17th Meeting of the Data Quality Working Group

Report on the Cross check of DQ chapters of S-1xx PSs

Agenda Item 4.1A

DQWG-17, VTC Event, 8 – 10 Feb 2022

1. INTRODUCTION/OVERVIEW

IHO

International Hydrographic Organization

According to the action DQWG 16/04,
the subWG comprised of Vice Chair, NL, SE and UNH has implemented the Cross check of DQ chapters of 8 published S-1xx PSs .

IHO S-100 based Product Specifications (S-1XX):		
S-101	ENC Product Specification (Edition 1.0.0, December 2018)	Link
S-102	Bathymetric Surface Product Specification (Edition 2.0.0, October 2019)	Link
S-111	Surface Currents Product Specification (Edition 1.0.0, December 2018)	Link
S-121	Maritime Limits and Boundaries Product Specification (Edition 1.0.0, October 2019)	Link
S-122	Marine Protected Areas (Edition 1.0.0, January 2019)	Link
S-123	Marine Radio Services (Edition 1.0.0, January 2019)	Link
S-127	Marine Traffic Management (Edition 1.0.0, December 2018)	Link
S-129	Under Keel Clearance Management (Edition 1.0.0, June 2019)	Link

IHO

International Hydrographic Organization

2. PRINCIPLES OF THE CROSS CHECK OF DQ CHAPTERS

The Cross check of DQ chapters follow principles as below:

a) The Cross check shall be carried out in accordance with S-97 Ed 1.1.0 - “IHO Guidelines for Creating S-100 Product Specifications”。
b) Only the DQ chapters of S-1xx PSs shall be checked and the results have been presented in the form of cross check matrix.
c) DQ elements included in the other parts of each S-1xx PS (such as S-111) shall be described in the form of Notes under its cross check matrix.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.1 Identify the DQ requirements for S-1xx PS in S-97

10 recommendations for the development of S-1xx based Product Specifications.

10 Recommendations
1.Completeness
2.Conceptual consistency
3. Domain consistency
4. Format consistency
5. Topological consistency
6. Positional Accuracy
7. Thematic Accuracy
8. Temporal Quality
9. Aggregation
10. Introduction to DQ Paragraph

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

3.1 Identify the DQ requirements for S-1xx PS in S-97

The recommended data quality measures of each S-1xx PS have been identified by the subWG according to S-97 part C Table C-7-1 – Recommended data quality measures.

Data Quality Measure	DQ measure / description	Applicable to spatial representation types	S-101	S-102	S-111	S-121	S-122	S-123	S-127	S-129
Completeness / Commission	numberOfExcessItems	All S-100 based PS	Y	Y	Y	Y	Y	Y	Y	Y
Completeness / Commission	numberOfDuplicateFeatureInstances	All S-100 based PS	Y	Y	Y	Y	Y	Y	Y	Y
Completeness / Omission	numberOfMissingItems	All S-100 based PS	Y	Y	Y	Y	Y	Y	Y	Y
Logical Consistency / Conceptual Consistency	numberOfInvalidSurfaceOverlaps	PS with geometric surfaces.	Y	Y	Y	Y	Y	Y	Y	Y
Logical Consistency / Domain Consistency	numberOfNonconformantItems	All S-100 based PS	Y	Y	Y	Y	Y	Y	Y	Y
Logical Consistency / Format Consistency	physicalStructureConflictsNumber	All S-100 based PS	Y	Y	Y	Y	Y	Y	Y	Y
Logical Consistency / Topological Consistency	rateOfFaultyPointCurveConnections	PS with curves.	Y	N	N	Y	Y	Y	Y	Y
Logical Consistency / Topological Consistency	numberOfMissingConnectionsUndershoots	PS with curves	Y	N	N	Y	Y	Y	Y	Y
Logical Consistency / Topological Consistency	numberOfMissingConnectionsOvershoots	PS with curves.	Y	N	N	Y	Y	Y	Y	Y
Logical Consistency / Topological Consistency	numberOfInvalidSlivers	PS with geometric surfaces.	Y	N	N	Y	Y	Y	Y	Y

