direction.

The creation of a technological infrastructure capable of providing data in different formats and also via API allows to increase this efficiency.

The identified datasets have a high value and, even if deployed read-

This level of policy can find obstacles in ensuring the level of coverage of the datasets.

The access to the land registry mainly finds obstacles in the change of government laws.

Its opening, even if in read-only mode, can bring a high level of efficiency to the land registry itself.

This level of policy requires a great deal of reorganization of the data collection processes that guarantee global coverage and a very high level of attention to detail and metadata.

The arrival at this level of policy allows a very high efficiency for the whole geospatial infrastructure of a member state.

---

only, can bring economic benefits to the data provider.

---

**Coherence**

The consistency between the policies is presented incrementally.

In some cases, the transition from one to the other can be more or less complicated in relation to the variables of resources, organizational processes, laws and the market generated by the datasets that each member state has to face

The coherence is primarily driven by the INSPIRE directive

This level of policy option may find inconsistencies with the national laws governing the cadastral data.

This level of policy requires an even broader commitment to cadastral data and a more in-depth analysis of some legal issues (e.g. the choice of CC0).

---

**Proportionality**

The policy options are proportional with the objectives identified. The differences in the intensity of the intervention reflect the needs and the current state of opening of the HVDs considered in scope.

The lower intensity option is fully proportional with the policy objectives. The recommendations in terms of licensing, formats, potentially are in line with the current situation affecting the Member States. In some cases, this policy option could be considered really "light".

The higher intensity option is proportional with the policy objectives. All the recommendations with an exception represented by the cadastral parcels data are in line with the objectives identified.

---

**Feasibility (legal and political)**

Legal and political feasibility mainly affects datasets that presents criticalities due to privacy concerns, GDPR issues. The licensing options

This option is ambitious, considering the difficulties deriving from the opening of the cadastral parcel, due in some cases, to the existing national laws. Legal issues are also related to the choice of the licence CC0. It's widely used, but many

could be very difficult to change (in the case of CCO), because of risks related to the

Member States disagree with this “upgrade”, because of the risk of losing the power of traceability, and facing consequential risks for the security.

Source: Deloitte compilation

Table 61 – Explanation of the performance values: Meteorological data

<b>Assessment Criteria</b>	<b>Weight which should be applied acc. to own research + expert interviews</b>	<b>PO1 (lower intensity intervention)</b>	<b>PO2 (higher intensity intervention)</b>
Effectiveness	There is a clear perceived value chain within meteorological data, starting from (historical) observations towards forecasting data. Micro level cost-benefits comparisons have always come out in favour of the benefits (in existing cases found to be between a factor 2 to 70 over costs) The data sets selected follow this value chain	(Validated and historical) observations are the very fundament of meteorological data, no open data effort could ignore them. Alerts are at the core of the public task, and already published by MS as information.  This option would already deliver significant effect, but less than when also adding radar observations and NWP model data.	This option encompasses the data along the full meteorological data value chain, by including radar scans and NWP model data.
Efficiency	Costs predominantly center on the creation of API infrastructure for data provision, and dealing with the	Alerts are already published as information, and the suggested re-usable formats are those already in use in the sector.	Datasets under this option (radar data, NWP model data, and unstructured digitized climate data, come with a heavier demand on infrastructure and for unstructured climate



	data volume that mostly NWP model data has. (Loss of revenue is considered in the political feasibility)	(Validated) observations data and structured historic data are suggested to be published in common and used formats, and do not put a heavy strain on infrastructure.	data with transformation costs. Using the European Meteorological Infrastructure for data provision may increase efficiency, specifically for smaller NWS.
Coherence	The policy options presented are consistent with the Directive, and with current practices and development in the sector.	The policy options presented are consistent with the Directive, and with current practices and development in the sector.	The policy options presented are consistent with the Directive, and with current practices and development in the sector.
Proportionality	Current differences in availability and terms of use across MS are an active barrier to creating re-use value, and holding back socio-economic growth (as evidenced by comparison with the open US market for same).	Creating a more common market and playing field is clearly helped by making these datasets available under the same conditions across the EU MS.	Creating a more common market and playing field is clearly helped by making these datasets available under the same conditions across the EU MS.
Legal and political feasibility	Political will is key in dealing with revenue loss and supporting needed investments. Legal changes are needed in some MS to remove mandated charges.	13 MS charge for observations data, 2 of which are fully dependent on revenue (though 1 is in the process of change). 9 MS have an open data policy and generally show benefits outweigh costs (where over time new tax revenue outpaces data revenue and provision costs)	Capabilities and resources needed for infrastructure investments may need central government support. Using the European Meteorological Infrastructure for public data provision may allow an effective use of resources.

Source: Deloitte compilation

Table 62 – Explanation of the performance values: Earth observation and environment

<b>Assessment Criteria</b>	<b>Weight which should be applied acc. to own research + expert interviews</b>	<b>PO1</b>	<b>PO2</b>
		<b>(lower intensity intervention)</b>	<b>(higher intensity intervention)</b>
		Reasoning/justification of the score	Reasoning/justification of the score
Effectiveness	The effectiveness correlates to the breadth and diversity of data available for EO/Environment. Use case generally are based on combinations of a range of data sets. All data considered falls within scope of INSPIRE, and is as such of high value almost by definition.	This option limits the data scope to environmental reporting priority data and the EMF INSPIRE theme. This supports limited use cases.	Including EO/Environment focused INSPIRE Themes as a whole covers the full variety of use cases.
Efficiency	The policy options efficiently build on existing INSPIRE infrastructures and practices, by adding an open data requirement that mostly means removal of current usage restrictions and terms of use. (Loss of revenue is considered under political feasibility)	Most of the environmental data is already being made available, albeit sometimes with usage restrictions. Adding open data requirements would not be a big step.	Building on existing INSPIRE infrastructure and practice by expanding the download facilities and removing restrictions is an efficient step. For some MS newly needed API infrastructure may introduce new costs.
Coherence	The policy options presented are consistent with the Directive, and with the INSPIRE recommendations, and thus with current practices and development in the sector.	The option is fully coherent with INSPIRE with regard to the data in scope. However because that scope is very limited it also limits the coherence with INSPIRE as a whole, as it leaves out all earth observation not pertaining to environmental data.	Fully coherent with INSPIRE
Proportionality	Current differences in availability and terms of use across MS are an active barrier to creating re-use value, and holding back socio-economic growth.	Consistent availability of these data across the MS does not require much effort, but does create EU-wide re-use potential, even if it only does so for a limited scope of data, and a limited scope of use cases.	Adding open data requirements to these INSPIRE Themes builds on extends the value of INSPIRE itself, and is a logical extension thereof. As there are hundreds of different datasets involved, including datasets that are charged for, efforts needed are less

			proportionate in this PO, although this policy option does resolve a much wider range of cross border barriers to re-use (because of the extended scope), than the lower intensity policy option.
Legal and political feasibility	Political will is key in dealing with revenue loss and supporting needed investments. Legal changes are needed in some MS to remove mandated charges.	Environmental data is already freely (gratis) accessible if not re-usable, and legal restrictions on some of the data (confidentiality and protection of e.g. sensitive species data), are covered as exceptions within the Directive. For some MS this policy option does too little however.	Given the wide scope, while being logical from the perspective of the data, theme and INSPIRE, there are differences across MS in what data is seen as easy to provide, and what data will incur a loss of revenue. So there will likely be political hesitance from different MS on different parts of the datasets involved. At the same time the lower intensity policy option doesn't reach far enough for other MS, balancing out the political feasibility for both policy options.

Source: Deloitte compilation

Table 63 – Explanation of the performance values: Statistics

<b>Assessment Criteria</b>	<b>Weight which should be applied acc. to own research + expert interviews</b>	<b>PO1 (lower intensity intervention)</b> Reasoning/justification of the score	<b>PO2 (higher intensity intervention)</b> Reasoning/justification of the score
<b>Effectiveness</b>	0.2	The policy can be easily implemented by the national data holders' organisations, due to the particular nature of the statistical datasets (open and freely available). The recommendations	While the implementation of this policy might raise some difficulties for the national data holders' organisations, the recommendations improve significantly the discoverability

<b>Assessment Criteria</b>	<b>Weight which should be applied acc. to own research + expert interviews</b>	<b>PO1 (lower intensity intervention)</b> Reasoning/justification of the score	<b>PO2 (higher intensity intervention)</b> Reasoning/justification of the score
		proposed have good potential to improve the re-use of datasets.	and reuse of the statistical datasets. The policy recommendation is more user-oriented and provides improved access to datasets for further use and re-use.
<b>Efficiency (B/C-ratio)</b>	<i>0.4</i>	The costs for implementing the recommendations remain low on average for the national data holders' organisations. At the same time, an increase in benefits is estimated to occur. However, since the changes proposed are minimal, the increase of benefits will remain limited.	The costs for implementing the recommendations will vary significantly across Member States, from low to medium-high (or even high) depending on the current status of the national data holders' organisations. However, the changes proposed will significantly increase the benefits as discoverability and re-use of the datasets is improved.
<b>Coherence</b>	<i>0.1</i>	The policy initiative is attainable for data holders' organisations. The recommendations of the policy initiative are in line with the other EU policy measures.	The policy initiative remains attainable for data holders' organisations, with various degree of efforts. The recommendations proposed remain in line with the other EU policy measures.
<b>Proportionality</b>	<i>0.1</i>	The recommendations are in line with the policy objectives. However, it matches only partially	The recommendations are in line with the policy objectives and the intensity fits better with the

<b>Assessment Criteria</b>	<b>Weight which should be applied acc. to own research + expert interviews</b>	<b>PO1 (lower intensity intervention)</b> Reasoning/justification of the score	<b>PO2 (higher intensity intervention)</b> Reasoning/justification of the score
		the objectives identified.	objectives identified.
<b>Feasibility (legal and political)</b>	0.2	The implementation of the policy option is feasible from both legal and political perspective, as the changes proposed are minimal and most of them already fulfilled by EU Member States.	The implementation of the recommendations could result in significant additional costs for countries where the technological readiness lags behind. As consequence, the political support might vary when it comes to specific options.

Source: Deloitte compilation

Table 64 – Explanation of the performance values: Mobility

Assessment criteria	PO1 (lower intensity intervention)	PO2 (higher intensity intervention)
<b>Effectiveness</b>	<p>In the lower intensity option for mobility, only a limited share of the datasets in scope are included HVDs (inland waterway infrastructure data and INSPIRE transport networks). Besides, these datasets are already widely accessible in the EU. However, the formalisation of their openness through open licenses and the establishment of APIs, as well as harmonisation of formats is expected to still slightly increase the effectiveness compared to the current baseline.</p>	<p>In the higher intensity option, in addition to the datasets in scope of the lower intensity option, all other national datasets related to transport networks (beyond those listed in INSPIRE) would be in scope. This is a significant increase in the scope of HVDs considered for the mobility thematic area, which would result in a very effective option compared to the current baseline. In fact, this would lead to the formal openness of all possible 'mobility' related datasets which are in the scope of this study (i.e. not included in ITS Directive).</p>
<b>Efficiency (B/C-ratio)</b>	<p>The lower intensity option is an option which is expected to produce limited, but positive benefits compared to the baseline, while generating limited costs to the EU Member States. Indeed, inland water way data is mostly covered via EU CEF funding already, while for INSPIRE transport networks the costs will be linked to the infrastructure to put in place only i.e. no new data transformation, processing, maintenance costs as harmonisation in these areas is already achieved through INSPIRE.</p>	<p>While the costs involved in the higher intensity option are expected to be high – as all other national datasets would also have to be transformed, processed, documented and maintained following the recommended principles, the benefits of this option are expected to strongly outweigh the costs. Indeed through the openness of such wide range of datasets, their reuse potential will be brought to its fullest, which in turn is expected to pave the way for a wide range of new services and applications having significant impacts from an economic, social and environmental point of view.</p>
<b>Coherence</b>	<p>The policy option is fully coherent with the current regulatory aquis, as the publication</p>	<p>The policy option is close to fully coherent with the current regulatory aquis, as the</p>

---

recommendations are totally aligned with the existing prescriptions of the RIS and INSPIRE Directives.

publication recommendations are totally aligned with the existing prescriptions of the RIS and INSPIRE Directives, however, as the option also encompasses datasets dealt with by national level legislation, some slight coherence issues may be considered in certain Member States where those datasets could currently be published in very different circumstances.

---

**Proportionality**

The policy option is fully proportional as it focuses on datasets already dealt with by well-established EU legislation and as its results could not be achieved by Member States individually, , thereby justifying EU action in this domain.

The higher intensity intervention options is less proportional than the lower intensity option in the sense that it encompasses EU intervention on currently nationally governed data. However, its results could not be achieved by Member States individually, thereby justifying EU action in this domain.

---

**Feasibility (legal and political)**

The policy option is regarded as fully feasible, as it only formalises practices which are already well-established throughout the EU. The obligation to the implementation of APIs to enhance the reuse of these datasets is not regarded as legally or politically sensitive.

Compared to the lower intensity option, the higher intensity option is considered as politically more sensitive, as it will involve the inclusion of currently nationally governed data under the EU intervention. However, as the measures for publication of the 'other' national transport networks datasets are less stringent than for INSPIRE transport networks datasets, this option is still considered as fairly feasible.

---

*Source: Deloitte compilation*

The scores/values assigned to each of the two policy options for each of the five criteria range from -3 (negative impact) to +3 (positive impact). The table below summarises the performance values, which are based on the explanations in the tables above.

Table 65 – Input matrix: performance values for the four intervention options

<b>Criteria</b>	<b>Performance range: -3 / +3</b>	<b>Lower intervention (PO1)</b>	<b>Higher intervention (PO2)</b>
<b>Company and company ownership</b>			
Effectiveness		1.50	2.50
Efficiency		2.50	1.50
Coherence		2.00	2.50
Proportionality		2.00	1.50
Feasibility		2.50	-1.00
<b>Geospatial</b>			
Effectiveness		1.75	2.75
Efficiency		3.00	1.75
Coherence		3.00	1.75
Proportionality		2.50	3.00
Feasibility		3.00	2.00
<b>Meteorological Data</b>			
Effectiveness		2.00	3.00
Efficiency		3.00	2.00
Coherence		2.00	3.00
Proportionality		3.00	3.00
Feasibility		1.00	1.00
<b>Earth observation and environment</b>			
Effectiveness		0.50	3.00
Efficiency		3.00	2.00
Coherence		2.00	3.00
Proportionality		2.50	2.00
Feasibility		2.50	2.50
<b>Statistics</b>			
Effectiveness		1.50	2.50
Efficiency		1.75	2.90



<b>Criteria</b>	<b>Performance range: -3 / +3</b>	<b>Lower intervention (PO1)</b>	<b>Higher intervention (PO2)</b>
Coherence		2.50	2.50
Proportionality		1.00	2.00
Feasibility		2.00	1.75
<b>Mobility</b>			
Effectiveness		1.00	2.50
Efficiency		1.50	1.50
Coherence		3.00	2.50
Proportionality		3.00	2.50
Feasibility		3.00	1.50

Source: Deloitte compilation

#### 4.1.1.2 Step 2: Outranking matrix

The aim of the outranking matrix is to summarise how the potential impacts of the baseline scenario and Policy Options compare against another for all possible pairs of Policy Options.

Practically, this means that for a given pair of options (e.g. lower intervention option (PO1) vs. higher intervention option (PO2)), the weightings<sup>506</sup> for each criterion are summed up for those criteria where PO1 is outranking PO2. In other words this means that the criteria in sum in relation to PO1 have a higher score than PO2, i.e. PO1 outranks PO2. Naturally, the combinations of PO1-PO1 and PO2-PO2 have received a score of 0 as it does not make sense to compare these.

The results are provided in the outranking matrix below.

Table 66 – Outranking matrix

<b>1. Company and company ownership</b>	<b>P1</b>	<b>P2</b>
P1	0.00	0.60
P2	0.40	0.00
<b>2. Geospatial</b>	<b>P1</b>	<b>P2</b>
P1	0.00	0.60
P2	0.40	0.00
<b>3. Meteorological data</b>	<b>P1</b>	<b>P2</b>
P1	0.00	0.20
P2	0.40	0.00

<sup>506</sup> Only the weightings are added. It makes no difference how much better each option is in respect of each of the criteria.

<b>4. Earth observation and environment</b>	<b>P1</b>	<b>P2</b>
P1	0.00	0.30
P2	0.40	0.00

<b>5. Statistics</b>	<b>P1</b>	<b>P2</b>
P1	0.00	0.30
P2	0.60	0.00

<b>6. Mobility</b>	<b>P1</b>	<b>P2</b>
P1	0.00	0.50
P2	0.30	0.00

Source: Deloitte

#### 4.1.1.3 Step 3: Permutation Matrix

As a third step, the outranking matrix is transformed into a permutation matrix. A permutation is a distinct combination of the Policy Options.

The aim of this step is to establish a decision matrix in order to select a final ranking of all the possible options that maximise pair-wise agreement (and minimise disagreement). The Policy Options are then scored by summing the elements from the outranking matrix for each policy pair which make up a given ranking of the Policy Options. The optimal ranking is the one with the highest score.

Since this study concerns two policy options, i.e. the lower intervention option and the higher intervention option, only two permutations (i.e.  $2 \times 1 = 2$ ) need to be assessed and compared towards each other for each of the six thematic areas.

An overview of all six possible permutations of the policy options is provided below.

Table 67 – Possible Permutations of the policy options

<b>#</b>	<b>1<sup>st</sup> element</b>	<b>2<sup>nd</sup> element</b>
1	PO1	PO2
2	PO2	PO1

Source: Deloitte

Next, the different so-called policy pairing<sup>507</sup> within these permutations need to be established in order to sum up the elements from the outranking matrix for each policy pair.

As each permutation consists of only the two Policy Options, i.e. two elements, only one pair of policy Options is possible within each permutation.

<sup>507</sup> The notion policy pairings reflects the mathematically possible combinations of the baseline scenario and the Policy Options. The notion has been used in order to comply with the Better Regulation Guidelines.

Table 68 – Possible Policy Pairings

#	1 <sup>st</sup> pairing
1	PO1PO2
2	PO2PO1

Source: Deloitte

For each of the policy pairings presented above, the summative weightings from the outranking matrix now have to be summed up. These sums are called coefficients of policy pairings.

The policy permutation with the highest total coefficient, based on the scores of each of the policy pairings, is going to be the optimal permutation of the baseline scenario and the Policy Options. This means that the first option within this permutation is the preferred Policy Option. Moreover, this permutation-based approach enables a decision not only on the preferred Policy Option but also ranks the less favoured Policy Options. According to this procedure of the MCA, the preferred policy option is the most favourable combination of coherence, effectiveness, efficiency, proportionality and feasibility.

#### 4.1.2 Results of the MCA and conclusions

Conducting the MCA with the performance values that were indicated in step 1, we obtained the following outcomes:

Table 69 – Outcome of MCA: preferred policy options

	Preferred PO
<b>Company &amp; Company Ownership</b>	PO1
<b>Geospatial Data</b>	PO1
<b>Meteorological Data</b>	PO2
<b>Earth Observation &amp; Environment</b>	PO2
<b>Statistics</b>	PO2
<b>Mobility</b>	PO1

Source: Deloitte

As a result, **Policy Option 1 (lower intensity intervention)** is identified as preferred option in the thematic areas of **Company and Company Ownership, Geospatial Data** and **Mobility**.

**Policy Option 2 (higher intensity intervention)** is the preferred option in the thematic areas **Meteorological Data, Earth Observation & Environment** and **Statistics**.

## 4.2 Cost estimations of the policy options

This subchapter provides an overview of the costs of each thematic area for both the lower and the higher intensity intervention option. It is important to note that the estimation provided in this subchapter

provides only a rough calculation based on the available data without asserting the claim to be complete and detailed in each aspect. Therefore, this cost estimation shall be understood to give a first initial understanding of the cost drivers for the Member States, stakeholders and datasets in scope by policy option.

Acknowledging this, it is necessary to build further work with more detailed cost calculations on the estimations provided below.

The general methodological approach used to determine the expected costs per thematic area is the following: First, the datasets in scope for each intervention option were determined to then estimate the number of stakeholders involved. It must be noted that for the vast majority of cases considered, the costs are only given for selected Member States and/or for selected areas. In cases where only selected cost (ranges) were available, these costs were used as an average and then extrapolated for the other Member States affected to the EU-27 level.

Due to the differences in scope, stakeholders and Member States affected, the detailed methodological approach is described for each thematic area separately.

#### **4.2.1 Company and company ownership**

In the thematic area company and company ownership, the stakeholders impacted can either be national authorities, courts, chambers of commerce or a public-private partnership. Depending on the Member State, the same authority might provide the data for different datasets in scope.

The main cost drivers for the thematic area company and company ownership can be categorised into infrastructural costs, data transformation costs, operational costs, other costs and lost income for data suppliers.

The most important costs are infrastructural cost which are costs related to infrastructural investments such as portals, APIs and servers. In the lower intensity intervention, around 75% of the Member States have to set up APIs for the provision of companies' information and establish API and bulk download. According to the study to support the review of Directive 2003/98/EC on the re-use of public sector information which was conducted by Deloitte in 2018, the costs of an API establishment range in between 30.000 EUR and 2.5 mio. EUR depending on the type of infrastructure and technical characteristics. On average, an API set up costs around 50.000 EUR. For the higher intensity intervention, the infrastructural costs would be the same as in the lower option but exacerbated as more data sets would be in scope.

Data transformation costs, the costs related to data processing including data cleaning, preparation of metadata, aggregation and anonymization are in between 4 and 10 full time equivalents (FTEs) employees for both options. For the higher intensity intervention option, more FTEs would be needed as more datasets would be in scope.

The operational costs, the cost related to data updates, replies to user requests, and corrections of errors in the datasets etc. can also be quantified to 4 to 10 FTEs for both intervention options.

Other costs, defined as any other costs such as legal advice on GDPR, training costs and other, largely depend on the circumstances of each Member State and therefore are not quantifiable. For the higher intensity intervention option, advice on GDPR will specifically be important in the company thematic area, as personal data will be shared.

Lost income for data suppliers, measured as the share of revenue related to the provision of HVD, highly depends on the datasets in scope. In the higher intensity intervention, all datasets in scope would be included as HVDs. Therefore, the total revenue would be lost. For the lower intervention option, in contrast, only a fraction of the total revenue would be lost as not all datasets in scope would be considered as HVDs. This means that the data holders would still be able to charge for some of the data points, and collect revenue. Revenues for selected countries can be found in the table below.

Table 70 - Revenues for selected Member States in the thematic area company and company ownership

<b>Country</b>	<b>Revenue (approximation/per year)</b>
Belgium	450 000 euro
Finland	2.7 million euro
France	6 million euro for INSEE and 2.4 million for INPI <sup>508</sup>
Germany	20 million
Estonia	2.6 million euro
Ireland	7 million euro
Italy	58-60 million euro
Poland	Around 30 000 euro (for specific services only, the data being provided for free)
Slovenia	1 million euro
The Netherlands	50 million euro

Source: Deloitte

According to the information available, the most important infrastructural costs (one-off costs) would be setting up an API. Costs estimation regarding the set-up of an API range between k€ 30 and € 2.5 million. The average was estimated at k€ 50. Based on the information, that approx. three quarter of the EU Member States would need to set-up an API under policy option 1 and policy option 2, the cost range is estimated between approx. k€ 600 (lower bound) and approx. € 50.6 million (upper bound) for the remaining (75%) EU 27 Member States in total. Based on the average costs for an API (k€ 50) total costs of € 1 million could be expected.

In addition, annual operational costs mainly related to data updates, replies to user requests, and corrections of errors in the datasets etc. can also be quantified to 4 to 10 FTEs for both intervention options. Taking into account average hourly labour costs in the EU27 according to Eurostat between € 20 for administrative and support service activities and approx. € 40 per hour in the ICT sector, total additional annual costs between € 3.2 million and € 16 million for the remaining (75%) EU 27 Member States in total could be expected for both policy options.

<sup>508</sup> Before the transition to an open data model in 2017, a compensation of 11 million euro was attributed to INSEE since but no compensation was foreseen for INPI.

### 4.2.2 Geospatial

With regard to geospatial datasets the loss of revenues should be regarded especially as it is a main cost driver. As far as the data on revenues are available, the Member States have a relatively wide range. As an example, revenues from the distribution of official basic geodata in Bavaria (Germany) amount to approximately € 14 million annually (approx. € 4 million from datasets of the National Survey, approx. € 10 million from datasets of the Real Estate Cadastre). With regard to cadastral parcels, the German Laender Hamburg and Berlin are in the range of € 1 million in terms of loss of revenue. Due to the federal structure, this is only a partial estimate of the overall revenues. In comparison, the estimated revenues in Ireland are in the range of € 4-5.5 million per annum, of these approximately € 1.5 million for opening addresses.

With regard to the costs, estimates based on the information available also indicate differences between the Member States. In Austria, the biggest cost factors concern data collection and maintenance. However, data collection and maintenance must take place anyway, regardless of any publication, and would therefore not be up for discussion. The additional costs for the public provision of these data are comparably low. Since this data has also to be kept internally independent of PSI, one-time setup costs of € 1000 per record are incurred for the provision under Open Data. The cost of the server infrastructure for the approx. 400 geodata sets for example, which the City of Vienna publishes under Open Data, is approx. € 50 000 per year.<sup>509</sup> For Sweden, costs related to geospatial data are estimated at SEK 553 million. Out of these SEK 553 million, SEK 511 million are losses of income from selling datasets and 42 million are investments and other costs that are needed to provide the data according to HVD regulations. SEK 426 million of the yearly loss of income from effected government authorities and municipalities comes from user fees from the private sector, and mainly from value added resellers and application providers. SEK 85 million are from fees that actors within the public sector charge each other for using data that are included in the Swedish proposal for HVDs.

For Latvia, costs are estimated at € 30-100 million for the entire geospatial sector. In Italy, costs are estimated at € 800 000 per annum and in Luxembourg € 10 000 per annum. For the Free State of Bavaria (Germany) annual expenditure of € 143 million (with revenues of € 5.9 million) for the department of the Bavarian Agency for Digitalisation, Broadband and Surveying (LDBV) (without state cadastral offices) are estimated. Estimated for the costs in The Netherlands amount to € 15 million. In Belgium, costs just for the geoportals of Flanders are approx. € 200 000.

### 4.2.3 Meteorological data

Regarding meteorological data, there is one data holder per Member State (acknowledging the fact that Spain and Italy have some regional organisations as well but also a national one that is within scope), which adds up in total to 27 stakeholders.

In the lower option, the costs are mostly determined by loss of revenue. High losses of revenue would be at the expense of political feasibility of the respective option. 13 Member States currently charge for observations data, 9 have an open data policy. The loss of revenue is in relative terms very high for Hungary and Austria as their budget depends heavily on it. A change of the revenue model in place may also need adapting existing laws with the accompanying effort, resources and time.

---

<sup>509</sup> The City of Vienna has been providing its (geo) data under Open Data since 2011 and potentially every user worldwide can access these data. Access to the Open Data geodata products via WMs and WFS is about 200 000 per year. Added to this are the address service accesses with 80 000 per day and the basemap.at WMTS accesses with up to 70 million per day.

In the higher intervention option, the costs for APIs are important. This, as radar and numerical weather prediction data can be quite voluminous, which requires heavier infrastructure.

When implementing the higher instead of the lower option, the costs for the following Member States are expected to rise: Croatia will have low-medium instead of low costs, France medium-high instead of medium costs, Finland, Ireland, Luxembourg and Greece will face high instead of medium costs, and Poland will face medium-high instead of medium costs. Denmark will have no change in losses of revenue as their transition is already planned irrespective of the PSI directive.

The low to high ranking is based on a relative comparison of the overall budget of the data holder. Low losses in revenues are losses up to 5% of the overall budget of the data holder, medium losses are 5-25% respectively, high losses consist in 25-45% of the overall budget for the provision of HVD of the data holder and very high losses are such losses above 45% of the overall budget of the data holder.

#### **4.2.4 Earth observation and environment**

For earth observation and environment, the lower intensity intervention option consists in publishing approximately 80 environment related data sets, whereas the higher intensity option adds an open data requirement to approximately 15 INSPIRE themes.

The differences between these options are large: As environmental data should be accessible for free already, however with a potential exception for commercial re-use, revenue loss plays no significant role in the lower intervention option. The opposite is true for the higher intervention option as earth observation data sets like land parcels, land use, ortho-imagery, elevation models are not always accessible for free.

Regarding the amount of data holders involved, there are also key differences between both options: For the lower intervention option, for environmental data at national level, there is usually one entity involved, though some regional or federal entities also exist in Germany, The Netherlands, Austria, Italy, and Spain. As the higher intervention option encompasses a much wider scope of data across INSPIRE themes, there will be probably at least 3-4 national entities involved in each Member State. Examples of such entities are environmental, hydrological, geological or geophysics agencies or land registries. Again, for the Member States mentioned in the lower intervention options, further subdivisions at the regional level could exist. For the higher intervention option, some overlaps in data holders with other areas exist. This could include the meteorological and also, potentially, the geospatial thematic area.

The higher intervention option has two cost drivers: first, infrastructure and API related cost and secondly, revenue loss. Of the 11 Member States, where information about costs was available, five countries (Cyprus, Denmark, Estonia, Finland and The Netherlands) would have low costs and six countries (Austria, Germany, Lithuania, Malta, Poland and Sweden) would have medium costs.

Again, low costs are defined as losses up to 5% of the overall budget for the provision of HVD of the data holder and medium losses are losses between 5-25% respectively.

The lower intervention option should be relatively low cost for all Member States, as this data must already be available through INSPIRE and the data sets do not entail complexities, thereby being limited to improving existing download and API facilities.

#### **4.2.5 Statistics**

For statistics, the main stakeholder affected in the Member States will be the National Statistical Institutes. Other public sector organisations as data providers might also be affected, but it is difficult to estimate the level of impact.

The main cost drivers depend on the intervention option. For the lower intervention option, cost drivers are the APIs provision and, although to a lesser extent, metadata provisions. In the case of metadata, the costs are estimated to remain on the low side, except for the countries where adjustments are needed to be implemented.

For the higher intervention option, the cost structure is, in addition to APIs development and metadata provisions, driven by the implementation of shared vocabulary and taxonomies and data linking. In case of shared vocabulary, the cost increase as compared to the lower intervention option is expected at low to medium cost as it is not clear to what extent this option might be already used by Member States. The second cost driver, data linking, is a rather cost-intensive option and thus is associated with high costs.

Regarding costs in statistics, the information on APIs is central, as it is the main driver for costs in both intervention options. APIs are already available in 16 Member States. In these Member States, the cost impact is therefore expected to be low. In the remaining 11 Member States, the cost impact is medium for Cyprus, the Czech Republic, and Romania and medium to high in Greece. For the other countries, no information is available regarding the cost impact.

In general, it is assumed that the provision of API and bulk download is associated with medium to high costs in both intervention options. Almost negligible as little to no costs occur for both intervention options are licence and terms of use for CC BY, the provision of an open format and the metadata and documentation.

Here, too, the amount of cost is defined as a relative share of the overall budget of the data holder. Low losses in revenues are losses up to 5% of the overall budget for the provision of HVD of the data holder, medium losses are 5-25% respectively, high losses consist in 25-45% of the overall budget of the data holder and very high losses are such losses above 45% of the overall budget of the data holder.

As the main cost driver in terms of one-off/infrastructural costs is the set-up of an API in both policy options, the overall costs could be assessed based on the number of Member States which would need to set-up an API. In total 11 of the 27 EU Member States would need to set-up an API in both policy options. Costs estimation regarding the set-up of an API range between k€ 30 and € 2.5 million. The average was estimated at k€ 50. Based on the information, that 11 of the EU Member States would need to set-up an API under policy option 1 and policy option 2, the cost range is estimated between approx. k€ 330 (lower bound) and approx. € 27.5 million (upper bound) for the remaining (75%) EU 27 Member States in total. Based on the average costs (k€ 50 per API set-up), total costs of k€ 550 for the remaining 11 of the EU27 Member States could be expected.

#### **4.2.6 Mobility**

For mobility, the lower intervention option covers inland waterway and river infrastructure data and INSPIRE Transport Networks. For inland waterway data, one main authority per country exists and the costs for all cost drivers have already been budgetised as part of an existing EU CEF funded project. Only 13 Member States are interconnected through waterway networks. The budgetised costs cover infrastructure data from these 13 Member States. Future, operational costs however must be borne by these 13 Member States themselves. The other Member States are not concerned by inland waterway data sets and have consequently no costs. For transport networks, all Member States are in scope. The main cost driver will be infrastructural costs for APIs, as the INSPIRE data is already provided for free and in the right formats.



The higher intervention option involves inland waterway and river infrastructure data, INSPIRE Transport Networks and all other transport networks data sets held by Member States. All cost elements of the lower policy option apply and in addition the following costs: for all other national transport networks-related datasets there may be, in addition to the infrastructure costs for APIs, costs for data transformation, lost income, and other operations and maintenance costs.

As the main cost driver is the set-up of an API, total costs for both policy options could be estimated. For inland waterways, 8 out of 21 Member States would be affected. Costs estimation regarding the set-up of an API range between k€ 30 and € 2.5 million. The average was estimated at k€ 50. Based on the information, that 8 of the EU Member States would need to set-up an API under policy option 1 and policy option 2, the cost range is estimated between approx. k€ 240 (lower bound) and approx. € 20 million (upper bound) for the remaining 8 of the EU 27 Member States in total. Based on the average costs (k€ 50 per API set-up), total costs of k€ 400 for the remaining 8 of the 21 Member States affected could be expected.

### 4.3 Modelling of the impacts of policy options

This section consists in comparing the expected macro-economic impacts of the lower and the higher intensity intervention and the preferred association of policy options on the overall economy and society compared to the baseline scenario.

#### 4.3.1 Methodological framework for the impact modelling

This section provides a brief explanation about the methodological approach for the macroeconomic analysis.

##### 4.3.1.1 Assumptions based on relevant literature and study findings

Our impact modelling assumptions are based on relevant literature and the findings of our study research so far. The table below provides an overview of the different studies that were considered as most relevant for the estimation of the baseline values.

However, several studies such as the latest study by the European Commission (2020) on the economic impact of open data point out that the exact reuse of open data and in particular PSI Data is barely documented. In addition, reuse can be endless and network effects can even multiply the economic value of PSI Data.

The market size of PSI data is therefore defined as the market size of products, services, and content improved or enabled by PSI Data.<sup>510</sup> However, which share of this value is attributed to open data can only be estimated. There exist several measures and methodological approaches to estimate the market value of PSI.

In this report, the market size is expressed in mEUR. In order to determine the market value of PSI, the Vickery 2011 study provided a main base for the prediction of the direct economic benefits of open data in the EU. Vickery 2011 looks at the impact of the adoption of the PSI Directive and extrapolate the sectoral and the national impact for the entire EU. Many studies after 2011, i.e. Deloitte 2018, refer to or are based on approaches or results from the Vickery 2011 study in order to determine the value of PSI.

Figure 4 – Our method: Assumptions used and literature

Study	Context	Assumptions about data economy/ open data
<b>Vickery 2011</b> , Review of Recent Studies on PSI Re-Use and Relate Market Developments	Part of the impact assessment of the first EU PSI Directive. Looks at the impact of the adoption of the PSI Directive in 2006, extrapolates sectoral + national impact research for the entire EU, based on GDP and ICT	Predicts a direct economic benefit of open data of 40 bio. EUR in the EU Predicts an indirect economic benefit of 140 bio. EUR in the EU. Predicts socio-economic benefits of 40 bio. EUR in the EU. Predicts EU market for government data in 2010 at about 32 bio. EUR.
<b>McKinsey 2013</b> , Open	Large scale report about the economic potential of data and	Estimates the total economic potential in the EU at 900 bio. Dollar, applying a bottom-up approach.

<sup>510</sup> See EC (2020), The economic impact of open data, pp. 18ff.

Study	Context	Assumptions about data economy/ open data
Data	open data with macro-economic estimates.	
<b>EU 2015</b> , Creating Value through open data	Macro-economic research into economic potential in the EU for 2016-20, building on Vickery 2011.	Predicts a direct economic value of open data of 75.7 bio. EUR in 2020. Predicts public sector savings in the EU of 1.7 bio. In 2020. Predicts an indirect value of open data in the EU from 265-286 bio. EUR in 2020.
<b>EU 2017</b> , Open Data Maturity in Europe	Yearly EU Data Portal Study	Predicted 325 bio. EUR of potential contribution of open data for 2016-2020 with 30.000 new jobs created in 2020.
<b>Deloitte 2018</b> , Reuse of Public Sector Information	Evaluation assessed the performance of the PSI Directive, whether it still responds to the stakeholders' needs and expectations and whether it fits the purpose of the next years.	Study based on the Vickery Study 2011 and assumptions of Eurostat and the Data Monitoring Tool. Assessment identified problem areas of re-use of PSI Data, among them costs of data re-use, availability, exclusive agreements

Source: Deloitte

Further, it is necessary to differentiate between the direct and the indirect impact of PSI Data. The direct market size refers to the monetized benefits that are realized in market transactions in revenues and gross value added (GVA). The indirect impact, in contrast, looks at the broader benefits of PSI, i.e. employment potential in downstream industries, new goods and services or increased productivity and efficiency.

The impact modelling will combine the direct and the indirect impact and refer to it as the total market size/ value of the PSI economy.

#### 4.3.1.2 Calculation of the baseline

The baseline has been calculated based on the extrapolation of the Vickery 2011<sup>511</sup> study and the forecasts of the European Data Market Monitoring Tool<sup>512</sup>. The European Data Market Monitoring Tool provides three forecast scenarios at the 2025 Horizon: a baseline scenario, a challenge scenario and a high growth scenario. The baseline scenario was chosen as the relevant scenario. The 2020 share of the PSI/HVD Market Size and Market Economy was derived applying the historic growth rates of the data economy to the value of the PSI related economy according to the Vickery Study.

In the year 2020, the outbreak of Covid-19 massively affected the European economy. Expected figures have been corrected to take into account the impact of this crisis.<sup>513</sup> Forecasts have been extrapolated to

<sup>511</sup> Vickery 2011, Review of Recent Studies on PSI Re-Use and Relate Market Developments.

<sup>512</sup> Data landscape, The European Data Market Monitoring Tool see: <http://datalandscape.eu/european-data-market-monitoring-tool-2018>

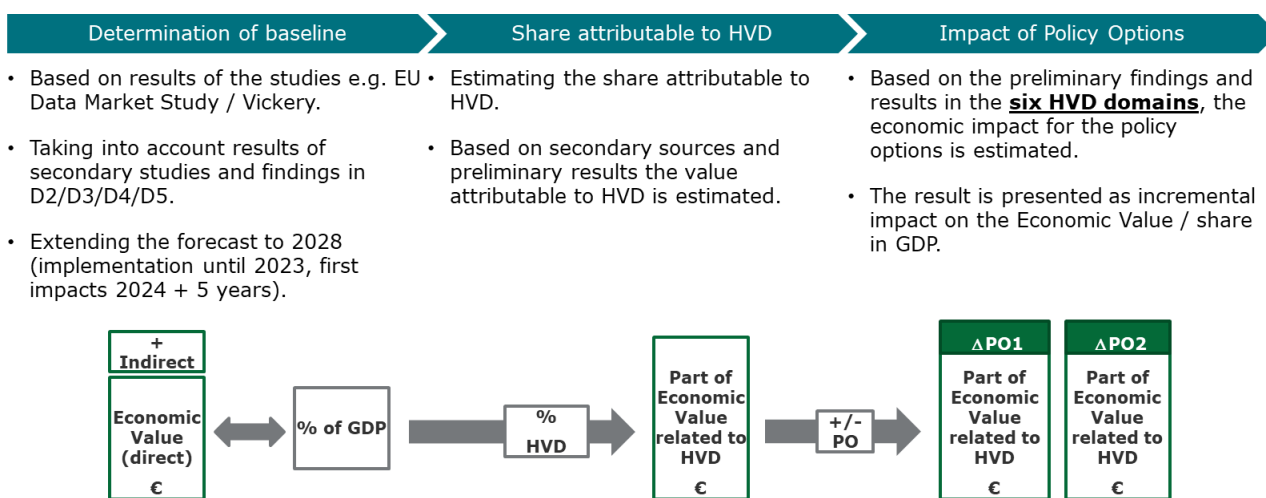
<sup>513</sup> The European Data Market Monitoring Tool already provides a Covid19 correction of the 2025 forecast. We have made further annual adjustments according to The Economist intelligence Unit data forecast of GDP, which includes Covid19 corrections for 2020.

provide a baseline until year 2028, taking into account the projected growth rates of the EU Data Monitoring Tool and Vickery for the short-term.<sup>514</sup> OECD long-term GDP forecast has been used in the last years of the reference period. In this regard, the forecast could be interpreted as being conservative, since the data economy might be expected to grow even faster than GDP in general.<sup>515</sup>

### 4.3.1.3 Top-down analysis

In order to obtain the economic impact of the PSI/HVD economy, in relation to its contribution to GDP, a top-down analysis has been performed. It was assumed, that the implementation of the policy options would take until 2023 and that no significant effects will occur before 2024. Consequently, the period 2024-2028 was chosen as reference period for estimating the impacts.

Figure 5 – Our method: Top-down approach



Source: Deloitte

- The data from the European Data Monitoring Tool provides a baseline for the Economic value of the data economy and relates it to GDP. We have used this data to calculate the baseline. Adjustments with regard to Covid-19 outbreak macroeconomic impact have been included.
- In order to define the relevant market size/size of the specific economic, that the policy intervention foreseen as part of this study can address, two conservative assumptions were made: The share attributable to high value data was based on the renowned Vickery study from 2011 that estimated the share attributable to the PSI economy for 2010. This value was extrapolated to the year 2020 and then adjusted for corrections with the baseline values of the data economy from the European Data Monitoring Tool. Relevant other studies and secondary literature, in particular the EU study (2015) on Creating Value through open data, the EU study (2017) on Open Data Maturity in Europe and also the Deloitte (2018) study on Reuse of Public Sector Information were considered for adjusting the Vickery 2011 baseline.

<sup>514</sup> The European Data Market Monitoring Tool provides a forecast until 2025. The forecast has been extrapolated by simply applying long-term GDP forecasts of the OECD and the Economist intelligence Unit for the individual EU27 Member States.

<sup>515</sup> The forecast for the growth of the EU data economy, however, ends in 2025. In order to calculate impacts beyond 2025 we have taken a conservative approach and calculated the impacts on the basis of the GDP growth rate forecast of the OECD. For this reason the impacts are based on a much lower per annum growth rate.

- For each of the six thematic areas, further assumptions have been made to understand in more details the magnitude of impact of the specific policy options. The economic impact of the different policy options/packages (on GDP) is measured for each indicator on an incremental basis (vs. the baseline). In this regard, ratios are estimated. The experts' assumptions are based on the findings of the interviews carried out and the literature studies.

The EU Data Monitoring Tool data for the data economy and the PSI baseline addressed are presented in the table below. The growth rates of the period 2026-2028 are based on the OECD long-term GDP forecast and are highlighted in dark green.

Figure 6 – Baseline estimates

HVD   Baseline and Economic Impact										
M€	2020	2021	2022	2023	2024	2025	2026	2027	2028	
<b>EU Data Monitoring Tool 2020 - baseline</b>										
Data revenues	71 050	75 866	81 008	86 499	92 362	98 623	100 144	101 711	103 321	
Data market value	62 244	65 795	69 584	73 628	77 948	82 564	83 837	85 149	86 497	
Data market value - direct impact	54 081	58 481	63 239	68 385	73 948	79 965	81 198	82 469	83 775	
growth rate %	(7.1%)	8.1%	8.1%	8.1%	8.1%	8.1%	1.5%	1.6%	1.6%	
<b>EU Data Monitoring Tool 2020 - high growth</b>										
Data revenues	71 050	80 943	92 215	105 055	119 684	136 350	138 453	140 620	142 846	
Data market value	62 244	69 320	77 236	86 097	96 020	107 139	108 791	110 494	112 243	
Data market value - direct impact	54 081	62 005	71 090	81 505	93 447	107 139	108 791	110 494	112 243	
growth rate %	(7.1%)	14.7%	14.7%	14.7%	14.7%	14.7%	1.5%	1.6%	1.6%	
% above baseline	-	6.0%	12.4%	19.2%	26.4%	34.0%	34.0%	34.0%	34.0%	
<b>Market size PSI - Vickery</b>										
Assumption: baseline growth rate: 7%	(7.1%)	6.5%	6.5%	6.5%	6.5%	6.5%	1.5%	1.6%	1.6%	
% PSI on data market value	90%	89%	87%	86%	85%	83%	83%	83%	83%	
	forecast based on EU Data Monitoring Tool					OECD GDP forecast				

Source: Deloitte compilation

Based on existing literature and on the study research, we estimated the market share of the six thematic areas in % of the PSI market for the baseline scenario. Consequently, we could estimate the direct impact in mEUR of each thematic area for the years 2024-2028 (implementation of policy options until 2023 + 5 years afterwards), keeping the shares constant in absence of any policy intervention.