IHO

International Hydrographic Organization

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

Data Quality Measure	DQ measure / description	Applicable to spatial representation types	S-101	S-102	S-111	S-121	S-122	S-123	S-127	S-129
Logical Consistency / Topological Consistency	numberOfInvalidSelfIntersects	PS with curves / geometric surfaces.	Y	N	N	Y	Y	Y	Y	Y
Logical Consistency / Topological Consistency	numberOfInvalidSelfOverlap	PS with curves / geometric surfaces.	Y	N	N	Y	Y	Y	Y	Y
Positional Accuracy / Absolute or External Accuracy	Root Mean Square Error	PS with objects that have coordinative values associated.	Y	Y	Y	Y	Y	Y	Y	Y
Positional Accuracy / Vertical Position Accuracy	linearMapAccuracy2Sigma	PS with objects that have a vertical coordinative values associated.	Y	Y	Y	Y	Y	Y	Y	Y
Positional Accuracy / Horizontal Position Accuracy	linearMapAccuracy2Sigma	PS with objects that have a horizontal coordinative values associated.	Y	Y	Y	Y	Y	Y	Y	Y
Positional Accuracy / Gridded Data Position Accuracy	Root mean square error of planimetry	PS with objects that have a gridded coordinative values associated.	N	Y	Y	N	N	N	N	N
Temporal Quality / Temporal Consistency	Correctness of ordered events or sequences, if reported.	PS with objects that have a time value associated.	Y	Y	Y	Y	Y	Y	Y	Y
Thematic Accuracy / Thematic Classification Correctness	miscalculationRate	All S-100 based PS.	Y	Y	Y	Y	Y	Y	Y	Y
Aggregation Measures / Aggregation Measures	DataProductSpecificationPassed	PS that a require a complete pass of all elements of a dataset/dataset series/spatial object types	Y	Y	Y	Y	Y	Y	Y	Y
Aggregation Measures / Aggregation Measures	DataProductSpecificationFailRate	PS that a require a complete pass of all elements of a dataset/dataset series/spatial object types	Y	Y	Y	Y	Y	Y	Y	Y

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2 The Cross check have been implemented one by one to verify whether these recommended DQ measures are included.

3.2.1 S-101

3.2.1.1 DQ chapters of S-101

The DQ chapter of S-101 is as follow:

6 Data Quality

6.1 Introduction

Data quality allows users and user systems to assess fitness for use of the provided data. Data quality measures and the associated evaluation are reported as metadata of a data product. This metadata improves interoperability with other data products and provides usage by user groups that the data product was not originally intended for. The secondary users can make assessments of the data product usefulness in their application based on the reported data quality measures.

For S-101 the following data quality elements have been included;

- Conformance to this Product Specification;
- Intended purpose of the data product;
- Completeness of the data product in terms of coverage;
- Logical Consistency;
- Depth Uncertainty and Accuracy;
- Positional Uncertainty and Accuracy;
- Thematic Accuracy;
- Temporal Quality;
- Aggregation measures;
- Validation checks or conformance checks including:
  - General tests for dataset integrity;
  - Specific tests for compliance against the S-101 data model.

For S-101 data quality is divided into two parts – data compliance, usability and integrity against all requirements of S-101; and bathymetric data quality.

6.1.1 Data Compliance and Usability

All S-101 datasets must be validated against the above data quality elements using conformance checks that are located in Annex C – ENC Validation Checks. As a minimum requirement, all datasets must conform to all checks that are categorized as "Critical" in Annex C.

S-101 datasets must conform to all mandatory elements of Annex A – Data Classification and Encoding Guide, where the word 'must' is used.

In addition to the above, dataset usability must be assessed against:

- Intended user requirements in regard to coverage, scale and specific content requirements as defined by the Producing Agency and key stakeholders;
- Conformance to established maintenance processes (see Section 8); and
- Overall compliance with the S-101 Product Specification, including context-specific evaluation of individual encoding instances for requirement of conformance to checks classified as "Error" and "Warning" in Annex C – ENC Validation Checks.

For dataset integrity requirements, see clause 11.6.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2.1.2 Result of cross check

Recommendations	S-101
1.Completeness	N
2.Conceptual consistency	N
3. Domain consistency	N
4. Format consistency	N
5. Topological consistency	N
6. Positional Accuracy	N
7. Thematic Accuracy	N
8. Temporal Quality	N
9. Aggregation	N
10. Introduction to DQ Paragraph	Y

Y=YES, N=NO, N/A=not application

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

NOTES:

a） In terms of “10. Introduction to DQ Paragraph”， S-101 follows the template provided by S-97 but makes a distinction of the positional accuracy to depth and positional (horizontal). We don’t agree that this is necessary.

3.2.1.3 DQ elements included in the other parts of S-101

No DQ element is included in the other parts of S-101.

6 Data Quality

6.1 Introduction

Data quality allows users and user systems to assess fitness for use of the provided data. Data quality measures and the associated evaluation are reported as metadata of a data product. This metadata improves interoperability with other data products and provides usage by user groups that the data product was not originally intended for. The secondary users can make assessments of the data product usefulness in their application based on the reported data quality measures.