Figure 7 – Baseline estimates – market shares

HVD   Baseline and Economic Impact										
M€	2020	2021	2022	2023	2024	2025	2026	2027	2028	
<b>Market size PSI - Vickery</b>										
Assumption: baseline growth rate: 7%	(7.1%)	6.5%	6.5%	6.5%	6.5%	6.5%	1.5%	1.6%	1.6%	
% PSI on data market value	90%	89%	87%	86%	85%	83%	83%	83%	83%	
	forecast based on EU Data Monitoring Tool					OECD GDP forecast				
<b>Market share [% of PSI Market]</b>										
1. Company and company ownership	6%	6%	6%	6%	6%	6%	6%	6%	6%	
2. Geospatial	34%	34%	34%	34%	34%	34%	34%	34%	34%	
3. Meteorological data	20%	20%	20%	20%	20%	20%	20%	20%	20%	
4. Earth observation and environment	15%	15%	15%	15%	15%	15%	15%	15%	15%	
5. Statistics	16%	16%	16%	16%	16%	16%	16%	16%	16%	
6. Mobility	9%	9%	9%	9%	9%	9%	9%	9%	9%	
<b>Total</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	<b>100.0%</b>	
<b>Market share</b>										
1. Company and company ownership	3 016	3 212	3 421	3 643	3 880	4 132	4 196	4 262	4 329	
2. Geospatial	16 541	17 616	18 761	19 980	21 279	22 662	23 011	23 372	23 742	
3. Meteorological data	9 730	10 362	11 036	11 753	12 517	13 331	13 536	13 748	13 966	
4. Earth observation and environment	7 200	7 668	8 166	8 697	9 263	9 865	10 017	10 174	10 335	
5. Statistics	7 784	8 290	8 829	9 402	10 014	10 664	10 829	10 998	11 173	
6. Mobility	4 378	4 663	4 966	5 289	5 633	5 999	6 091	6 187	6 285	
<b>Total</b>	<b>48 649</b>	<b>51 811</b>	<b>55 179</b>	<b>58 765</b>	<b>62 585</b>	<b>66 653</b>	<b>67 681</b>	<b>68 740</b>	<b>69 828</b>	

Source: Deloitte compilation

### 4.3.2 Impacts of policy options on the economy

#### 4.3.2.1 Direct economic impacts

The Impact Assessment support study took as the baseline the data market value as estimated by the EU Data Monitoring Tool of 54 081 mEUR for the baseline scenario.<sup>516</sup> These numbers take into account a correction linked to Covid-19 impact on the overall EU economy. Thereof, the market size of the PSI Data was estimated with the Vickery value from 2011 with corresponding corrections as described above. The baseline PSI market value in 2020 is estimated 48 649 mEUR. This value already takes into account a correction linked to the Covid-19 impact on the overall EU economy.

The growth rates of the baseline scenario for the PSI market are conservative: A growth rate of 6.5% for the years from 2021-2025 is applied. The literature indicates an average growth rate of 7% of PSI Data and for the data economy the EU Data Monitoring Tool has estimated a CAGR of app. 9% in general from 2002-2025.<sup>517</sup> From the years 2026-2028, we have assumed that the growth rate of the baseline PSI market follows the path of the growth rates as estimated by the OECD for the EU27 economy in general. This growth rate can be viewed as conservative, as it could be assumed that the data industry will continue to grow faster than the economy in general even after 2025. However, it also can be argued that the growth rate slows down a bit as the overall trend of the growth rate of the economy decreases.

Under these assumptions, the PSI market value in 2028 could increase to 69 828 mEUR.

The macroeconomic impacts of the policy packages are based on the respective growth rates for the different thematic areas. Each thematic area has other policy intervention options which cover a different scope, different policy areas and industries and therefore they are expected to have distinct growth rates.

Hence, we estimated the growth rates for the lower intervention option and the high intervention option for each thematic area separately and then summed up the impact of each of the six thematic areas to obtain the overall impact of the policy packages. All assumptions are based on study findings, expert interviews and estimations and desktop research.

This procedure allows us to differentiate the growth between different thematic areas and also account for already mature areas such as statistics, where – due to the maturity of the area – lower growth rates would be expected. Where a high potential of data re-use is expected, such as in geospatial data which is currently as being a more important high value data area, the growth is assumed to be higher. As the scope also widens from the lower to the higher intervention, it therefore was assumed that the growth rates are higher in the higher intervention option.

The **baseline growth rates** for each thematic area correspond to the baseline growth rate of the PSI market value of 6.5% annually. For the lower and the higher intervention option, the growth rates are, depending on the intensity and scope of the intervention options expected to be higher than the baseline growth rate of 6.5% annually. Therefore, three changes in growth rates were categorised: a slight change in the growth rate, a medium and a major change in the growth rate. The changes in the growth rates are a result of the scope and nature of the respective intervention option. For a slight change in the growth rate a differential of + 0.5-1.5% was added to the baseline growth rate. For a medium change in the growth rate, + 1.5-2.0%, depending on the scope and areas covered by the intervention option, was

---

<sup>516</sup> It must be noted that the European Data Market Monitoring Tool uses the “Value of the Data Market” as a proxy for the direct economic value. The Value of the Market is calculated based on revenues of data companies, excluding exports and including imports. At the very least, it should be noted that imports do not usually contribute directly to GDP, while they influence foreign GDP (while exports contribute to domestic GDP).

<sup>517</sup> See e.g. European Commission, European Data Portal (2020), The Economic Impact of Open Data: Opportunities for value creation in Europe. Study conducted by Capgemini

added. A major change in growth rate was asserted with a differential of + 2.0-3.0%-points as compared to the baseline growth rate.

As we assumed that the growth rate of the baseline PSI market follows for the years 2026-2028 OECD estimates, meaning that it will decrease by 5%, we accordingly assumed that the growth rates for each thematic area will also lower for the years 2026-2028. We therefore reduced the growth rates of the thematic areas for the years 2026-2028 by 5%-points accordingly. Here, again, we assume a conservative estimation for the years 2026-2028, acknowledging the fact that growth rate estimations for the PSI market are currently not provided by official EU sources such as the EU Data Monitoring Tool.

In **company and company ownership**, the lower intensity intervention option includes only two datasets in scope which would be, as described in the specific section above, the low-hanging fruits as the implementation would require little effort from member states. Hence, only a slight change of the growth rate in comparison to the baseline is expected and therefore a growth rate of 7.0% for the lower intensity intervention option was assigned. The higher intensity intervention option would include a wide scope of datasets and consequently a medium to major change of the growth rate in comparison to the baseline. Considering that the company and company ownership high value data set market is relatively small, a growth rate of 8.5% was assumed.

For the lower intervention option for **geospatial** high value data, only limited requirements of data will be necessary which requires only minimum changes in the current options. 4-5 high value datasets will be in scope with this option. It can be concluded that this intervention option would result in a slight to medium change in the growth rate of the geospatial sector, also because of the importance of geospatial data and its expected impact to affected stakeholders. Taking this into account, a growth rate of 7.5 % for the lower intervention option of geospatial high value data was assumed. The higher intervention option is the far-reaching option, whereas the main differences to the lower intensity option refer to the licence, the APIs and few changes on the granularity and key attributes. Considering this medium change in comparison to the baseline, a growth rate of 8.5% was attributed to the higher intensity intervention option.

The **meteorological** lower intensity intervention option encompasses three datasets, namely observations, climate and digitised structured historical climate data. This intervention stays close to what is already common, but leaves room for Member States to adopt more advanced technology. Hence, this intervention option would go along with a medium change of the growth rate in comparison to the baseline. A growth rate of 8.0% was therefore attributed to the lower intensity intervention option of meteorological high value data. In the meteorological higher intensity option, three additions are made, unstructured historical data, radar data and numerical weather prediction model data. As all data sets are highly voluminous, this intervention option can be considered as a major change in the growth rate in comparison to the baseline. Therefore, a growth rate of 9.5% was attributed to the higher intervention option of meteorological datasets.

The lower intensity intervention option of **earth environment and observation** includes an open data obligation to the environmental reporting and observation data which means mostly removing reuse restrictions. This constitutes an expected medium change of the growth rate in comparison to the baseline, for which a growth rate of 8.0% was assigned. The higher intensity intervention option involves that the scope of data involved is broadened by adding additional INSPIRE themes relevant to earth and observation. This intervention option implies a medium to major change in comparison to the baseline. A growth rate for the higher intensity intervention option of 9.0% was therefore assigned.

For **statistics**, the lower intensity intervention option requires only a set of minimum changes to the current publication options available which is expected to have low to no impact on the institutions and

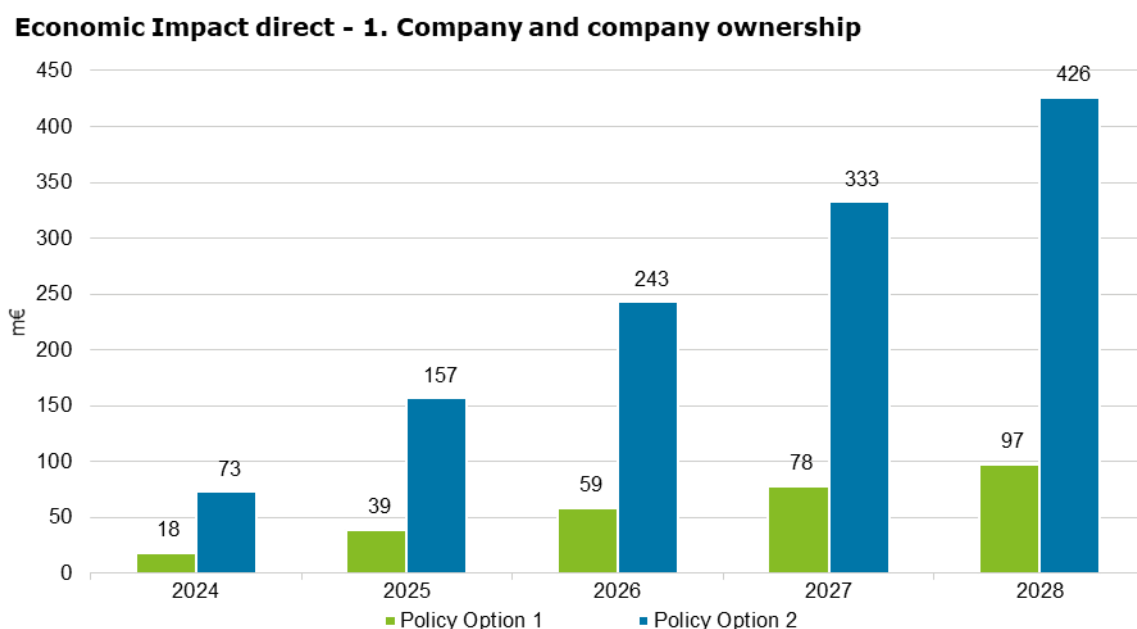
stakeholders affected. Hence, no to slight changes of the growth rate of statistics in comparison to the baseline is expected for the lower intensity intervention option. Therefore, a growth rate of 7.0% for the lower intensity intervention option for statistics was assumed. The higher intensity intervention option implies new changes in terms of measures for publication. Improved APIs and metadata files, development of controlled vocabularies and taxonomies are expected which therefore constitutes a medium change in comparison to the baseline, also because the market of high value data in statistics is already quite advanced relative to the other thematic areas. A growth rate of 8.0% for the higher intensity intervention option in statistics is consequently assumed.

For **mobility**, the lower intervention option covers, as stated in chapter 3.1, only the bare minimum of the datasets and 2 out of 4 categories. This signifies a slight to medium change of the growth rate in comparison to the baseline. A growth rate for this slight to medium change of 7.5% was therefore assigned for the lower intervention option of mobility. The higher intervention option, in contrast, is more ambitious than the previous one as all four categories of datasets are included in the high value data list. Hence, this can be interpreted as a medium change in the growth rate in comparison to the baseline and therefore a growth rate of 8.0% for the higher intervention option in mobility was assigned.

With regard to the **economic impact (incremental economic value)**, the Policy Option 2 (higher intensity intervention) creates the most impact on the total economic value. This result is logical, as the higher intensity intervention will affect more datasets in each of the six thematic areas. The incremental impact for the higher intensity intervention is estimated to be between approx. 2 and 4 times higher than for policy option 1. The different results for policy option 1 and policy option 2 are presented in the following graphs for each thematic area.

#### 4.3.2.1.1 Company and company ownership

Figure 8 – Direct economic impact per Policy Option by thematic area – company and company ownership

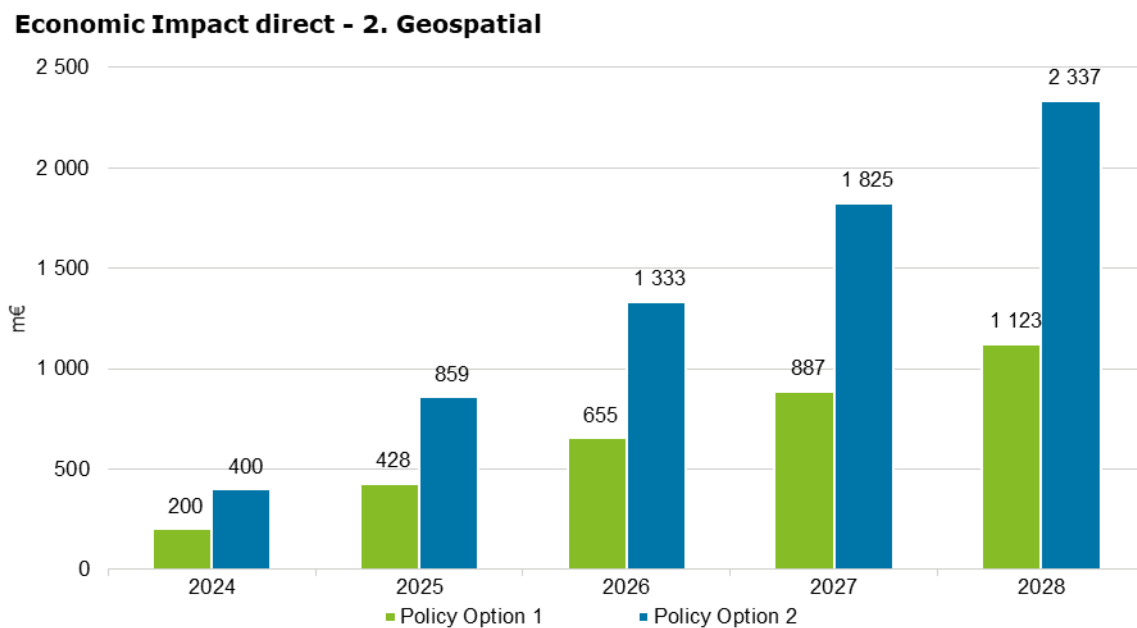


Source: Deloitte estimation



### 4.3.2.1.2 Geospatial

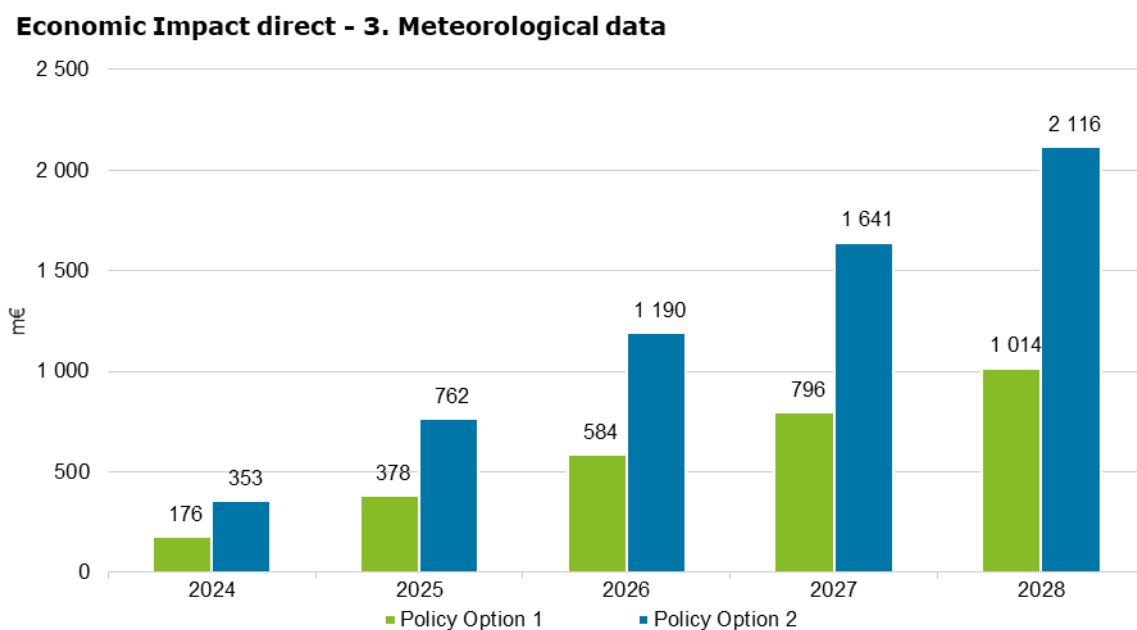
Figure 9 - Direct economic impact - geospatial



Source: Deloitte estimation

### 4.3.2.1.3 Meteorological

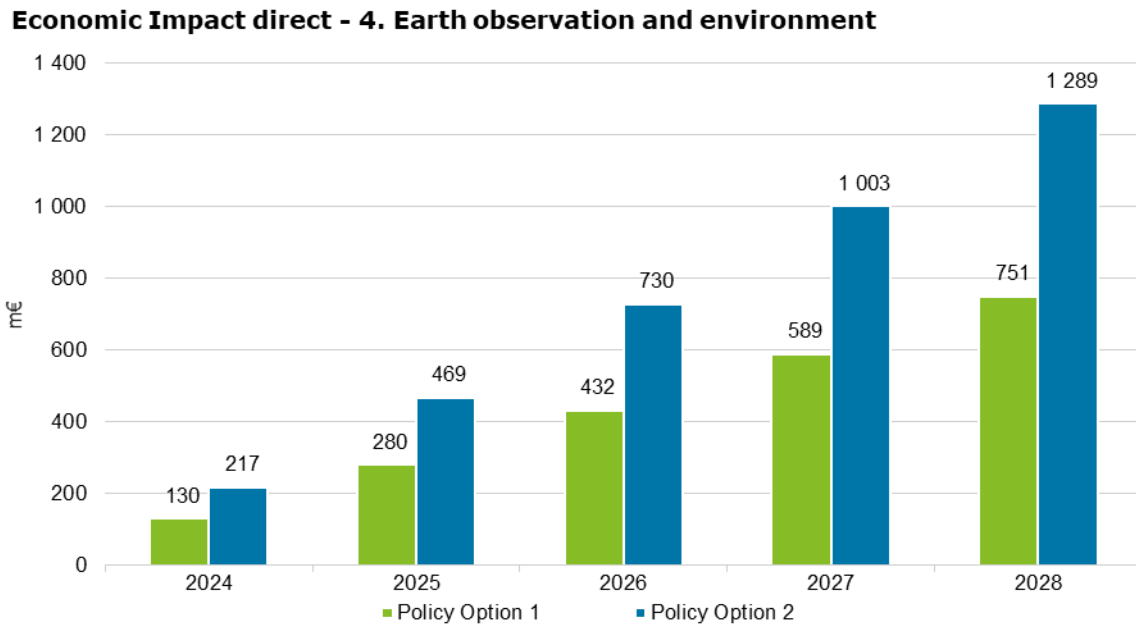
Figure 10 - Direct economic impact - meteorological data



Source: Deloitte estimation

#### 4.3.2.1.4 Earth observation and environment

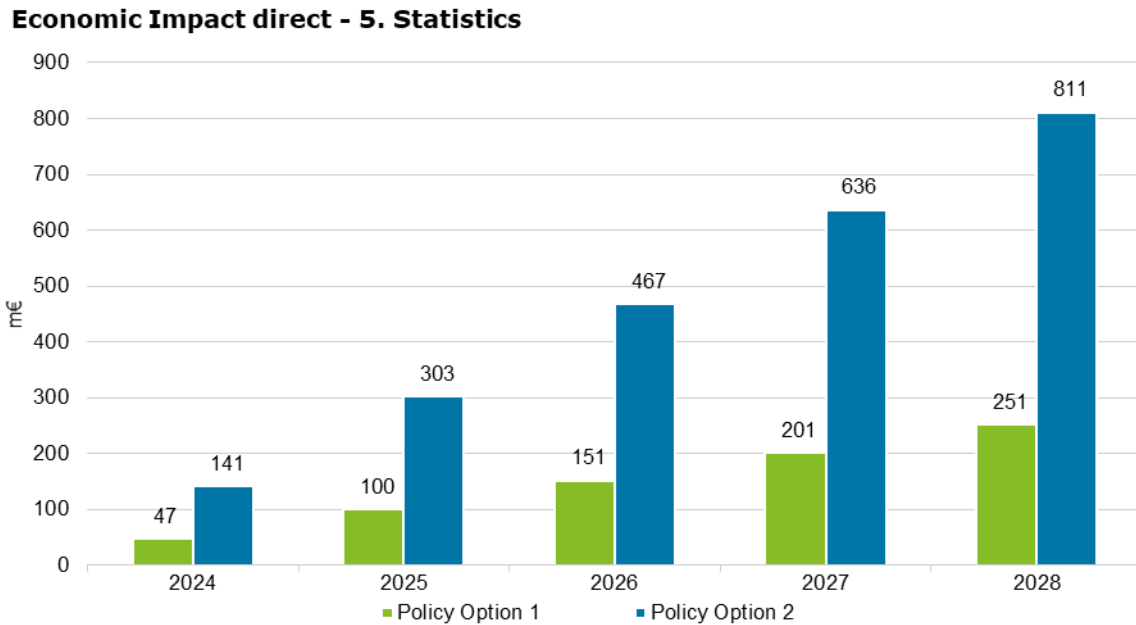
Figure 11 - Direct economic impact - earth observation and environment



Source: Deloitte estimation

#### 4.3.2.1.5 Statistics

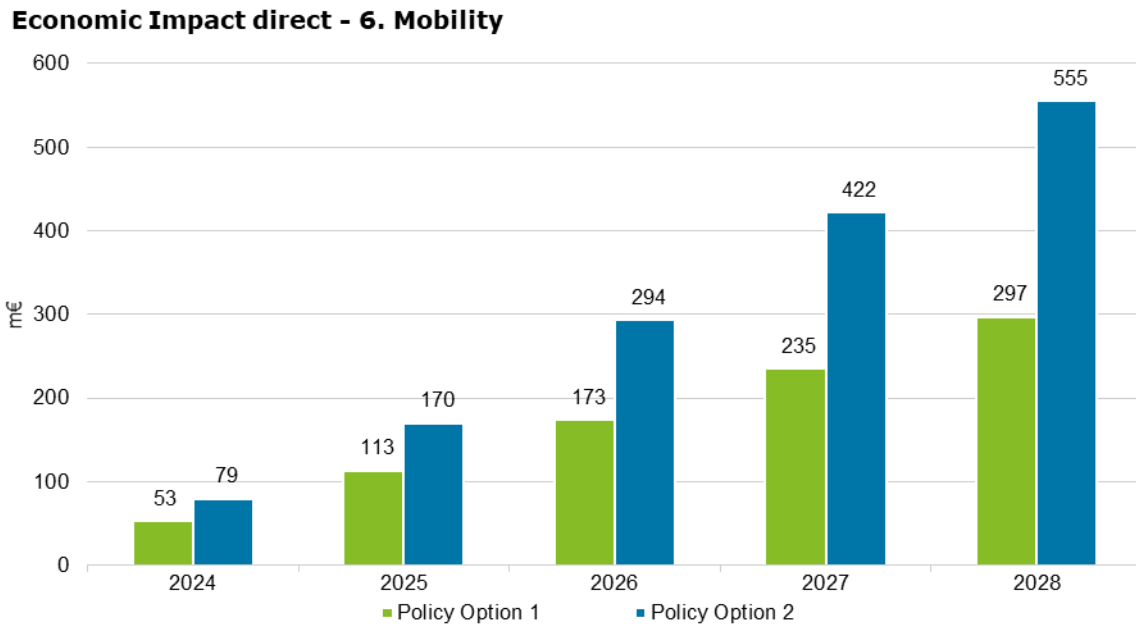
Figure 12 - Direct economic impact statistics



Source: Deloitte estimation

4.3.2.1.6 Mobility

Figure 13 - Direct economic impact - mobility

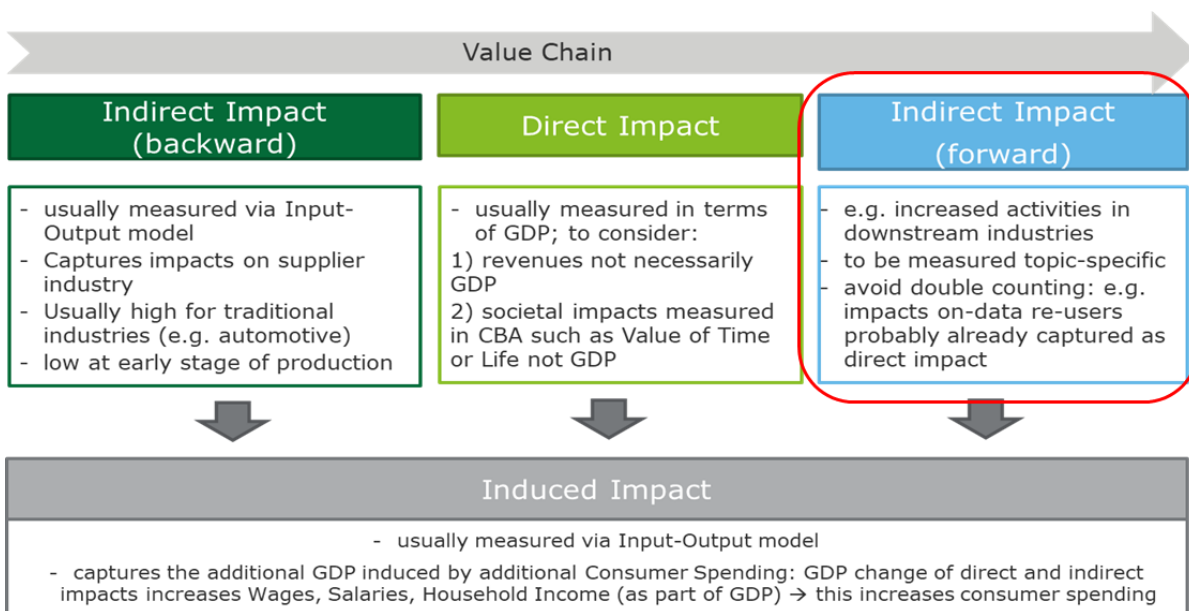


Source: Deloitte estimation

4.3.2.2 Indirect economic impacts

In order to fully reflect on the overall impact on the economy, indirect impacts can be calculated based on the estimates of the direct impacts. In general, different types of multipliers can be considered in the context of an economic impact assessment. An overview of the different multipliers potentially applicable is presented below.

Figure 14 – Overview of direct, indirect, and induced impacts



Source: Deloitte.

The classical indirect impact included in economic impact assessment is upstream/backward oriented and focuses on the impact of supplier industries.<sup>518</sup> These impacts are expected to be relatively low for the data industry compared to traditional industries of the manufacturing sector, because data companies (direct activities of the data economy) including PSI/HVD data activities are usually expected to be positioned at an early stage of value chain.<sup>519</sup> Contrary, the indirect effect on downstream industries (forward) is expected to be more significant, because products and services based on PSI/HVD data are considered to be used in various other industry sectors. Taking this into account, the analysis focuses on the indirect (forward) impact on downstream industries, which is considered to be the major indirect impact.<sup>520</sup> <sup>521</sup> To measure these impacts with regard to the PSI/HVD economic activity, results of the EU Data Monitoring have been analysed.<sup>522</sup> As a result, for the indirect (forward) impact a magnitude between ca. 2.5 and 3.0 has been identified.

Table 71 - Multipliers

<b>HVD   Baseline and Economic Impact</b>									
<b>M€</b>	<b>2020</b>	<b>2021</b>	<b>2022</b>	<b>2023</b>	<b>2024</b>	<b>2025</b>	<b>2026</b>	<b>2027</b>	<b>2028</b>
<b>EU Data Monitoring Tool Multipliers (% of direct impact)</b>									
<u>Baseline</u>									
Direct Impact	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Indirect Backward Impact	0.06	0.06	0.06	0.06	0.06	0.05	0.05	0.05	0.05
Indirect Forward Impact	2.79	2.76	2.74	2.71	2.68	2.66	2.66	2.66	2.66
Induced Impact	1.83	1.97	2.12	2.29	2.47	2.66	2.66	2.66	2.66
<b>Total Impact</b>	<b>5.68</b>	<b>5.79</b>	<b>5.91</b>	<b>6.05</b>	<b>6.20</b>	<b>6.37</b>	<b>6.37</b>	<b>6.37</b>	<b>6.37</b>
<u>High Growth</u>									
Direct Impact	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Indirect Backward Impact	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
Indirect Forward Impact	2.79	2.84	2.89	2.94	2.99	3.04	3.04	3.04	3.04
Induced Impact	1.83	2.09	2.39	2.74	3.13	3.58	3.58	3.58	3.58
<b>Total Impact</b>	<b>5.68</b>	<b>5.99</b>	<b>6.34</b>	<b>6.73</b>	<b>7.18</b>	<b>7.68</b>	<b>7.68</b>	<b>7.68</b>	<b>7.68</b>

Source: Deloitte compilation based on EU Data Monitoring Tool

To estimate the indirect (forward) economic impact, multipliers at the lower bound between 2.6 and 2.8 have been applied, which can be considered as conservative approach. The results which estimate the indirect economic impact for the different Policy Options per thematic area are presented below.

<sup>518</sup> Those types of indirect effects are usually captured by Input-Output analysis/models. The impact is considered high especially for traditional industries of the manufacturing sector, like e.g. the automotive industry.

<sup>519</sup> Within the Study on the EU Data Market the indirect (backward) impact was estimated to be remarkably low, at around only 6%. Taking this into account, the indirect (backward) multiplier has not applied in addition, because the indirect forward multiplier estimated in the same study is in contrary already relatively high.

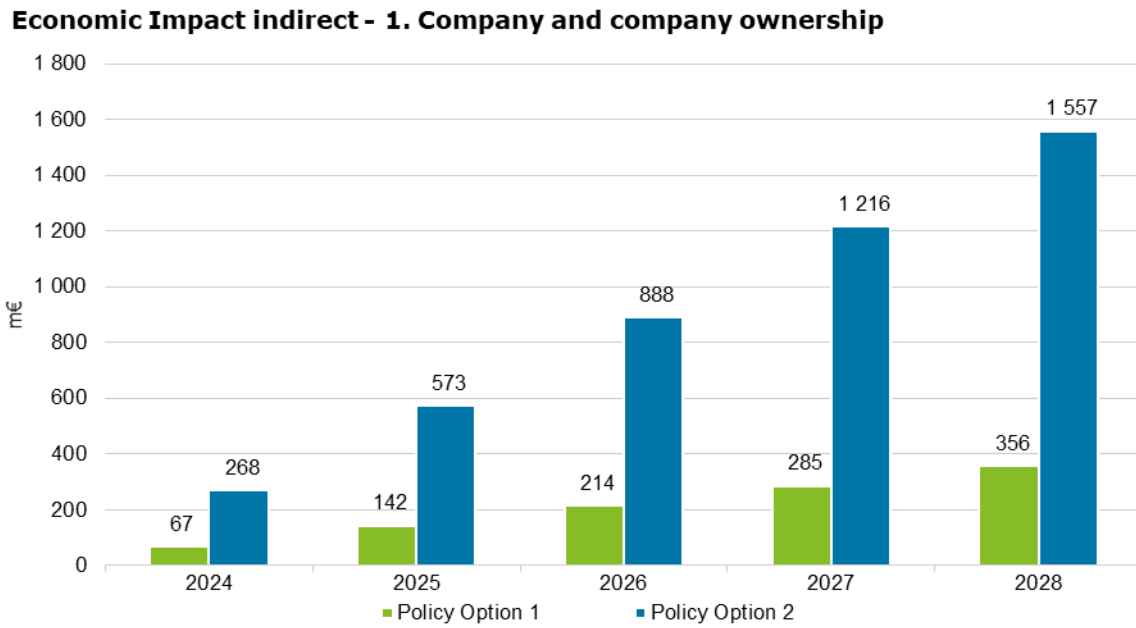
<sup>520</sup> These impacts are specific and usually cannot be measured by general, standard economic models like e.g. input-output-models.

<sup>521</sup> In addition, induced impacts could be considered in an economic impact assessment. However, following the classical definition, the induced effects are defined as so called type II multipliers. These multipliers not only account for the direct and indirect impacts, but they also account for induced impacts based on the purchases (consumption) made by employees. In this regard, they reflect general economic effects induced by increased consumer activities, which are triggered by increased employment and labor income as a result of the direct and indirect effects. These broader impacts have been excluded in this analysis.

<sup>522</sup> The European Data Monitoring implicitly includes several types of multipliers, including indirect and induced impacts, which estimate impacts on the supplier industries and the overall economy generated through additional income and consumption (both could be classically estimated using e.g. Input-Output models), as well as indirect forward impacts, which estimate the effects downstream in the economy. To stay conservative, the later one have been considered here based on the European Data Monitoring Tool, since those impacts are expected to be of major interest. The European Data Monitoring Tool in this regard estimates coefficients between 2.6 in the baseline as a lower bound and 3.0 in the high growth scenario as an upper bound.

### 4.3.2.2.1 Company and company ownership

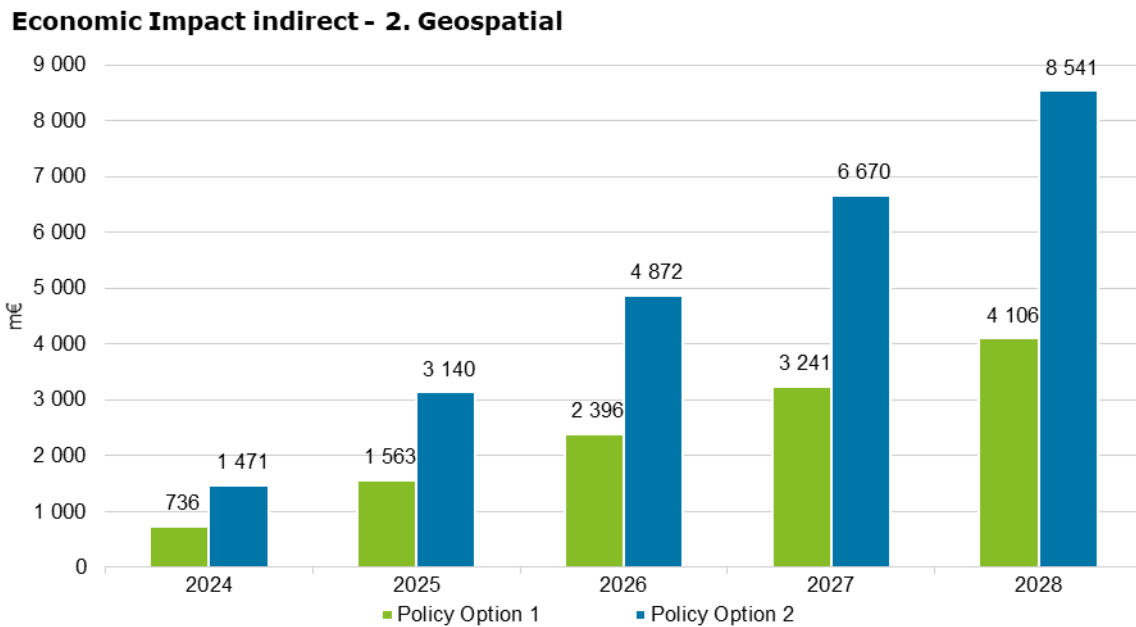
Figure 15 – Indirect economic impact – Company and company ownership



Source: Deloitte estimation

### 4.3.2.2.2 Geospatial

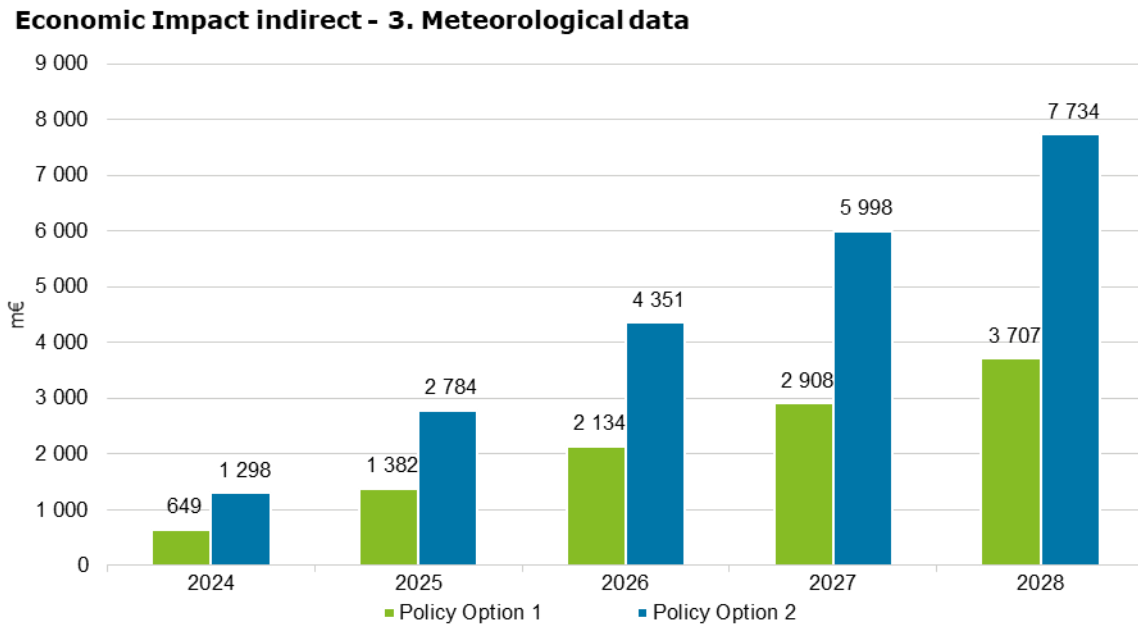
Figure 16 - Indirect economic impact - Geospatial



Source: Deloitte estimation

### 4.3.2.2.3 Meteorological

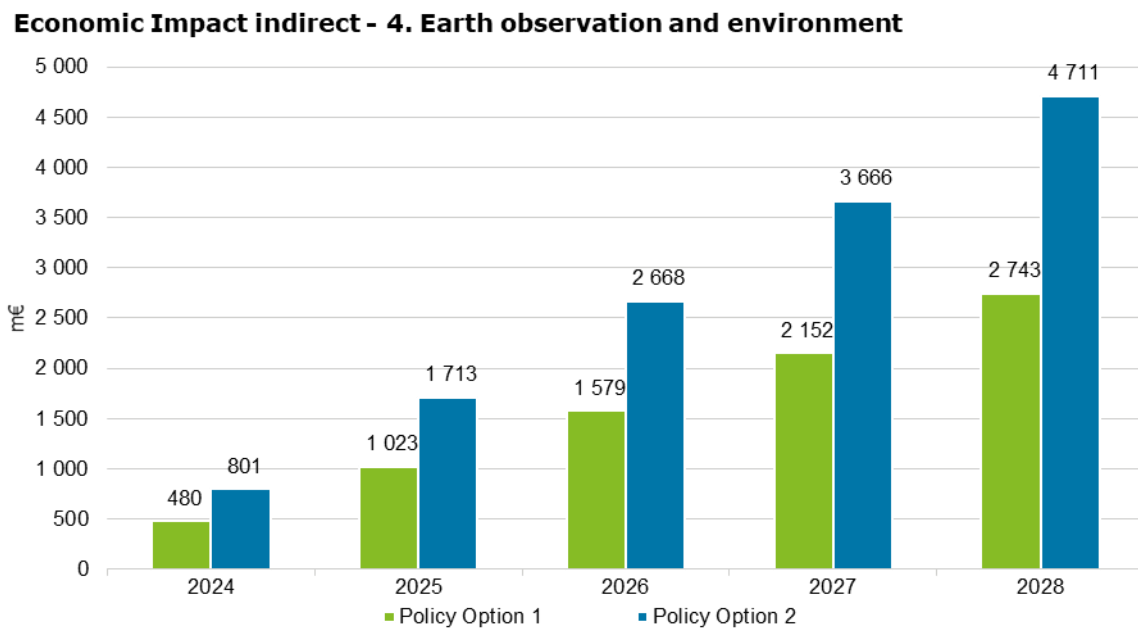
Figure 17 - Indirect economic impact - Meteorological Data



Source: Deloitte estimation

### 4.3.2.2.4 Earth observation and environment

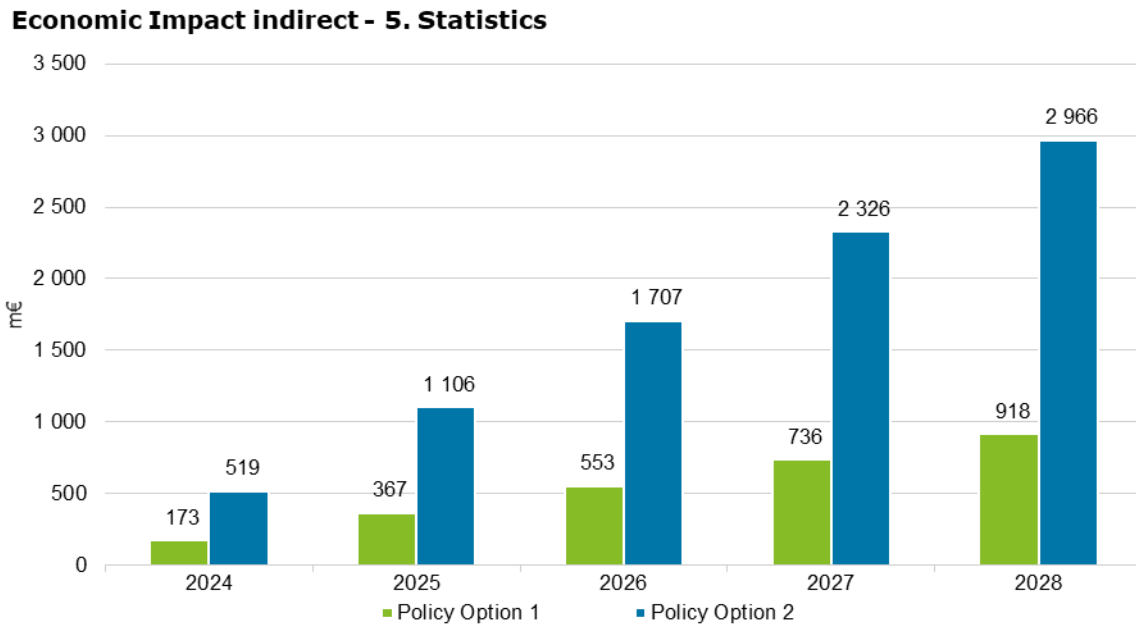
Figure 18 - Indirect economic impact - Earth observation and environment



Source: Deloitte estimation

#### 4.3.2.2.5 Statistics

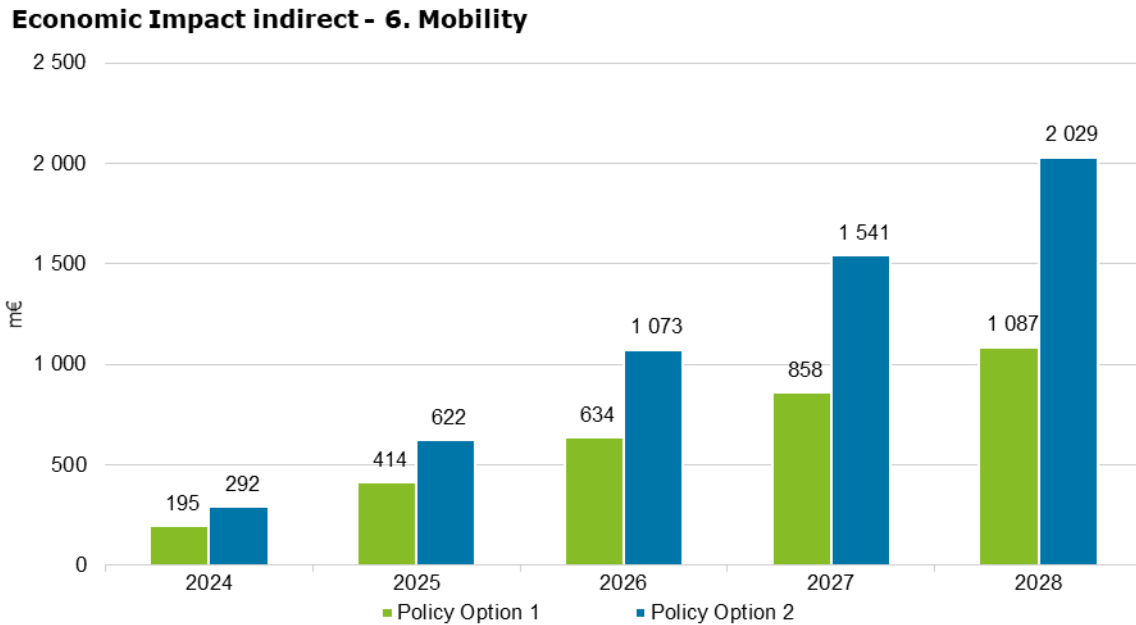
Figure 19 - Indirect economic impact - Statistics



Source: Deloitte estimation

#### 4.3.2.2.6 Mobility

Figure 20 - Indirect economic impact - Mobility



Source: Deloitte estimation

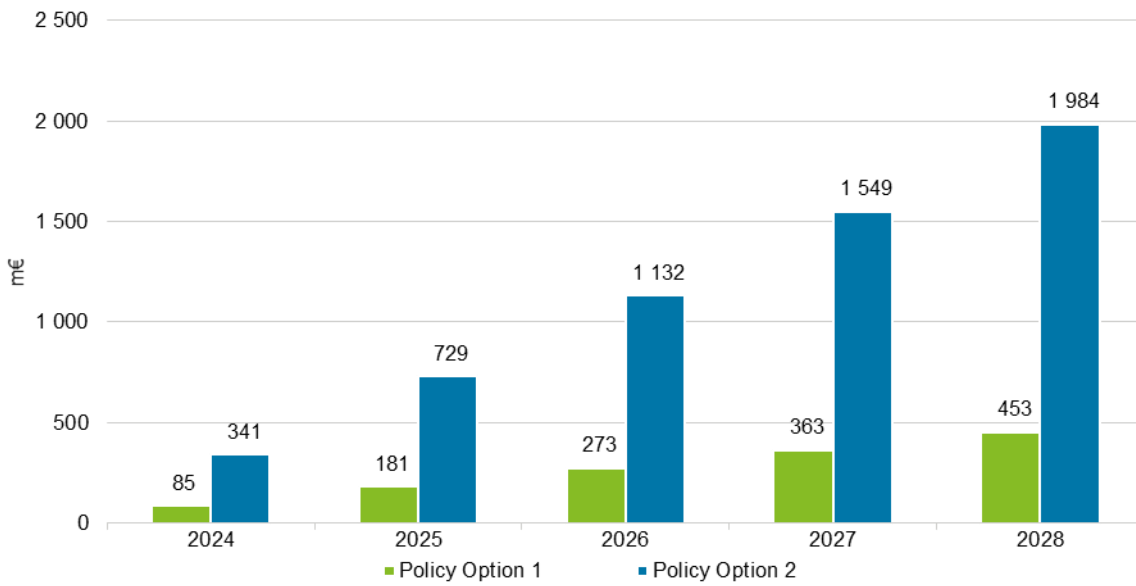
### 4.3.2.3 Total economic impacts

To estimate the total economic impact, the direct impact and indirect effects presented above are aggregated. The results are estimated on an incremental basis, calculating the difference of the economic value of the PSI/HVD industry against the baseline (per Policy Option and by thematic area). The results are presented below as total economic impact per policy option by thematic area.