For S-101 the following data quality elements have been included;

- Conformance to this Product Specification;
- Intended purpose of the data product;
- Completeness of the data product in terms of coverage;
- Logical Consistency;
- Depth Uncertainty and Accuracy;
- Positional Uncertainty and Accuracy;
- Thematic Accuracy;
- Temporal Quality;
- Aggregation measures;
- Validation checks or conformance checks including:
  - General tests for dataset integrity;
  - Specific tests for compliance against the S-101 data model.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

3.2.2 S-102

3.2.2.1 DQ chapters of S-102

The DQ chapter of S-102 is as follow:

6 Data Quality 23
6.1 Completeness 23
6.1.1 Commission 23
6.1.2 Omission 23
6.2 Logical Consistency 23
6.2.1 Conceptual Consistency 23
6.2.2 Domain Consistency 23
6.2.3 Format Consistency 23
6.3 Positional Accuracy 24
6.3.1 Accuracy of a Time Measurement 24
6.3.2 Gridded Data Positional Accuracy 24
6.3.3 Relative Internal Positional Accuracy 24
6.4 Temporal Accuracy 24
6.4.1 Temporal Consistency 24
6.4.2 Temporal Validity 24

6.5 Thematic Accuracy 24
6.5.1 Thematic Classification Correctness 24
6.5.2 Non-Quantitative Attribute Accuracy 24

3.2.2.2 Result of cross check

Recommendations	S-102
1.Completeness	Y
2.Conceptual consistency	Y
3. Domain consistency	Y
4. Format consistency	Y
5. Topological consistency	N/A
6. Positional Accuracy	Y
7. Thematic Accuracy	Y
8. Temporal Quality	Y
9. Aggregation	N
10. Introduction to DQ Paragraph	N

Y=YES, N=NO, N/A=not application

NOTES:

a） In terms of “10. Introduction to DQ Paragraph”， S-102 doesn’t follow the template provided by S-97 but only includes the first paragraph of the template.

3.2.2.3 DQ elements included in the other parts of S-102

No DQ element is included in the other parts of S-102.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

6 Data Quality

Data quality allows users and user systems to assess fitness for use of the provided data. Data quality measures and the associated evaluation are reported as metadata of a data product. This metadata improves interoperability with other data products and provides usage by user groups that the data product was not originally intended for. The secondary users can make assessments of the data product usefulness in their application based on the reported data quality measures.

6.1 Completeness

6.1.1 Commission

The S-102 bathymetric grid has a high-level of completeness regarding commission, due to the fact that the issuing hydrographic office has deemed the grid to contain all the necessary data and/or considered all contributing factors required to make a navigationally valid product. These factors are recorded in the metadata for the file.

6.1.2 Omission

The S-102 bathymetric grid has a high level of completeness in regards to omission, due to the fact that the issuing hydrographic office will have noted any major discrepancies or negative quality factors in the applicable fields of the metadata for the file.

6.2 Logical Consistency

6.2.1 Conceptual Consistency

The conceptual consistency of S-102 grids is maintained through this and related specifications which are conceptually consistent with the accepted standards.

6.2.2 Domain Consistency

The domain consistency of S-102 grids is maintained through the definition of their primary purpose, which is safety of navigation. The data contained can also be used derivatively for other scientific/fields domains (secondary purposes). All processes used in primary purpose generation is geared solely towards the satisfaction of safety of navigation concerns.

6.2.3 Format Consistency

The formatting consistency of S-102 grids is maintained due to the overriding encoding (HDF5) defined

6.3 Positional Accuracy

6.3.1 Accuracy of a Time Measurement

Temporal aspects of bathymetric grids are confined to elements of the vertical control processes. These aspects are addressed during the formulation and application of vertical control processes applied by the various hydrographic offices. Details of these processes will be included in the Lineage portion of the metadata defined in section 12 of this Product Specification.

6.3.2 Gridded Data Positional Accuracy

Gridded positional accuracy is defined by the precision of the positional reference used to specify its location within its spatial projection. These positional references are contained within the spatial metadata of the S-102 grid. Nodes within a bathymetric grid have an absolute position with no horizontal error with vertical values that are calculated for that position by the processes and procedures used by each hydrographic office during the creation of the S-102 grid. Appropriate selection of both the origin reference points and positional resolution are important and are another factor in gridded positional accuracy.

6.3.3 Relative Internal Positional Accuracy

The internal positional accuracy is defined as the precision of the location of each node within the S-102 grid. The position of each node within the grid is referenced by a row and column combination. The metadata for the S-102 defines a gridded resolution along both the X and Y axis of the grid. This absolute position of a node within the spatial projection of the grid is calculated using the row/column and the X/Y resolution. In this case, the accuracy is controlled by the precision used in defining these resolutions.