#### 4.3.2.3.1 Company and company ownership

Figure 21 – Total economic impact – Company and company ownership

#### Economic Impact total - 1. Company and company ownership

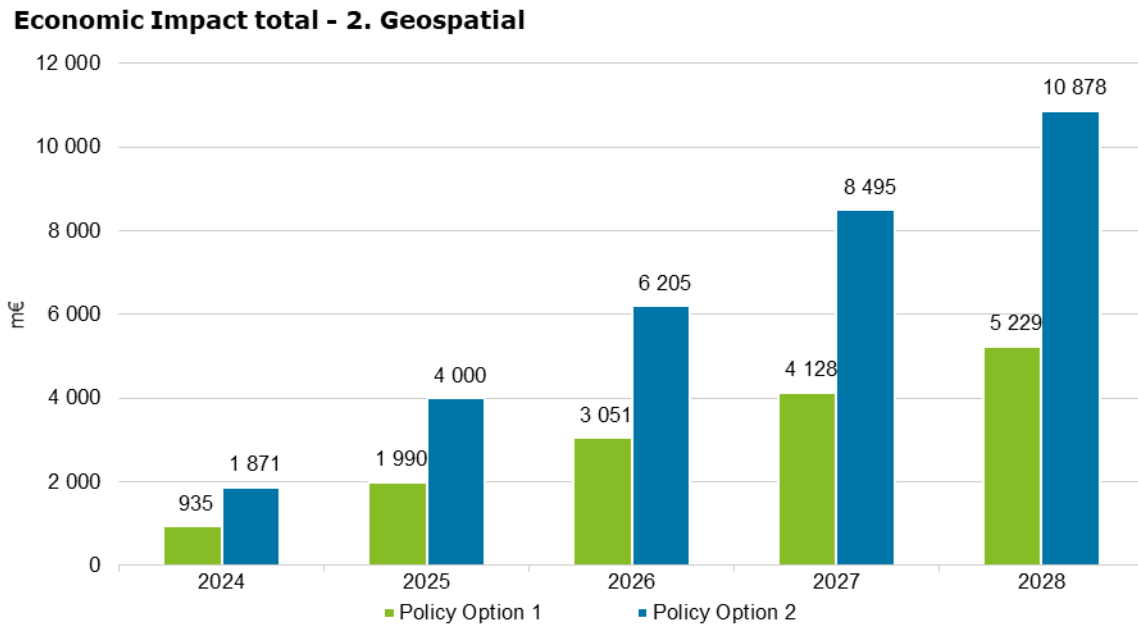


Source: Deloitte estimation



### 4.3.2.3.2 Geospatial

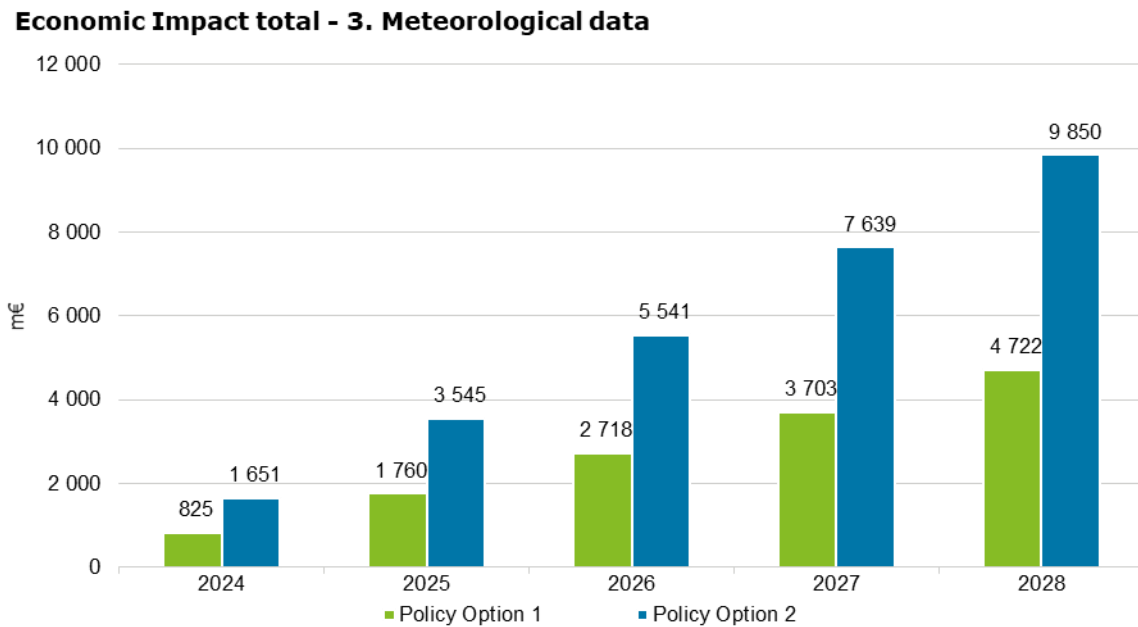
Figure 22 - Total economic impact - Geospatial



Source: Deloitte estimation

### 4.3.2.3.3 Meteorological

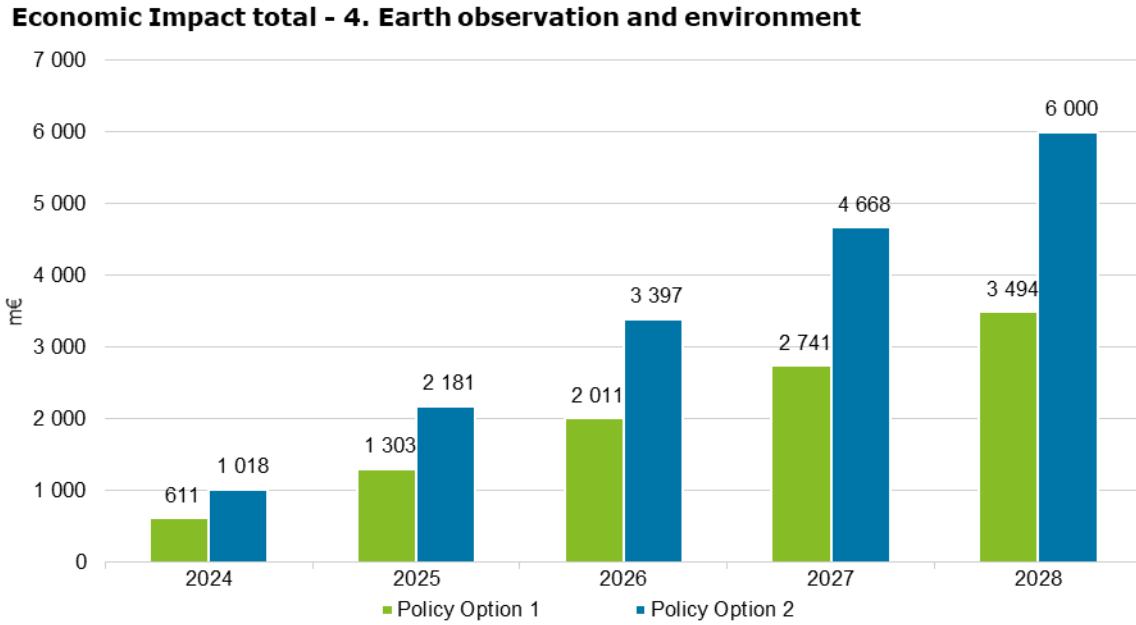
Figure 23 - Total economic impact - Meteorological Data



Source: Deloitte estimation

4.3.2.3.4 Earth observation and environment

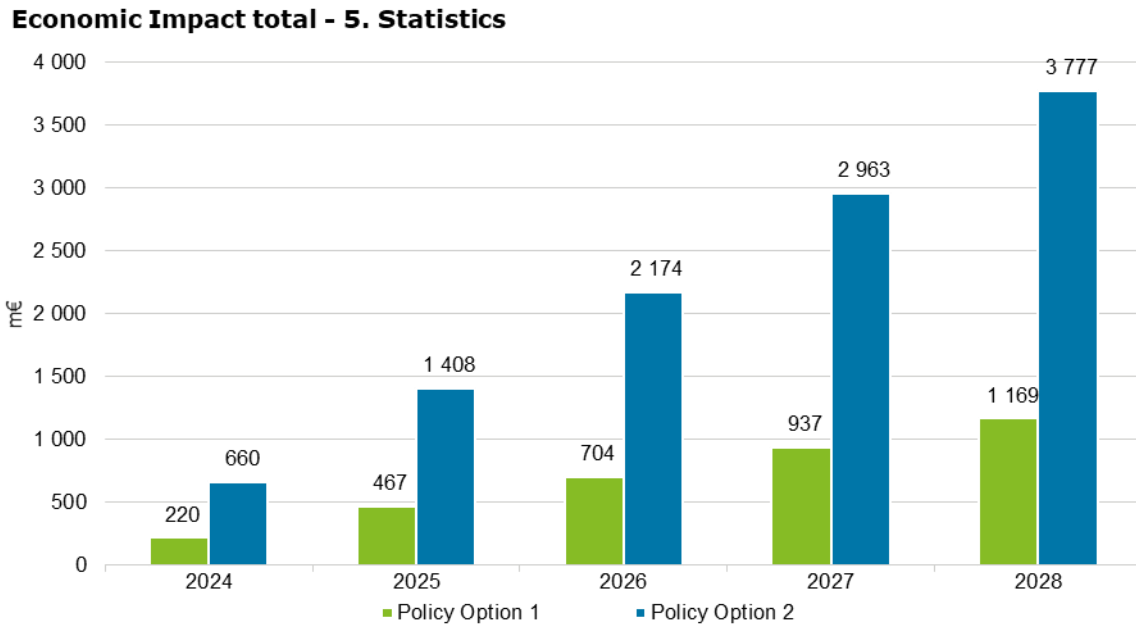
Figure 24 - Total economic impact - Earth observation and environment



Source: Deloitte estimation

4.3.2.3.5 Statistics

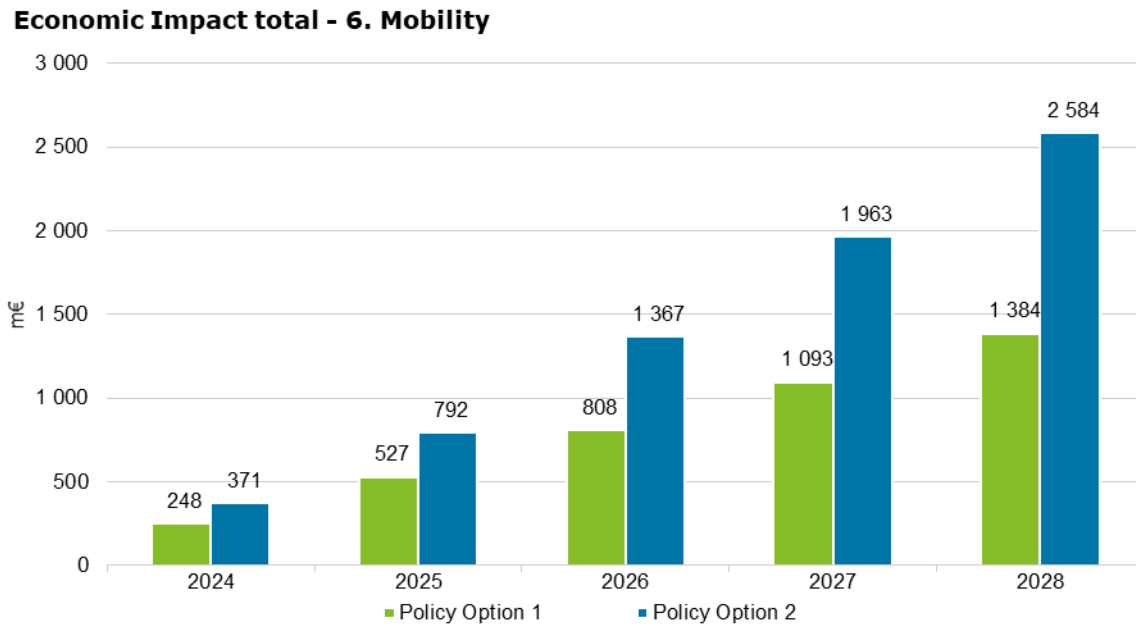
Figure 25 - Total economic impact - Statistics



Source: Deloitte estimation

### 4.3.2.3.6 Mobility

Figure 26 - Total economic impact - Mobility



Source: Deloitte estimation

### 4.3.3 Impacts of policy options on employment

Based on the macroeconomic impacts we have estimated the impact of the policy options on additional economic and socio-economic indicators. To estimate the impact on these indicators, coefficients in terms of GDP-ratios have been used based mainly on official data provided by Eurostat.

The employment indicator includes the total number of persons additionally employed, directly and indirectly including part-time and self-employed. To calculate the total number of additionally employed people, the coefficient of employment as per mEUR gross value added was determined.

The following graphs display the estimated incremental effect on the number of persons employed in the area of PSI/HVD. We have estimated the impact based on the total economic impact and an average coefficient regarding the number of persons employed. The employment coefficient has been calculated as a weighted average number of persons employed per millions of Euro Gross Value Added in the ICT sector in the EU27 2018.<sup>523</sup> The employment coefficient indicates the per-ratio increase in employment (number of persons employed) throughout the economy which result from an increase in GDP/GVA.

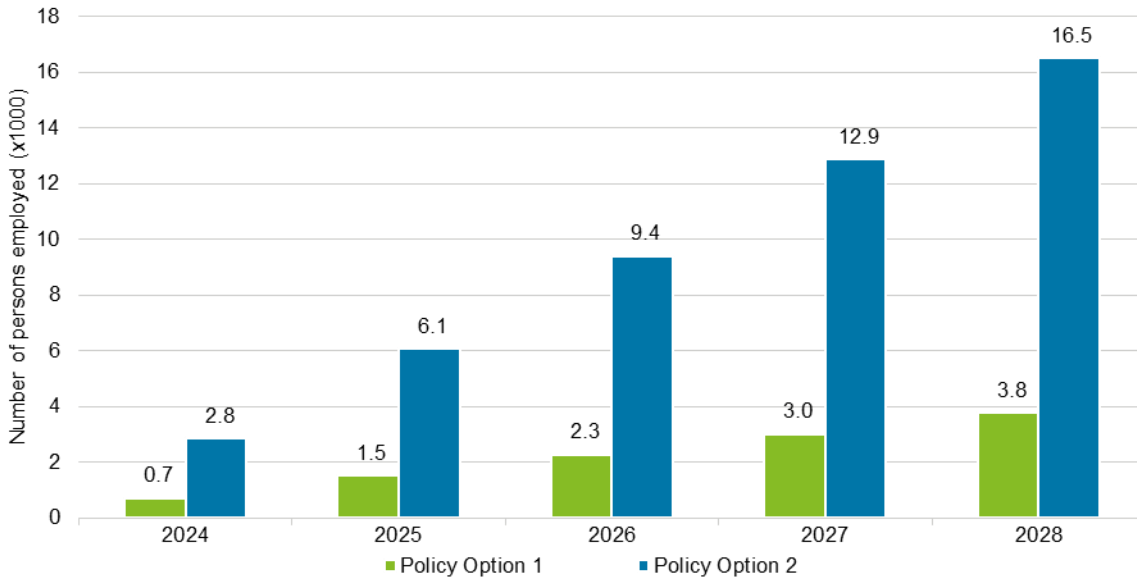
The following figures illustrate the employment impact, defined as incremental number of persons employed per policy option by thematic area.

<sup>523</sup> With regard to the indirect effects included however, the employment coefficient of the ICT sector can only serve as a proxy, as the industries will be affected. However, since the concrete breakdown of the industries affected per thematic area is unknown, the ICT coefficient serves as a proxy since the activities will be data related.

### 4.3.3.1 Company and company ownership

Figure 27 – Employment impact – Company and company ownership.

#### Employment Impact - 1. Company and company ownership

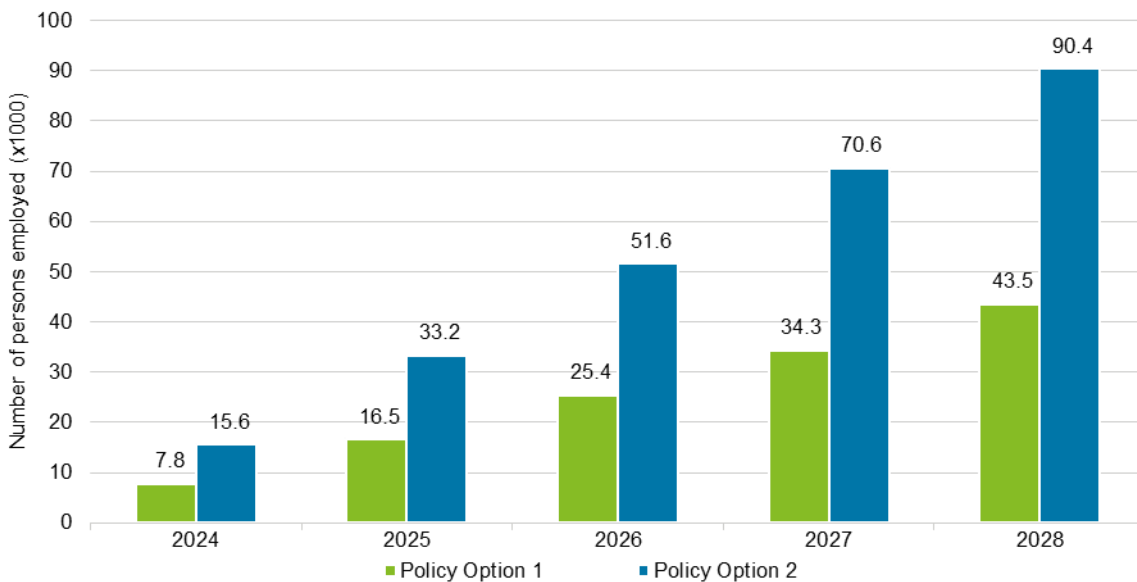


Source: Deloitte estimation

### 4.3.3.2 Geospatial

Figure 28 - Employment impact - geospatial

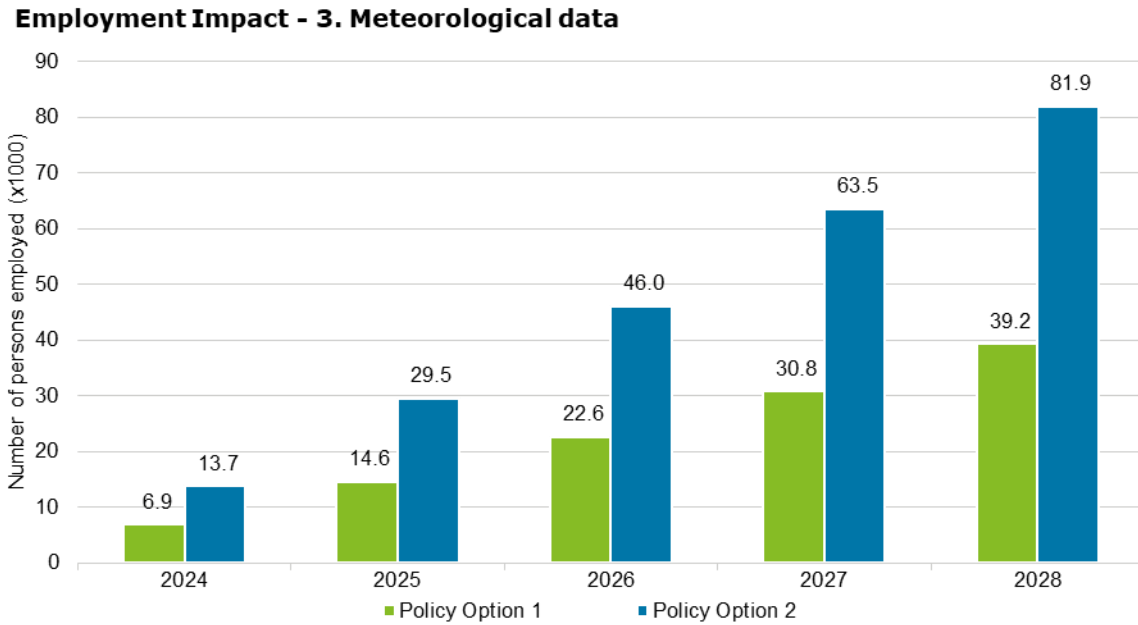
#### Employment Impact - 2. Geospatial



Source: Deloitte estimation

### 4.3.3.3 Meteorological

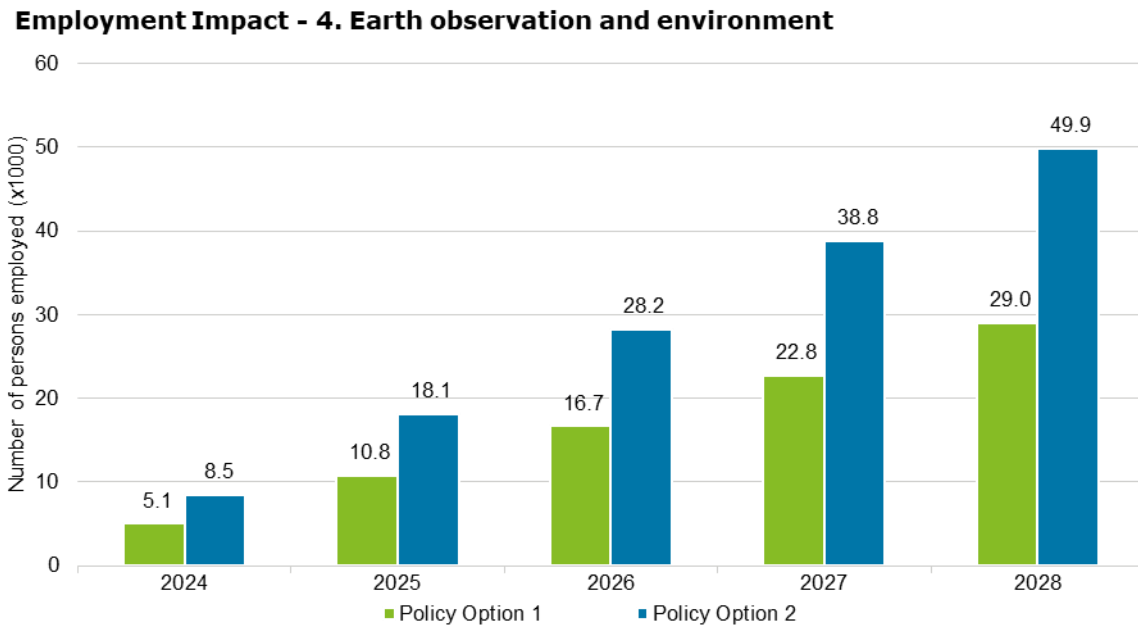
Figure 29 - Employment impact - meteorological data