6.4 Temporal Accuracy

6.4.1 Temporal Consistency

Temporal aspects of bathymetric grids are confined to elements of the vertical control processes. These aspects are addressed during the formulation and application of vertical control processes applied by the various hydrographic offices. Details of these processes will be included in the Lineage portion of the metadata defined in section 12 of this Product Specification.

6.4.2 Temporal Validity

Temporal aspects of bathymetric grids are confined to elements of the vertical control processes. These aspects are addressed during the formulation and application of vertical control processes applied by the various hydrographic offices. Details of these processes will be included in the Lineage portion of the metadata defined in section 12 of this Product Specification.

6.5 Thematic Accuracy

6.5.1 Thematic Classification Correctness

For S-102 bathymetric grids there are two classifications of data values, which are land and water. There are two considerations for accessing classification correctness when using the grid. The first is that values given in the depth layer of the S-102 grid are based on the associated hydrographic offices chosen vertical datum. Should another value in relation to a different vertical datum be required, a series of correctors would need to be applied. Secondly, when considering the data values, the value stored in the corresponding uncertainty node must be considered. This uncertainty value is a +/- value and when assessing the classification correctness must be applied. The new value(s) generated when applied may cause a change in the classification.

6.5.2 Non-Quantitative Attribute Accuracy

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

3.2.3 S-111

3.2.3.1 DQ chapters of S-111

The DQ chapter of S-111 is as follow:

In each case, missing data or points over land are tagged with a unique value. A Surface Current coverage data set is complete when the grid coverage value matrix contains direction and speed values or the null value for every vertex point defined in the grid, and when all of the mandatory associated metadata is provided. See Annex D – Tests for Completeness (Normative).

Logical Consistency ensures that the data are stored in a consistent manner: The HDF structure used to hold the data was designed to enforce such consistency. In addition, the placement of current arrow symbols is consistent with the accepted coastline so that the centroid of the arrow is place within the water domain (see Figure 9.1), and if the water depth is zero, the symbol is not shown.

Thematic Accuracy insures that the values represented (speed and direction) are representative of the true situation. Measurement and modeling errors may put limits on these values.

Aggregation describes global quality values related to a particular dataset. For surface currents, each dataset will be evaluated separately.

Usability will be continually assessed through user and manufacturer response to the symbols and analysis presented in the latest Product Specification.

6 Data Quality

6.1 Assessment of data

Data quality allows users and user systems to assess fitness for use of the provided data. Data quality measures and the associated evaluation are reported as metadata of a data product. This metadata improves interoperability with other data products and provides usage by user groups that the data product was not originally intended for. The secondary users can make assessments of the data product usefulness in their application based on the reported data quality measures.

The prescribed precision (see Annex A – Data Classification and Encoding Guide) of current speed (0.01 kn) and direction (0.1 arc-deg) is close to the perceived accuracy of the data, but the increased precision is useful for time integration of current vectors and for the computation of spatial gradients (that is, non-navigational uses).

Important factors in the quality of surface current data for navigation consists of the quality of

- the observed data;
- the predicted/forecast data;
- the positional data; and
- the time stamp.

Factors determining the accuracy of the data are shown in Table 6.1. Information of the quality of the components of the data is normally available in field survey reports, QC analyses, or other technical reports.

Table 6.1 – Data types and accuracy factors

Type of Data	Factors Influencing Accuracy
Observed Current	Accuracy of the sensors Processing techniques
Predicted/forecast Current	Quality of input data Timeliness of input data Mathematical modelling techniques Accuracy of harmonic constants
Horizontal Position	Accuracy of geolocation techniques Model grid accuracy
Vertical Position	Accuracy of vertical datum
Time stamp	Sensor accuracy Data time tagging accuracy

Data quality measures for the entire data set are described in clause 10.2.3. These include, *horizontalPositionUncertainty*, *verticalUncertainty*, and *timeUncertainty*. The additional data quality measures *speedUncertainty* and *directionUncertainty* are described in clause 10.2.4.

6.2 Additional components

Additional data quality measures include Completeness, Logical Consistency, Thematic Accuracy, Aggregation, and Usability.

Completeness consists of commission and omission of data. For surface current data in gridded form, there is likely to be an excess of data for a region. For observed or historical, there is likely to be a dearth of data.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

3.2.3.2 Result of cross check

Recommendations	S-111
1.Completeness	Y
2.Conceptual consistency	N
3. Domain consistency	N
4. Format consistency	N
5. Topological consistency	N/A
6. Positional Accuracy	Y
7. Thematic Accuracy	Y
8. Temporal Quality	N
9. Aggregation	Y
10. Introduction to DQ Paragraph	N

Y=YES, N=NO, N/A=not application

NOTES:

a）In terms of “10. Introduction to DQ Paragraph”， S-111 doesn’t follow the template provided by S-97 but only includes the first paragraph of the template.

b） S-111 does include an “Logical Consistency” , however, it has nothing to do with the proposed in S-97.