Source: Deloitte estimation

### 4.3.3.4 Earth observation and environment

Figure 30 - Employment impact - earth observation and environment

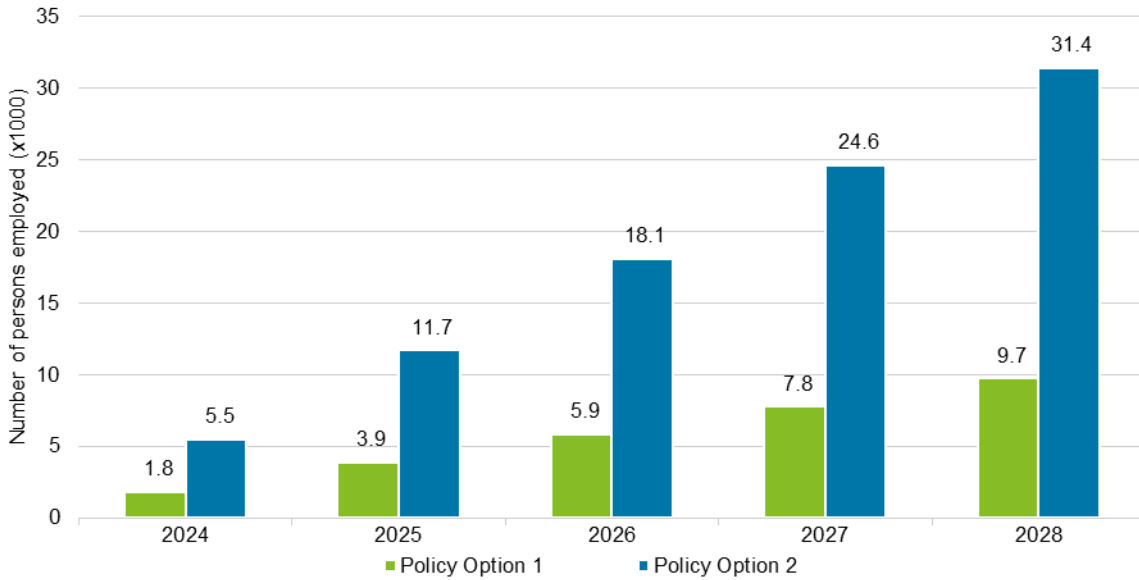


Source: Deloitte estimation

### 4.3.3.5 Statistics

Figure 31 - Employment impact - Statistics

#### Employment Impact - 5. Statistics

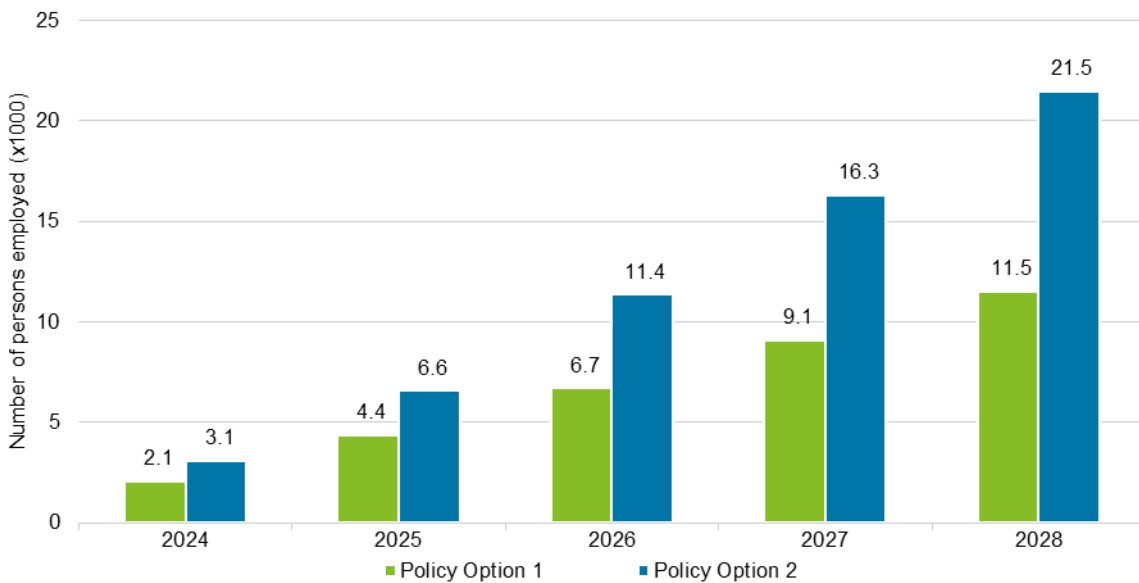


Source: Deloitte estimation

### 4.3.3.6 Mobility

Figure 32 - Employment impact - Mobility

#### Employment Impact - 6. Mobility



Source: Deloitte estimation

#### 4.3.4 Impacts of policy options on governmental revenues

In addition to the economic impact and employment impact, public administrations are expected to benefit in general from increased general revenues and taxation of the increased overall economic activities.

According to the definition of Eurostat<sup>524</sup>, the governmental revenue is the sum market output, of taxes, net social contributions, sales, other current revenues and capital transfer revenues. Total taxes are composed of taxes on production and imports, current taxes on income and wealth and capital taxes. The net social contribution is composed of actual social contributions by employers and households and the imputed social contributions, households' social contribution supplements and social insurance scheme service charges. Other current revenues consist of the categories property income earned, other subsidies on production received and current transfers. Combining these categories of governmental revenue, a weighted coefficient of EU27 by GDP is obtained. Following the calculations of Eurostat, this coefficient has and approximately value of 46% of GDP for the EU27. It should be noted, that part of this is related to governmental output, including market output, output for own final use and payments for non-market output, which could be linked to increased economic activity, but does not represent governmental inflows from taxes, social security payments or similar revenues. Hence it should be noted, that the indicator reflects the revenues for governments in general, induced by the stimulation of GDP due to the policy interventions.<sup>525</sup>

To estimate the incremental impact with regard to governmental revenues, the average total governmental revenues as % of GDP have been applied as published by Eurostat. The following figures depicts the estimated individual contributions of the thematic areas by Policy Options on the development of the additional (incremental) governmental revenues in the 27 EU Member States by thematic area.

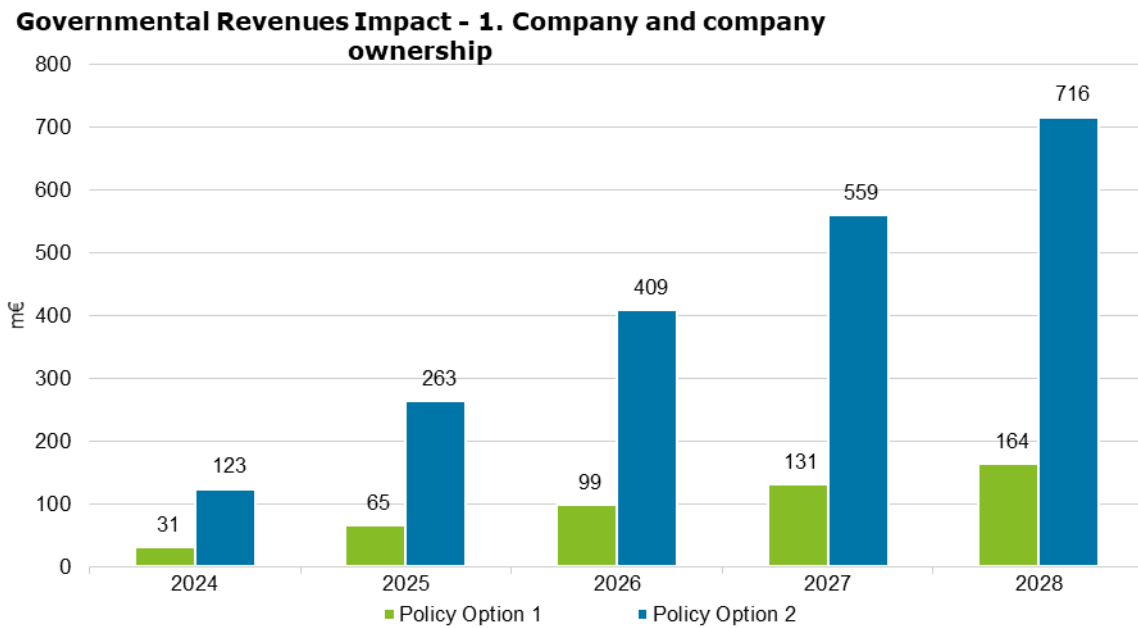
---

<sup>524</sup> Eurostat 2020, Statistics Explained, Glossary: government revenue and expenditure. [https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Government\\_revenue\\_and\\_expenditure](https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Government_revenue_and_expenditure)

<sup>525</sup> It must be noted, however, that this total governmental revenue includes – as defined in the European System of Accounts 2010 – also the market output, output for own final use and payments for non-market production. As this definition is a rather broad concept and as the macroeconomic effect of the introduction of the Policy Packages depends on a lot yet unknown factors, market output, output for own final use and payments for non-market production cannot be predicted as precisely as the other variables of governmental revenues. Excluding the categories mentioned, the adjusted governmental revenues would lower to approximately 38% of GDP according to OECD estimates. OECD, 2020, Comparative Statistics: Governmental Revenue. <https://stats.oecd.org/Index.aspx?DataSetCode=REV>

#### 4.3.4.1 Company and company ownership

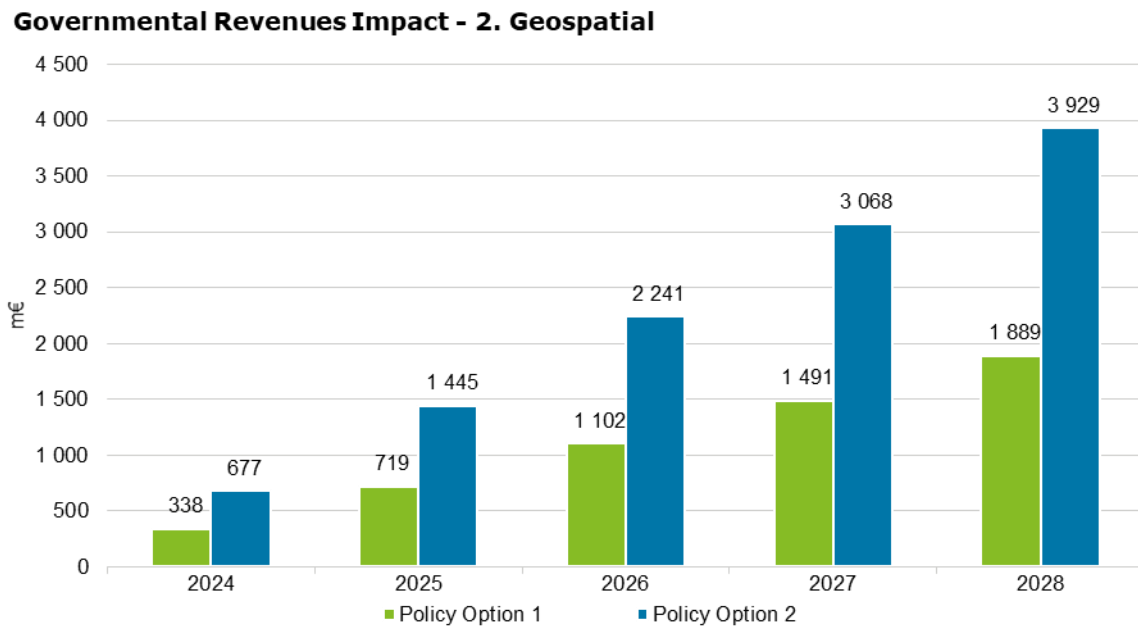
Figure 33 – Governmental revenues impact – Company and company ownership



Source: Deloitte estimation

#### 4.3.4.2 Geospatial

Figure 34 - Governmental revenues impact - Geospatial

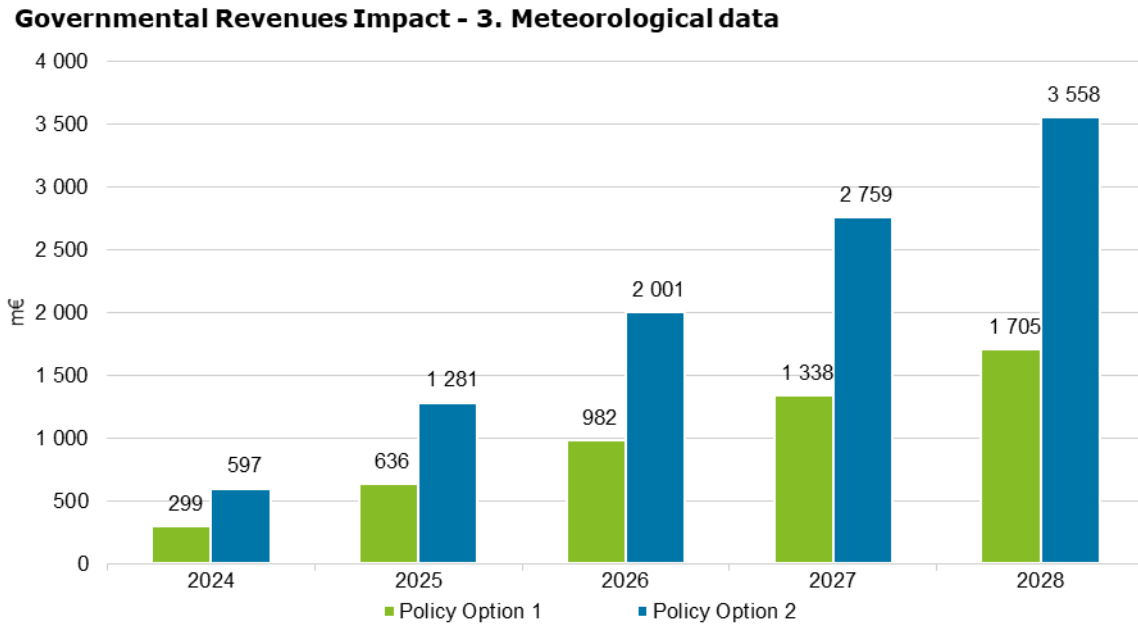


Source: Deloitte estimation



### 4.3.4.3 Meteorological

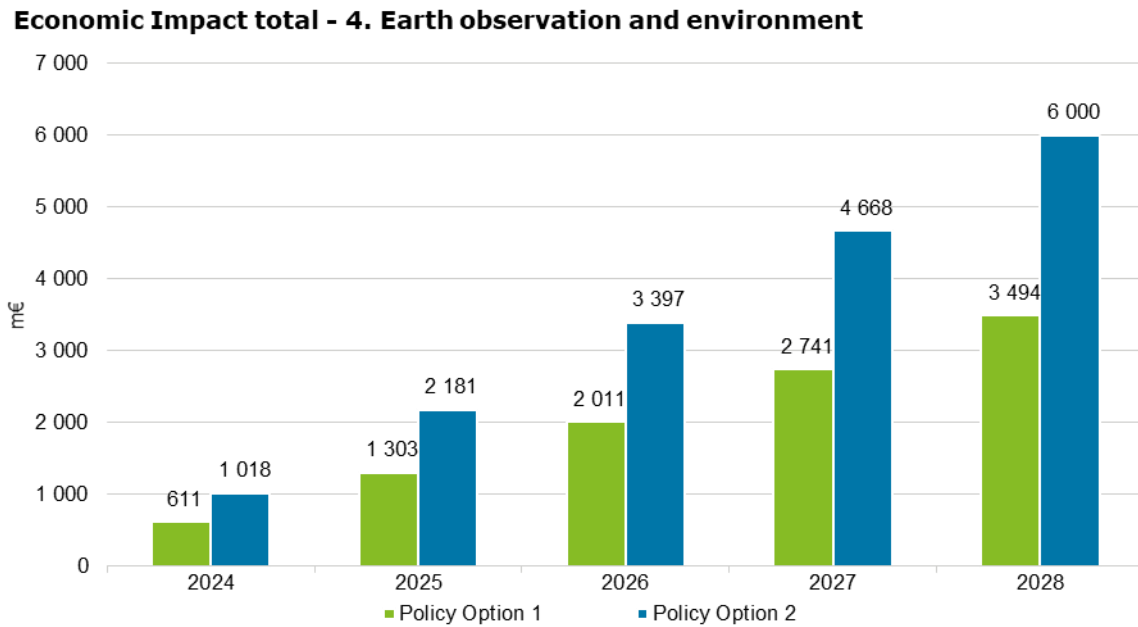
Figure 35 - Governmental revenues impact - Meteorological data



Source: Deloitte estimation

### 4.3.4.4 Earth observation and environment

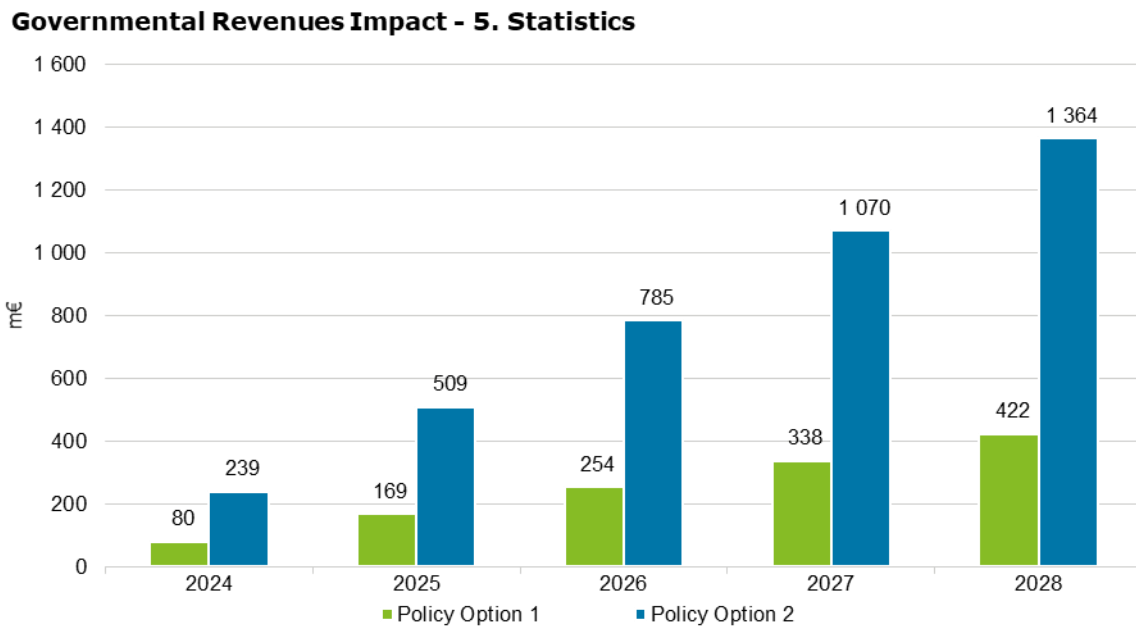
Figure 36 - Governmental revenues impact - Earth observation and environment



Source: Deloitte estimation

#### 4.3.4.5 Statistics

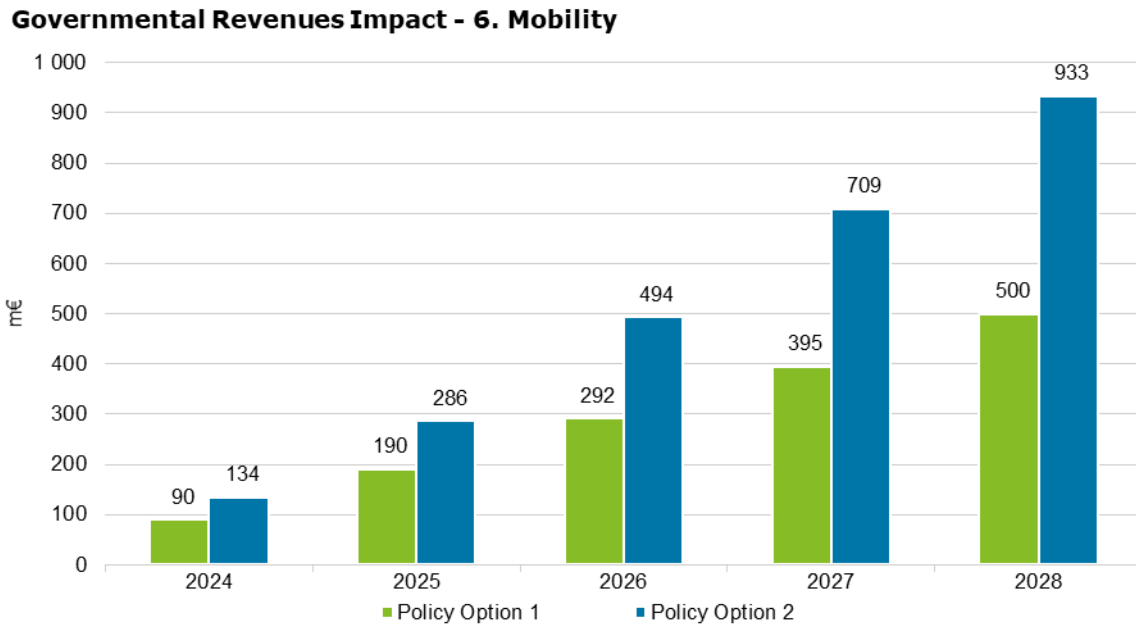
Figure 37 - Governmental revenues impact - Statistics



Source: Deloitte estimation

#### 4.3.4.6 Mobility

Figure 38 - Governmental revenues impact - Mobility



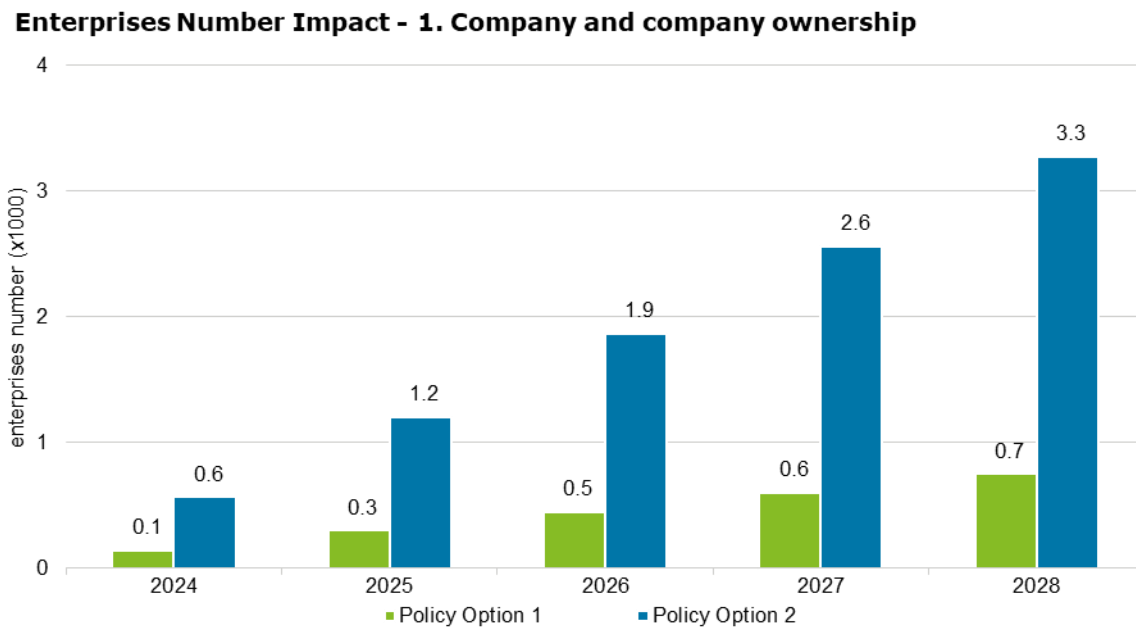
Source: Deloitte estimation

### 4.3.5 Impacts of policy options on the number of enterprises

A further indicator to be included in the impact assessment is the number of additional enterprises, which would be created statistically. This variable was calculated with a weighted coefficient for the EU27 ICT-service sector, representing the ratio of enterprises per GVA/GDP. For the weighted coefficient it was assumed that an average firm in the EU27 ICT-sector has 6 employees, respectively a statistical ratio of ca. 2 enterprises per 1 million € GVA.<sup>526</sup> However, the results should be interpreted as a proxy and represent a statistical value. It should also be noted, that a part of the impact refers to indirect impacts, which are linked to downstream activities in other industries. The economic impact can also occur in existing companies in the form of expanding their activities.

#### 4.3.5.1 Company and company ownership

Figure 39 – Enterprises number impact – Company and company ownership

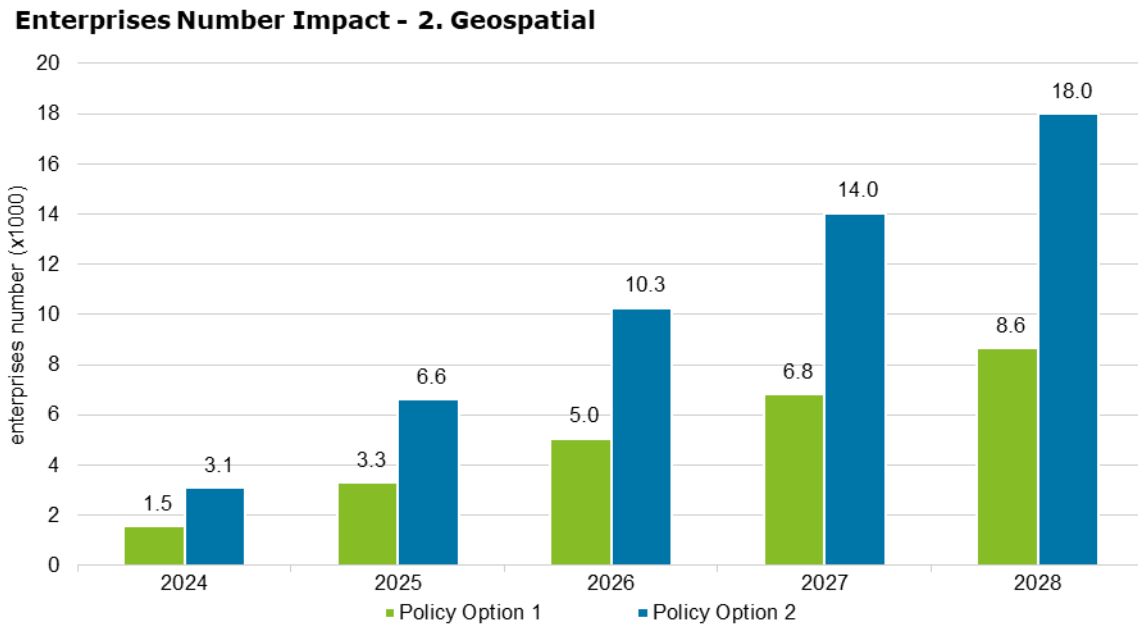


Source: Deloitte estimation

<sup>526</sup> The coefficient has been calculated as average of the years 2013 – 2017 for the total ICT-services sector in the EU27. With regard to the forecast period, the ratio should usually be adjusted, according to projected inflation. However, for the ICT industry in total, the HICP index has even been decreasing steadily in the recent years. Against this background we used a constant employment ratio for the forecast period.

### 4.3.5.2 Geospatial

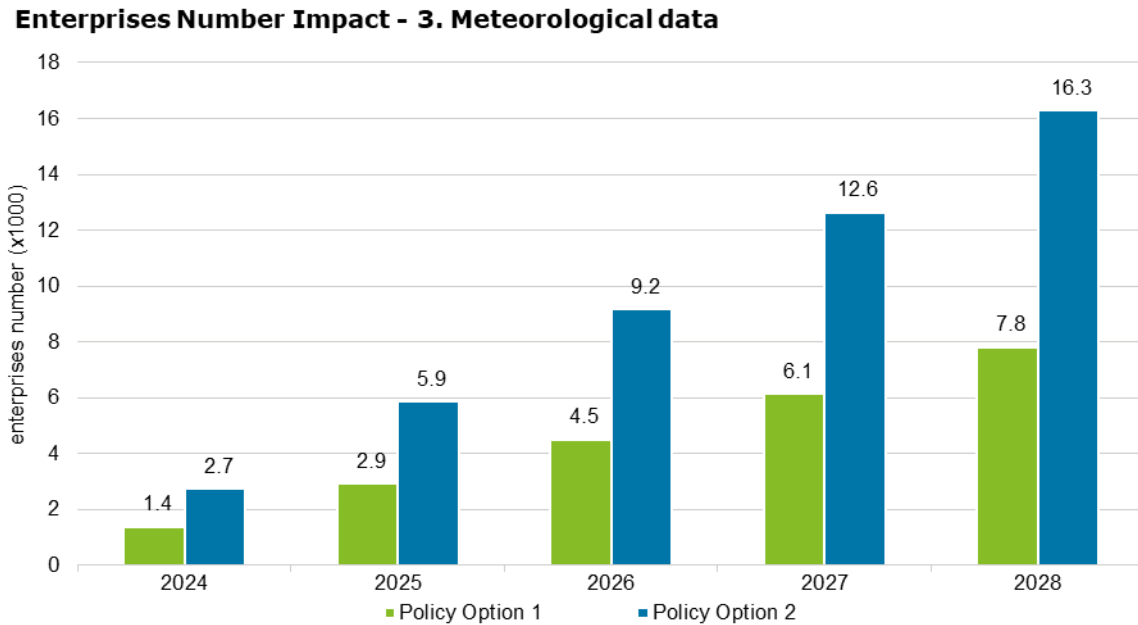
Figure 40 - Enterprises number impact - Geospatial



Source: Deloitte estimation

### 4.3.5.3 Meteorological

Figure 41 - Enterprises number impact - Meteorological data

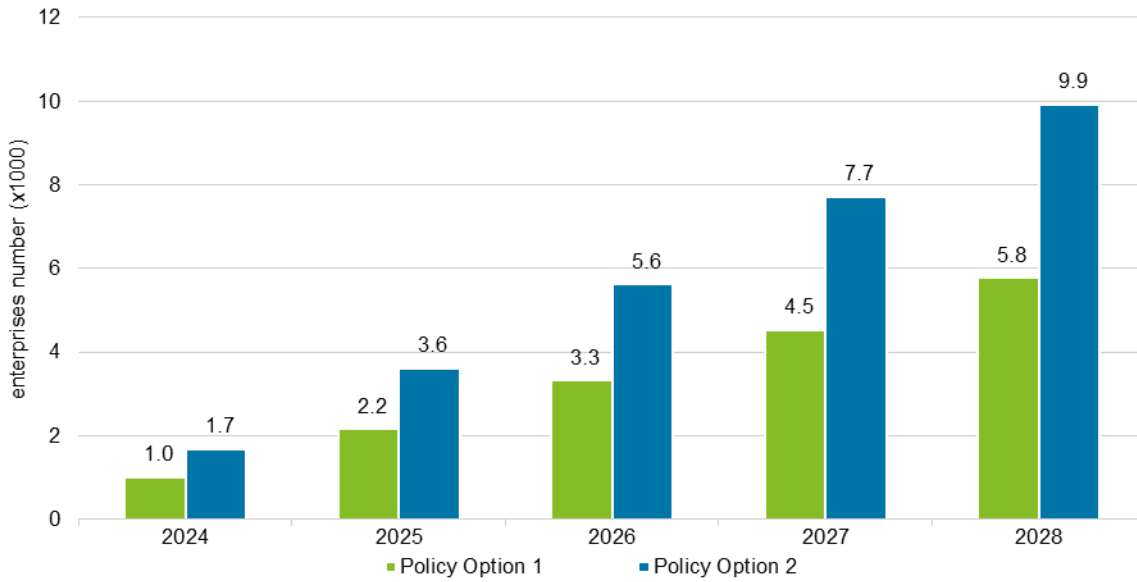


Source: Deloitte estimation

#### 4.3.5.4 Earth observation and environment

Figure 42 - Enterprises number impact - Earth observation and environment

##### Enterprises Number Impact - 4. Earth observation and environment

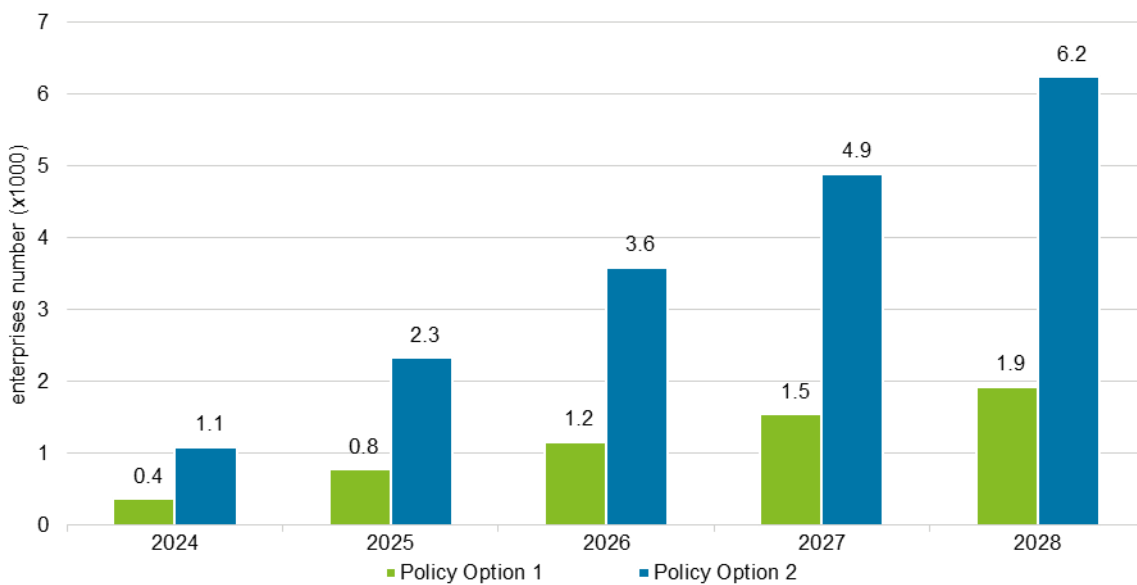


Source: Deloitte estimation

#### 4.3.5.5 Statistics

Figure 43 - Enterprises number impact - Statistics

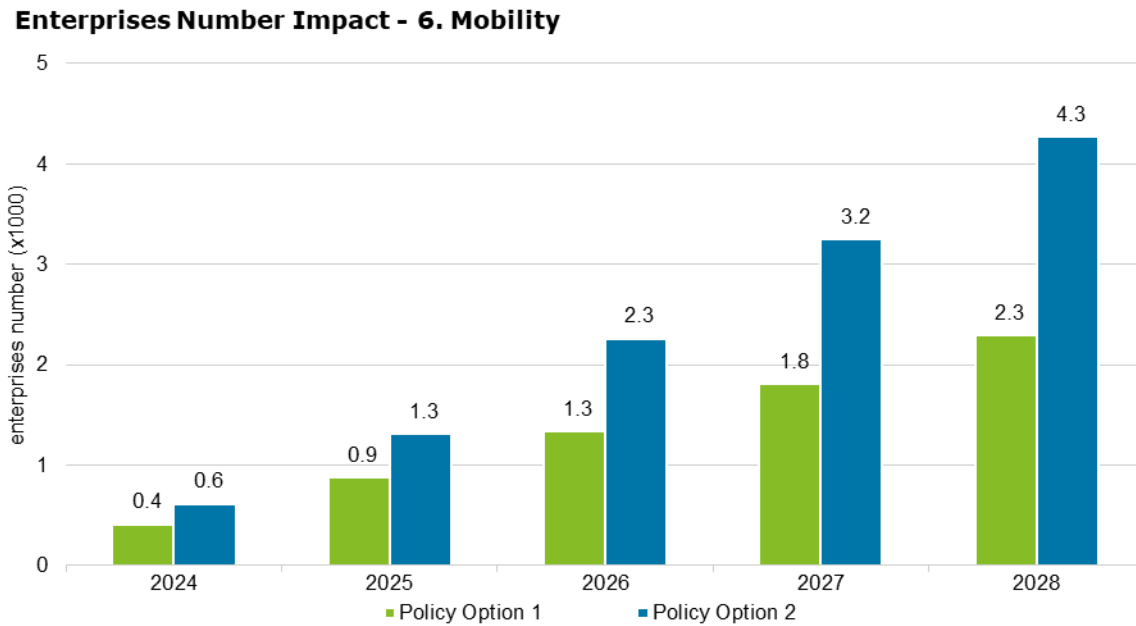
##### Enterprises Number Impact - 5. Statistics



Source: Deloitte estimation

### 4.3.5.6 Mobility

Figure 44 - Enterprises number impact - Mobility



Source: Deloitte estimation

## 4.4 Identification of policy intervention packages and their aggregated impacts

This section draws conclusions of the previous impact assessment and outlines different policy packages based on the MCA and the macroeconomic assessment.

### 4.4.1 Possible policy intervention packages

The assessment of impacts on the overall economy and society can only be performed at an aggregated level, by creating policy intervention packages composed of one policy option per thematic area. Based on the multi-criteria analysis performed under chapter 4.1.2, taking into account the interdependences between the policy options three policy packages were identified:

- **Policy package 1 consists of a lower intensity intervention:** this option would entail that a few datasets considered in scope of the six thematic areas, or only some data points from a given dataset, are included as HVD. For example in the case of company and company's ownership, this package could imply to only include as HVD those already made available by the Member States.
- **Policy package 2 contains a higher intensity intervention:** as indicated in its name, this policy package is a far-reaching one. Still considering the company and company's ownership example, this policy package would entail that all datasets presented in this report under this thematic area are part of the PSI Directive as HVD.
- **Policy package 3 is a mixed intervention:** based on the results drawn in the chapter 4.1.2 (Results of the MCA and conclusions), where a multi-criteria-analysis (MCA) was conducted in order to determine for each thematic area the respective preferred option, a mixed intervention is the intervention where the most favourable policy option as a result of the MCA for each areas is selected. This yielded in a higher

intervention for the thematic areas meteorological data and statistics and a lower intervention for the other areas.

**Table 72 – Policy Packages composition**

	Company & Company Ownership	Geo-spatial	Meteoro-logical Data	Earth observation & environment	Statistics	Mobility
<b>Policy Package 1</b>	Lower intervention	PO1	PO1	PO1	PO1	PO1
<b>Policy Package 2</b>	Higher intervention	PO2	PO2	PO2	PO2	PO2
<b>Policy Package 3</b>	Mixed intervention	PO1	PO1	PO2	PO2	PO1

Source: Deloitte.

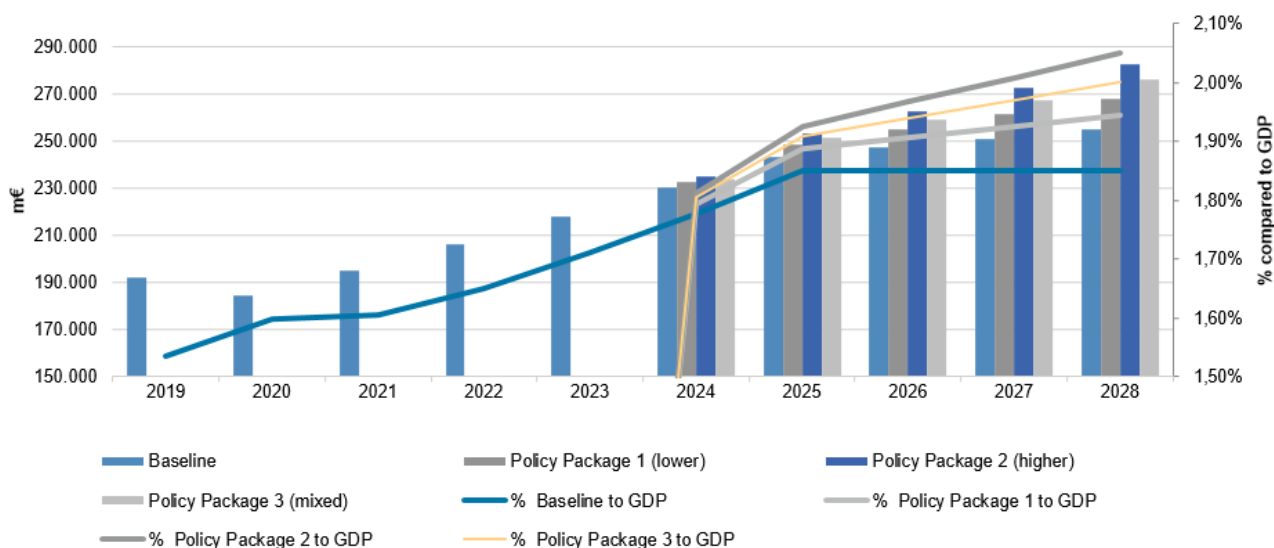
#### 4.4.2 Impacts of the policy intervention packages at macro-level

Finally, the economic impact is assessed on an aggregated level for the six thematic areas under consideration. In this regard, policy packages can be formulated. As a simple approach, all Policy Options being lower or higher, could be bundled to form a lower or a higher intervention policy package. In addition, based on the results of the Multi-Criteria Analysis, the mixed policy package can be assessed.

Our assessment is that all policy packages (lower, higher, mixed intervention) are creating positive impacts at the macroeconomic level by positively impacting the value of the economy for the EU27 from a projected autonomous growth of the direct and indirect impacts from an estimated 184 bEUR in 2020 to a projected 255 bEUR in 2028 (1.60% to 1.85% of GDP).

Figure 45 – Economic Value of PSI/HVD per policy Package in m€ and compared to EU27 GDP

#### Impact on the Economic Value of the PSI/HVD Policy Packages compared to GDP



Source: Deloitte estimation

The policy package 2 (higher intervention Policy Option in all six thematic areas) creates the most impact on the total economic value of the economy. The value of the PSI/HVD related economy is estimated to grow to bEUR 282 in 2028, compared to bEUR 255 in the baseline scenario. Compared to GDP, the ratio expected to increase from 1.85 % to 2.05 % in the year 2028.

Table 73 - Economic Value of PSI/HVD per policy package in mEUR and compared to EU27 GDP

#### Economic Value/GDP [m€]

	2024	2025	2026	2027	2028
Baseline	230.417	243.632	247.389	251.262	255.239
% <b>Baseline to GDP</b>	<b>1,78%</b>	<b>1,85%</b>	<b>1,85%</b>	<b>1,85%</b>	<b>1,85%</b>
Policy Package 1 (lower) <span style="float:right">✔</span>	232.716 <span style="float:right">✔</span>	248.523 <span style="float:right">✔</span>	254.899 <span style="float:right">✔</span>	261.442 <span style="float:right">✔</span>	268.157 <span style="float:right">✔</span>
% <b>Policy Package 1 to GDP</b>	<b>1,79%</b>	<b>1,89%</b>	<b>1,91%</b>	<b>1,93%</b>	<b>1,94%</b>
Policy Package 2 (higher) <span style="float:right">✔</span>	235.066 <span style="float:right">✔</span>	253.570 <span style="float:right">✔</span>	262.948 <span style="float:right">✔</span>	272.680 <span style="float:right">✔</span>	282.778 <span style="float:right">✔</span>
% <b>Policy Package 2 to GDP</b>	<b>1,81%</b>	<b>1,93%</b>	<b>1,97%</b>	<b>2,01%</b>	<b>2,05%</b>
Policy Package 3 (mixed) <span style="float:right">✔</span>	234.032 <span style="float:right">✔</span>	251.353 <span style="float:right">✔</span>	259.358 <span style="float:right">✔</span>	267.637 <span style="float:right">✔</span>	276.199 <span style="float:right">✔</span>
% <b>Policy Package 3 to GDP</b>	<b>1,80%</b>	<b>1,91%</b>	<b>1,94%</b>	<b>1,97%</b>	<b>2,00%</b>

Source: Deloitte estimation

This result is logical, as with increased scope, increased re-use is expected. It will affect more stakeholders which further create more economic value. However, the efficiency is slightly lower with this policy package as higher costs are expected.

**A solution could be to implement a mixed policy package.** For the mixed policy package 3, which is in accordance with the results of the multi-criteria analysis in chapter 4.1.2, the value of the PSI/HVD related economy is estimated to grow to bEUR 276 in 2028. The ratio to GDP in 2028 is expected to increase to 2.00 % compared to 1.85 % in the baseline scenario.