3.2.3.3 DQ elements included in the other parts of S-111

Conceptual consistency, Domain consistency, Positional Accuracy and Temporal Quality are included in Annex D.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2.4 S-121

3.2.4.1 DQ chapters of S-121

The DQ chapter of S-121 is as follow:

6 Data Quality

6.1 Introduction

Data quality allows users and user systems to assess fitness for use of the provided data. Data quality measures and the associated evaluation are reported as metadata of a data product. This metadata improves interoperability with other data products and provides usage by user groups that the data product was not originally intended for. The secondary users can make assessments of the data product usefulness in their application based on the reported data quality measures.

For S-121 the following data quality elements have been included;

- Completeness of the data product in terms of coverage;
- Logical Consistency;
- Positional Accuracy;
- Temporal Accuracy;
- Thematic Accuracy;
- Data Set Validation checks.

6.2 Completeness

An S-121 product contains a quality geometric and attributive description (current, accurate, consistent), homogeneous and standardised of the entire set of MLB managed by the organization which issued the data set.

6.2.1 Redundant or oversampled data

MLB data is official data and therefore redundant and oversampled data elements are not permitted. It is up to the data producer to ensure consistency with official source documentation and correct attribution as such in the S-121 dataset.

Duplicate instances of features (defined by equal geometric realization, feature and attribute definition and temporal validity) are not permitted. Multiple instances of feature classes must be differentiated by means of attributes or metadata. Different versions of a feature instance are permitted using the versioning attributes to distinguish which is valid and any particular time.

6.2.2 Omission

The description of jurisdictional zones through the definition of MLB is often done through conceptual descriptions which may be complete or incomplete. For the sake of clarity features defined in the Feature Catalogue have very strict geometric primitives to ensure their geometries match the feature's definition in related legislation.

In order to also support conceptual description, these features also may be described textually or by reference to other features. Certain features such as geodetic lines or loxodromes, may have densification requirements (which may be separately documented).

6.3 Logical consistency

6.3.1 Domain consistency

Attribute values must be validated to ensure they are within the defined range. The allowed ranges are defined in the Feature Catalogue.

6.3.2 Format consistency

Data encoded in one of the S-121 data formats; either GML or the Explicit Text Format defined in Annex B should be validated in accordance with the format specification.

6.3.3 Topological consistency

The dataset validation tests the geometric consistency of a dataset should reflect S-100 Level 3a geometry.

6.4 Positional accuracy

Authoritative data has by definition absolute accuracy because it represents a defined position without qualification. Authoritative data may be described in treaties, national legislation and/or deposits in compliance with UNCLOS. These absolute accuracy positions may be included in a data set along with derived data.

Official data has by definition absolute accuracy because it depicts a determination of a state's Maritime Limits and Boundaries where no process exists to make it authoritative or where a technical process has been required to realize the data; although errors may exist in the data no disadvantage will apply to a user that acts in good faith. Official data may be described in treaties, national legislation and/or deposits in compliance with UNCLOS and become authoritative data. These absolute accuracy positions may be included in a data set along with derived data. Official data may be normalized to a single spatial reference system.

Data derived from authoritative or official data has the accuracy of the transformation or other process used to generate the data. A location may be represented by a point that has an absolute accuracy in one Coordinate Reference System, but has a different transformed value in another CRS. Since multiple treaties may have been agreed at different times there may be multiple locations with absolute accuracy in a MLB data set that use different CRS. If all of these points are transformed to a common CRS, so that they can be managed in a GIS and used, then each point may have an error resultant from the transformation process. The transformation source and the resultant error may be referenced in a source reference within the structure for each transformed point.

The accuracy of derived lines, such as transformed or densified lines shall also carry an accuracy statement in a referenced source for the transformation process.

6.5 Temporal accuracy

Constituent of data features may be established at different dates. A validity attribute is mandatory for all features. A single creation date may be assigned to a component feature derived from different sources. The source description shall be used to describe the difference between the actual establishment date of a feature and any date assigned to a group.

6.6 Thematic accuracy

6.6.1 Thematic classification correctness

Each of the standardized features will have an IHO code to serve as a linkage to proper depiction based on the IHO S-100 Feature Concept Dictionary. When a feature type will be realised into a specific instance the corresponding IHO code shall be linked to the object to ensure that the proper thematic classification is used. Through such automation, no further verification is needed.

6.6.2 Non-quantitative attribute accuracy

The method used for evaluating the accuracy of the non-quantitative attribute values with respect to reality is determined by the type of data and its method of acquisition, and thus rely on the metadata information and data description provided at the time of the dataset integration.

6.6.3 Quantitative attribute accuracy

The method used for evaluating the accuracy of the quantitative attribute values with respect to reality is determined by the type of data and its method of acquisition, and may be calculated in accordance with clause 5.3 or may rely on the metadata information.

The presence of inconsistencies or discrepancies in information provided by different states does not indicate the existence of a dispute among the relevant states.

6.7 Dataset Validation checks

A set of dataset validation checks are included as Appendix A of this Product Specification. These form a minimum set of data validation tests and do not provide assurance that features meet an intended level of logical or topological consistency. This section of the Product Specification contains guidance to ensure a minimum level of consistency is achieved across all S-121 datasets.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

3.2.4.2 Result of cross check

Recommendations	S-121
1.Completeness	Y
2.Conceptual consistency	N
3. Domain consistency	Y
4. Format consistency	Y
5. Topological consistency	Y
6. Positional Accuracy	Y
7. Thematic Accuracy	Y
8. Temporal Quality	Y
9. Aggregation	N
10. Introduction to DQ Paragraph	Y

NOTES:

- Y=YES, N=NO, N/A=not application

3.2.4.3 DQ elements included in the other parts of S-121

No DQ element is included in the other parts of S-121.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2.5 S-122

3.2.5.1 DQ chapters of S-122

The DQ chapter of S-122 is as follow:

9 Data Quality

9.1 Introduction

S-122 products must be tested with the S-122 specific checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are a mix of data format validation checks,

conformance to standard checks and logical consistency checks. The checks are listed in Annex E.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

3.2.5.2 Result of cross check

Recommendations	S-122
1.Completeness	N
2.Conceptual consistency	N
3. Domain consistency	N
4. Format consistency	N
5. Topological consistency	N
6. Positional Accuracy	N
7. Thematic Accuracy	N
8. Temporal Quality	N
9. Aggregation	N
10. Introduction to DQ Paragraph	N

Y=YES, N=NO, N/A=not application

NOTES:

a）In terms of “10. Introduction to DQ Paragraph”， S-122 does include an introduction, however, this is their own and has nothing to do with the proposed in S-97.

b) In fact, the whole DQ chapter of S-122 is only related to data validation.

3.2.5.3 DQ elements included in the other parts of S-122

No DQ element is included in the other parts of S-122.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2.6 S-123

3.2.6.1 DQ chapters of S-123

The DQ chapter of S-123 is as follow:

9 Data Quality

9.1 Introduction

S-123 products must be tested with the S-123 specific checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are a mix of data format validation checks, conformance to standard checks and logical consistency checks. The checks are listed in Annex E.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

3.2.6.2 Result of cross check

Recommendations	S-123
1.Completeness	N
2.Conceptual consistency	N
3. Domain consistency	N
4. Format consistency	N
5. Topological consistency	N
6. Positional Accuracy	N
7. Thematic Accuracy	N
8. Temporal Quality	N
9. Aggregation	N
10. Introduction to DQ Paragraph	N

Y=YES, N=NO, N/A=not application

NOTES:

a）In terms of “10. Introduction to DQ Paragraph”， S-123 does include an introduction, however, this is their own and has nothing to do with the proposed in S-97.

b) In fact, the whole DQ chapter of S-123 is only related to data validation and is the same as S-122.

3.2.6.3 DQ elements included in the other parts of S-123

No DQ element is included in the other parts of S-123.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2.7 S-127

3.2.7.1 DQ chapters of S-127

The DQ chapter of S-127 is as follow:

9 Data quality

9.1 Introduction

Data quality allows users and user systems to assess fitness for use of the provided data. Data quality measures and the associated evaluation are reported as metadata of a data product. This metadata improves interoperability with other data products and provides usage by user groups that the data product was not originally intended for. The secondary users can make assessments of the data product usefulness in their application based on the reported data quality measures.

For S-127 the following data quality elements have been included:

- Conformance to this Product Specification;
- Intended purpose of the data product;
- Completeness of the data product in terms of coverage;
- Logical Consistency;
- Positional Uncertainty and Accuracy;
- Thematic Accuracy;
- Temporal Quality;
- Aggregation measures;
- Elements specifically required for the data product (none currently identified for S-127);
- Validation checks or conformance checks including:
  - General tests for dataset integrity;
  - Specific tests for a specific data model;

9.2 Quality measure elements

The data quality measures recommended in S-97 (Part C) and their applicability in S-127 are indicated in Table 9.1 below. NA indicates the measure is not applicable. This table reproduces the first 4 columns of the data quality checklist recommended elements and replaces the final column with descriptions of the scope of the element in the context of S-127 datasets.