The total value of the economy in 2028 with the mixed policy package is bEUR 276 represents an incremental impact of bEUR 21 in comparison to the baseline. This value exceeds the effect of implementing the lower policy package 1 (bEUR 268) by bEUR 8.

The total effect is composed of a direct and an indirect forward effect. In 2028, the indirect forward effect of implementing the mixed policy package is bEUR 201. This signifies an increase of bEUR 16 in comparison to the baseline (bEUR 185). The direct effect of the mixed policy package is bEUR 75, representing an incremental impact of bEUR 6 as compared to the baseline in 2028.

In addition to the analyses above, which estimates the economic value of the PSI related economy and its relation to GDP for the three policy packages, the incremental impact can be analysed. The incremental impact on the following indicators is estimated:

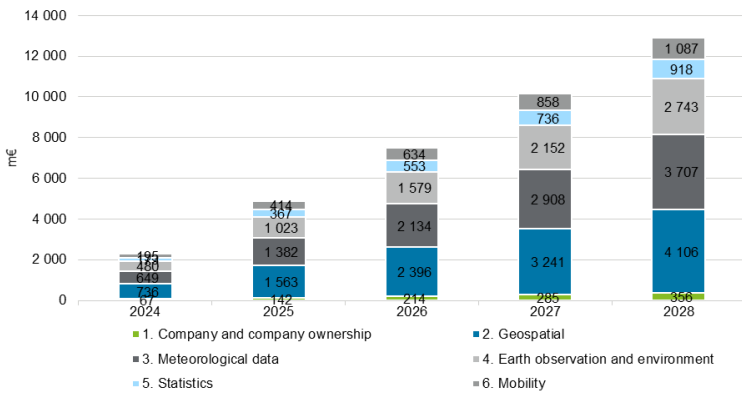
- Economic value (incremental GDP contribution),
- Employment (incremental number of persons employed including part-time),
- Governmental revenues (incremental statistical percentage of GDP),
- Number of enterprises (additional enterprises created statistically).

The following figures illustrate the incremental impact, calculated as the delta to the baseline.

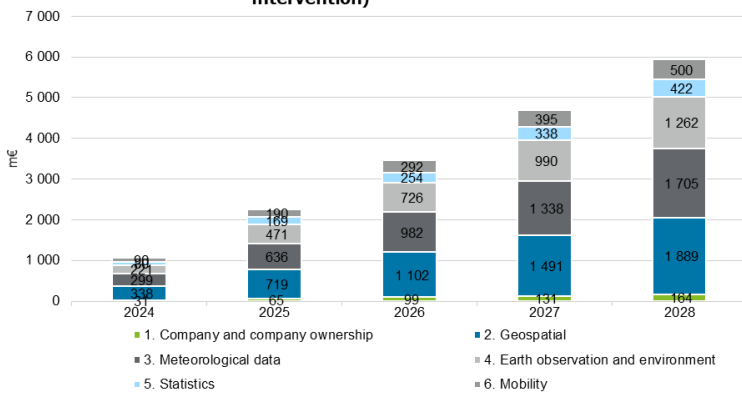


Figure 46 – Incremental impacts policy package 1 (lower intensity intervention)

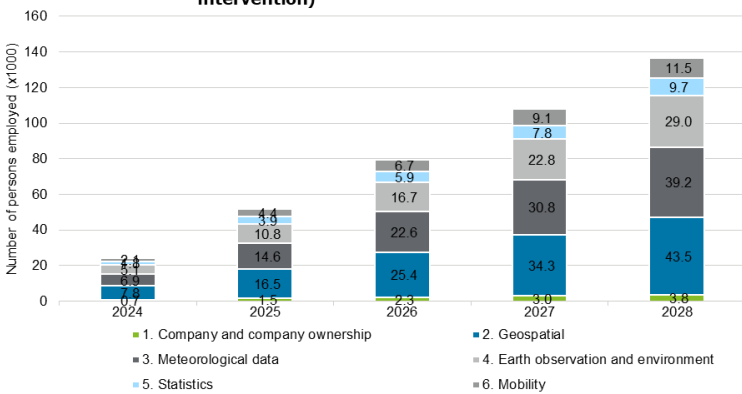
**Economic Impact - Policy Package 1 (lower intensity intervention)**



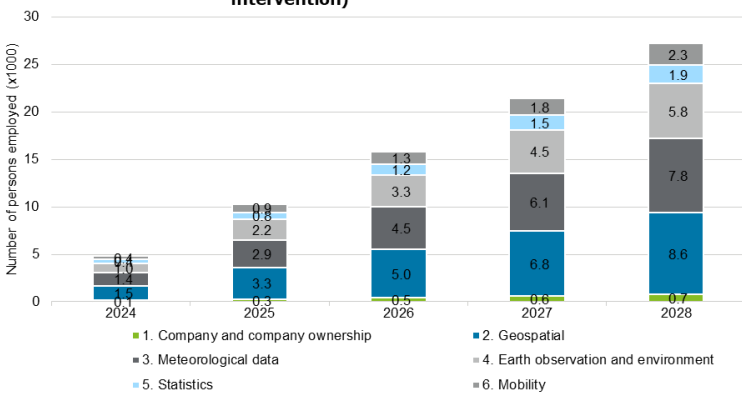
**Governmental Revenues Impact - Policy Package 1 (lower intensity intervention)**



**Employment Impact - Policy Package 1 (lower intensity intervention)**



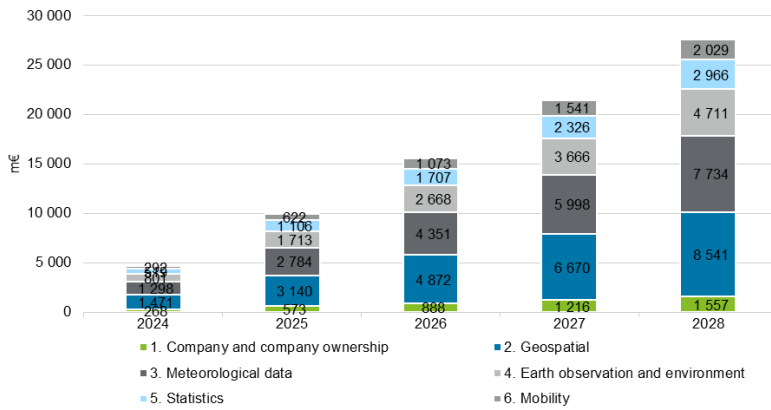
**Enterprises Number Impact - Policy Package 1 (lower intensity intervention)**



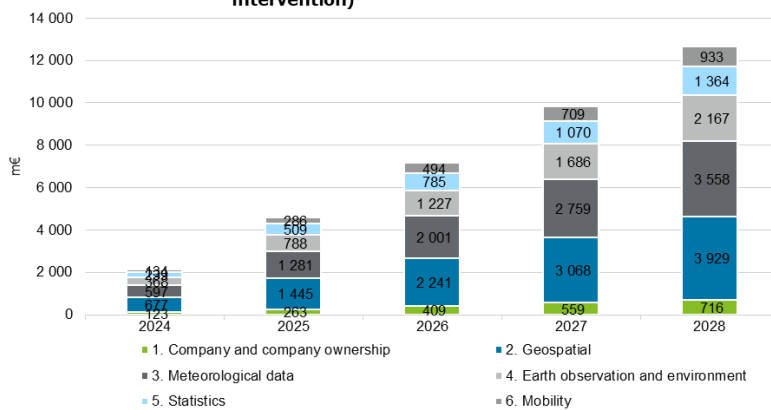
Source: Deloitte estimation

Figure 47 – Incremental impacts policy package 2 (higher intensity intervention)

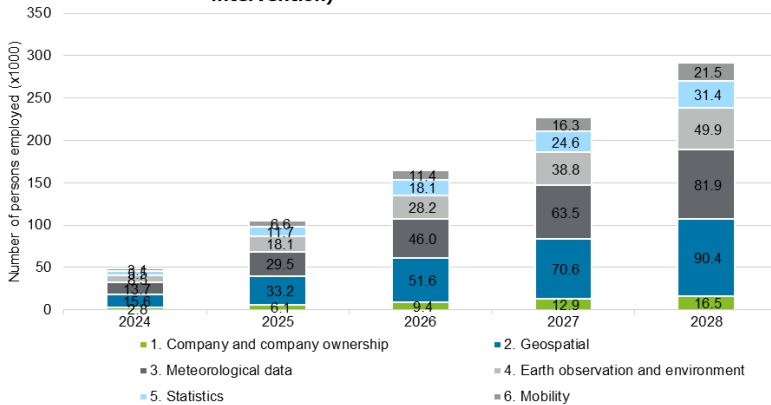
**Economic Impact - Policy Package 2 (higher intensity intervention)**



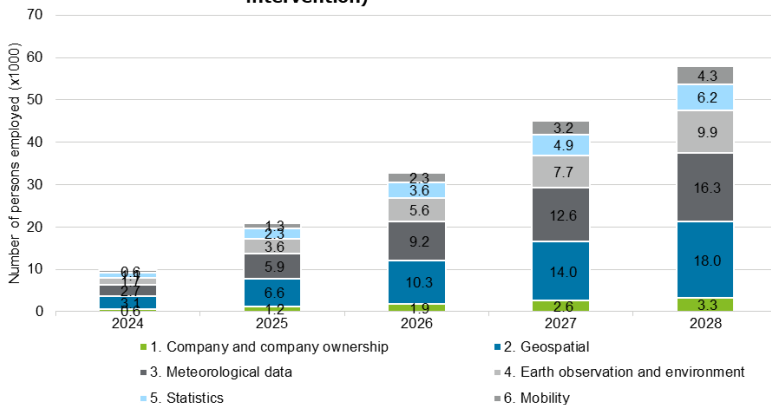
**Governmental Revenues - Policy Package 2 (higher intensity intervention)**



**Employment Impact - Policy Package 2 (higher intensity intervention)**



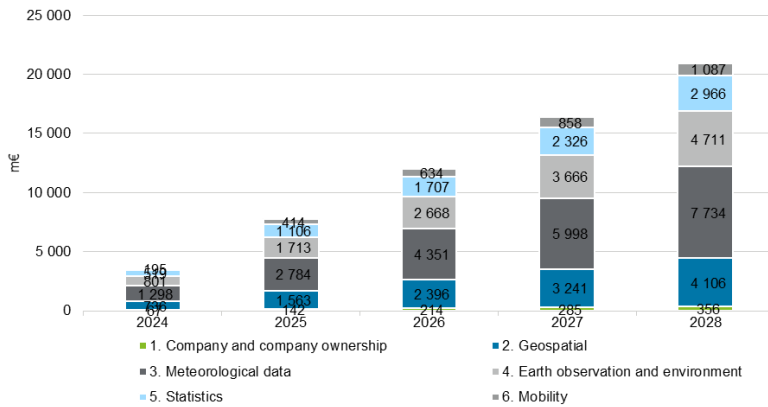
**Enterprises Number Impact - Policy Package 2 (higher intensity intervention)**



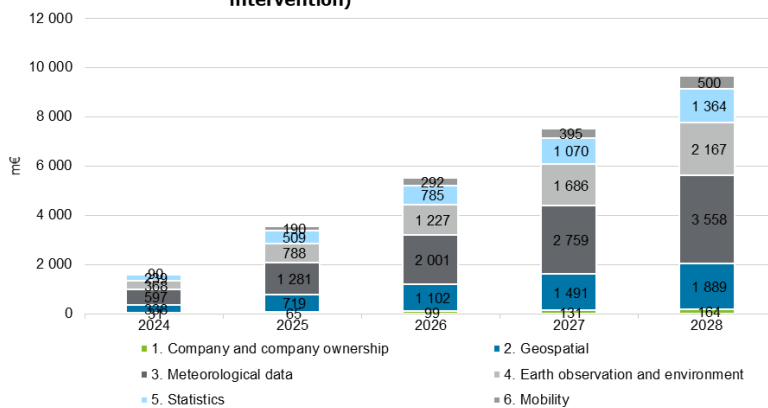
Source: Deloitte estimation

Figure 48 – Incremental impacts policy package 3 (mixed intensity intervention)

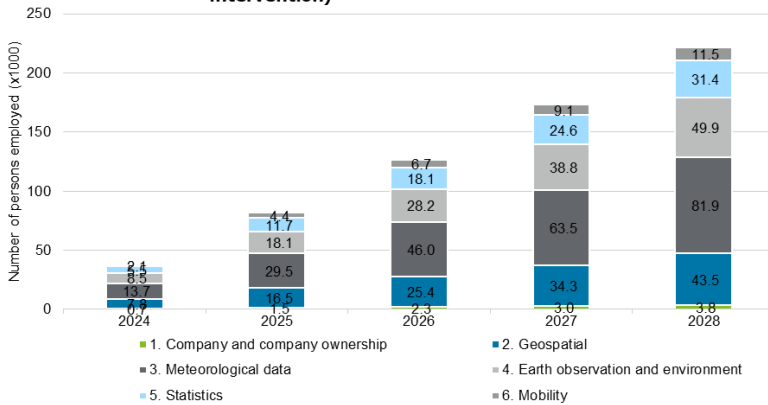
**Economic Impact - Policy Package 3 (mixed intensity intervention)**



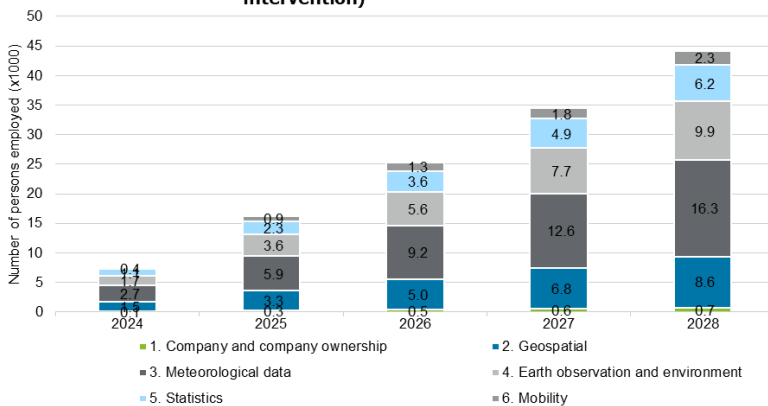
**Governmental Revenues - Policy Package 3 (mixed intensity intervention)**



**Employment Impact - Policy Package 3 (mixed intensity intervention)**



**Enterprises Number Impact - Policy Package 3 (mixed intensity intervention)**

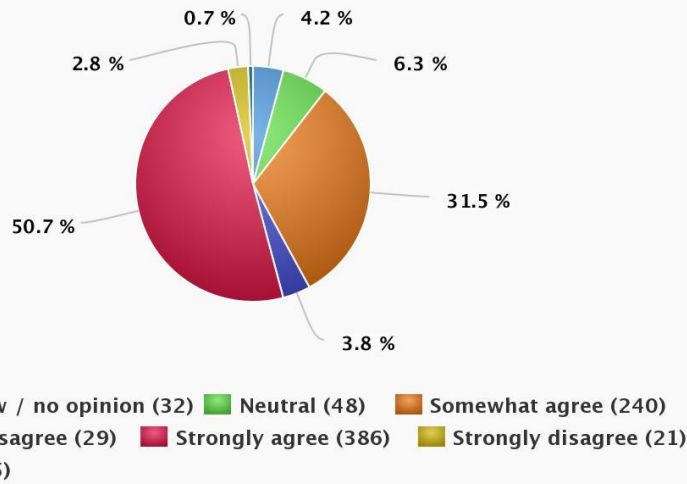


Source: Deloitte estimation

# Annexes

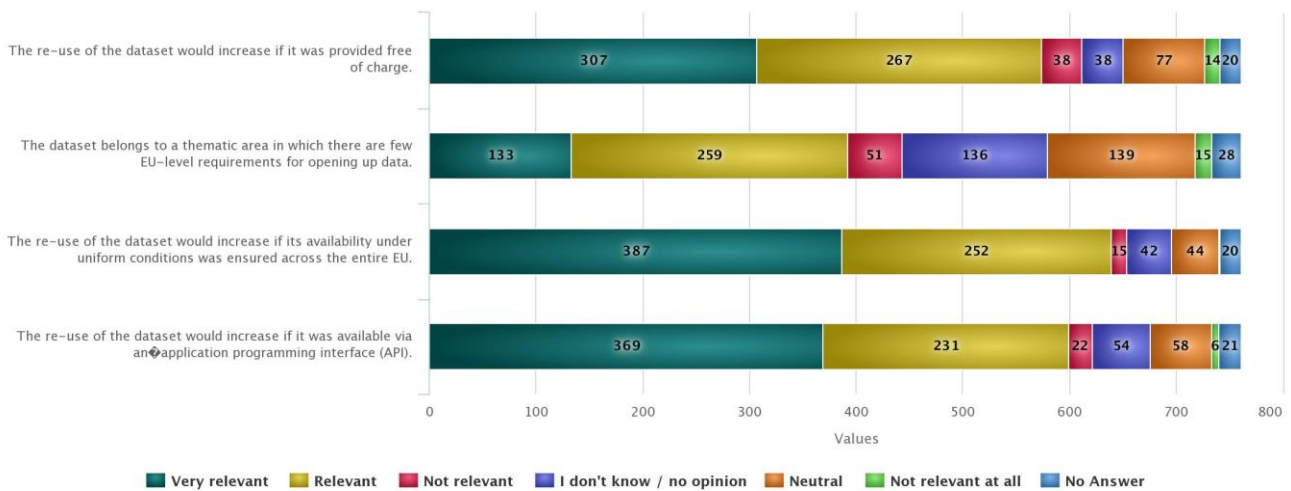
## Annex A – Used graphs for the analysis of the Open Public Consultation

'The establishment of a list of high-value datasets, to be made available free of charge, without restrictions and via APIs, is a good way to ensure that public sector data has a positive impact on the EU's economy and society.' To what extent do you agree with this statement?



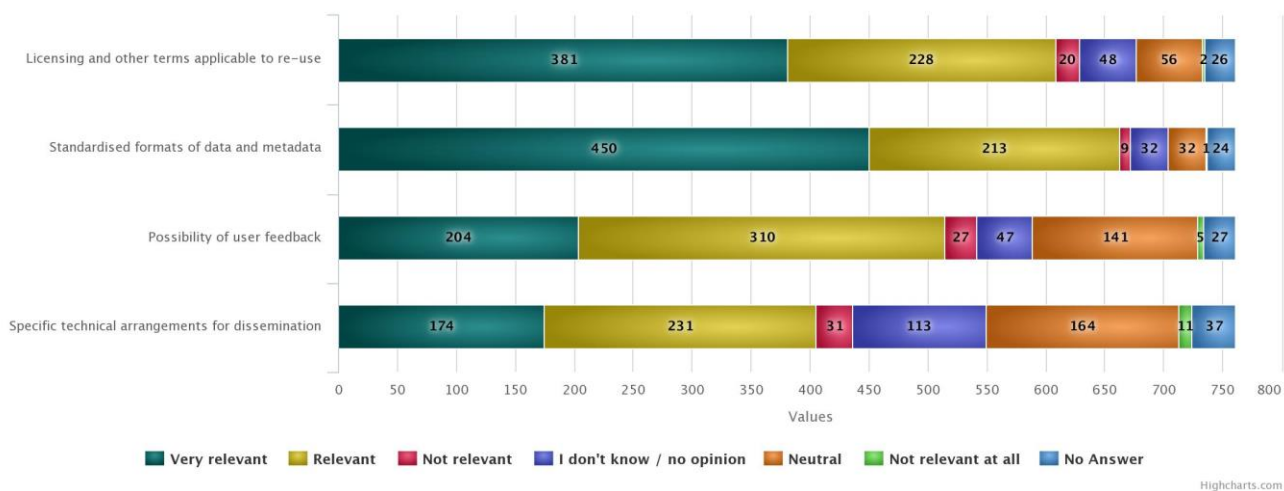
Highcharts.com

Apart from the potential to generate socio-economic benefits, please indicate the relevance of the following additional factors to be taken into account when selecting datasets for the future list of high value datasets:

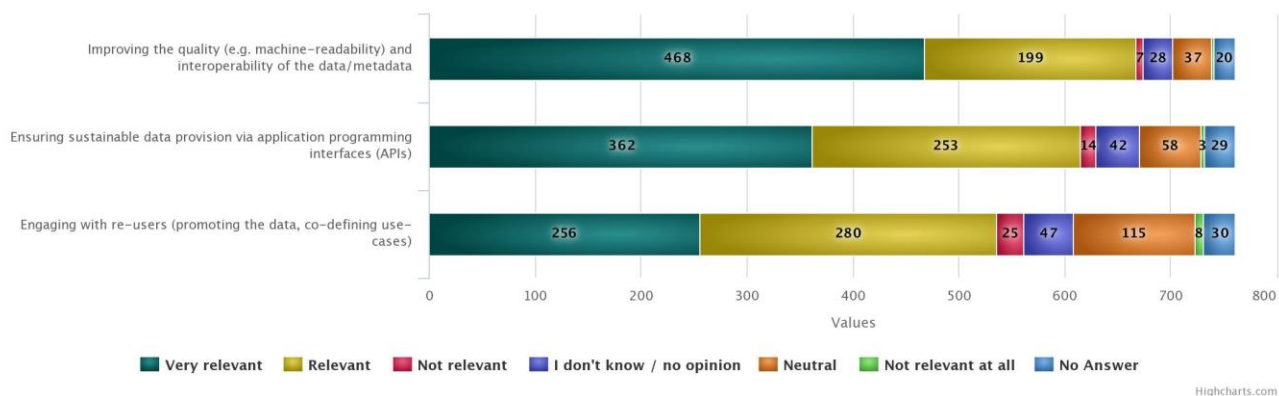


Highcharts.com

Under the Open Data Directive, specific high-value datasets will have to be available free of charge, in a machine-readable format, provided via APIs and, where relevant, provided as a bulk download. Please indicate the relevance of each of the other arrangements indicated below to improve the re-usability of specific high-value datasets.



EU programmes may provide funding to enhance the availability and re-use of high-value datasets across Europe. For each of the following activities, please indicate how relevant it is to support them.







## GETTING IN TOUCH WITH THE EU

### In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: [https://europa.eu/european-union/contact\\_en](https://europa.eu/european-union/contact_en)

### On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

– by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

– at the following standard number: +32 22999696 or

– by email via: [https://europa.eu/european-union/contact\\_en](https://europa.eu/european-union/contact_en)

## FINDING INFORMATION ABOUT THE EU

### Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: [https://europa.eu/european-union/index\\_en](https://europa.eu/european-union/index_en)

### EU publications

You can download or order free and priced EU publications at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see [https://europa.eu/european-union/contact\\_en](https://europa.eu/european-union/contact_en)).

### EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

### Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.





Publications Office

doi: 10.2759/493091  
ISBN 978-92-76-25267-2