No.	Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Scope in S-127
1	Completeness / Commission	Excess data present in a dataset, as described by the scope.	numberOfExcessItems / This data quality measure indicates the number of items in the dataset, that should not have been present in the dataset.	dataset/dataset series	All features and info types
2	Completeness / Commission	Excess data present in a dataset, as described by the scope.	numberOfDuplicateFeatureInstances / This data quality measure indicates the total number of exact duplications of feature instances within the data.	dataset/dataset series	All features and info types
3	Completeness / Omission	Data absent from the dataset, as described by the scope.	numberOfMissingItems / This data quality measure is an indicator that shows that a specific item is missing in the data.	dataset/dataset series/spatial object type	All features and info types
4	Logical Consistency / Conceptual Consistency	Adherence to the rules of a conceptual schema.	numberOfInvalidSurfaceOverlaps / This data quality measure is a count of the total number of erroneous overlaps within the data. Which surfaces may overlap and which must not is application dependent. Not all overlapping surfaces are necessarily erroneous.	spatial object / spatial object type	Features with surface geometry; spatial objects of type surface
5	Logical Consistency / Domain Consistency	Adherence of the values to the value domains.	numberOfNonconformantItems / This data quality measure is a count of all items in the dataset that are not in conformance with their value domain.	spatial object / spatial object type	All features and info types
6	Logical Consistency / Format Consistency	Degree to which data is stored in accordance with the physical structure of the data set, as described by the scope	physicalStructureConflictsNumber / This data quality measure is a count of all items in the dataset that are stored in conflict with the physical structure of the dataset.	dataset/dataset series	All features and info types
7	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	rateOfFaultyPointCurveConnections / This data quality measure indicates the number of faulty link-node connections in relation to the number of supposed link-node connections. This data quality measure gives the erroneous point-curve connections in relation to the total number of point-curve connections.	spatial object / spatial object type	Features with curve geometry; spatial objects of curve types
8	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	numberOfMissingConnectionsUndershoots / This data quality measure is a count of items in the dataset within the parameter tolerance that are mismatched due to undershoots.	spatial object / spatial object type	Features with curve geometry; spatial objects of curve types
9	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	numberOfMissingConnectionsOvershoots / This data quality measure is a count of items in the dataset within the parameter tolerance that are mismatched due to overshoots.	spatial object / spatial object type	Features with curve geometry; spatial objects of curve types

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

No.	Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Scope in S-127
18	Thematic Accuracy / ThematicClassification Correctness	Comparison of the classes assigned to features or their attributes to a universe of discourse.	miscalculationRate / This data quality measure indicates the number of incorrectly classified features in relation to the number of features that are supposed to be there. [Adapted from ISO 19157] This is a RATE which is a ratio, and is expressed as a REAL number representing the rational fraction corresponding to the numerator and denominator of the ratio. For example, if there are 1 items that are classified incorrectly and there are 100 of the items in the dataset then the ratio is 1/100 and the reported rate = 0.01.	dataset/dataset series/spatial object type	All features and info types
19	Aggregation Measures / AggregationMeasures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationPassed / This data quality measure is a boolean indicating that all requirements in the referred data product specification are fulfilled.	dataset/dataset series/spatial object type	Dataset as a whole
20	Aggregation Measures / AggregationMeasures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationFailRate / This data quality measure is a number indicating the number of data product specification requirements that are not fulfilled by the current product/dataset in relation to the total number of data product specification requirements.	dataset/dataset series/spatial object type	Dataset as a whole

Table 9.1 - IHO recommended quality elements and their relevance to S-127

9.3 Test methods

Test methods consist of executing the relevant tests from Annex E (Validation Checks) for each quality element in Table 9.1 and counting the number of instances in the dataset which fail the checks for that quality element.

Note that in some cases "executing the relevant test" may involve comparing the encoded S-127 dataset to the source material by visual means (e.g., for measures 17 and 18). For tests requiring visual comparison of encoded data to source material, sampling methods may be used if the volume of data precludes checking all the relevant data objects.

9.3.1 Accuracy computations

Recommendations for Positional Accuracy / Absolute or External Accuracy:
Maximum RMSE (horizontal) = E / 10000

Where:
E = Denominator of intended scale of mapping

9.4 Data quality testing and reporting

S-127 products must be tested with the S-127 specific checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are a mix of data format validation checks, conformance to standard checks and logical consistency checks. The checks are listed in Annex E.

Production and certification processes for S-127 data should include a standalone quality report which provides full information on the original results (with evaluation procedures and measures applied).

The dataset or exchange set metadata that is distributed with the exchange set may describe only the aggregated result with a reference to the original results described in the standalone quality report. The aggregated Data Quality result provides an indication if the dataset has passed conformance to the Data Product Specification.

Data Quality Measure Aggregation results should be included to indicate if the dataset/dataset series have passed the Product Specifications. The elements which must be included are described in Table 9.2.

Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Applicable to spatial representation types
Aggregation Measures / AggregationMeasures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationPassed / This data quality measure is a boolean indicating that all requirements in the referred data product specification are fulfilled.	dataset	All features and information types of the dataset
Aggregation Measures / AggregationMeasures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationFailRate / This data quality measure is a number indicating the number of data product specification requirements that are not fulfilled by the current product/dataset in relation to the total number of data product specification requirements.	dataset	All features and information types of the dataset

Table 9.2 - Elements of data quality aggregated report (extract from S-97 Part C checklist)

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2.7.2 Result of cross check

Recommendations	S-127
1.Completeness	Y
2.Conceptual consistency	Y
3. Domain consistency	Y
4. Format consistency	Y
5. Topological consistency	Y
6. Positional Accuracy	Y
7. Thematic Accuracy	Y
8. Temporal Quality	Y
9. Aggregation	Y
10. Introduction to DQ Paragraph	Y

NOTES:

Y=YES, N=NO, N/A=not application

-

3.2.7.3 DQ elements included in the other parts of S-127

No DQ element is included in the other parts of S-127.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

International Hydrographic Organization

3.2.8 S-129

3.2.8.1 DQ chapters of S-129

The DQ chapter of S-129 is as follow:

13 Data Quality

13.1 Introduction

Areas where UKCM services are in place are typically covered by very high definition and up-to-date bathymetry, and have available predicted and observed (that is, real time) met-ocean data.

Bathymetric, tidal and other met-ocean data used to generate products in compliance with this Product Specification are provided by official sources using quality assured processes outside the scope of this Product Specification. This information is, therefore, assumed to be of high quality and guaranteed by the processes employed by the relevant authorities.

Quality of UKCM products used for navigation depends on the combined quality of many inputs including observed and forecast data (for example bathymetry, tide, water levels, currents, tidal streams etc.) and ship positional data. Quality information relevant to the many data inputs used by a UKCM service is difficult to produce as a meaningful metric of UKCM product quality able to be understood by the mariner.

Data validity is time-limited and is another reason why it is not practical to provide a meaningful measure of data quality for UKCM products.

UKCM products are generated containing margins that account for uncertainty to guarantee the safety of individual ships for stated periods of validity.

3. IMPLEMENTATION PROCESS OF THE CROSS CHECK OF DQ CHAPTERS

IHO

International Hydrographic Organization

3.2.8.2 Result of cross check

Recommendations	S-129
1.Completeness	N
2.Conceptual consistency	N
3. Domain consistency	N
4. Format consistency	N
5. Topological consistency	N
6. Positional Accuracy	N
7. Thematic Accuracy	N
8. Temporal Quality	N
9. Aggregation	N
10. Introduction to DQ Paragraph	N

Y=YES, N=NO, N/A=not application

3.2.8.3 DQ elements included in the other parts of S-129

No DQ element is included in the other parts of S-129.

NOTES:

In terms of "10. Introduction to DQ Paragraph", S-129 does include an introduction, but it is to mainly explain why DQ cannot be evaluated in the context of UKC (as it should be already validated based on other product specifications). However, S-97 states: "All S-100-based Product Specifications should include comprehensive ways of capturing information about the quality of the data."

IHO

International Hydrographic Organization

4. CROSS CHECK MATRIX

Recommendations	S-101	S-102	S-111	S-121	S-122	S-123	S-127	S-129
1.Completeness	N	Y	Y	Y	N	N	Y	N
2.Conceptual consistency	N	Y	N	N	N	N	Y	N
3. Domain consistency	N	Y	N	Y	N	N	Y	N
4. Format consistency	N	Y	N	Y	N	N	Y	N
5. Topological consistency	N	N/A	N/A	Y	N	N	Y	N
6. Positional Accuracy	N	Y	Y	Y	N	N	Y	N
7. Thematic Accuracy	N	Y	Y	Y	N	N	Y	N
8. Temporal Quality	N	Y	N	Y	N	N	Y	N
9. Aggregation	N	N	Y	N	N	N	Y	N
10. Introduction to DQ Paragraph	Y	N	N	Y	N	N	Y	N

Y=YES, N=NO, N/A=not application

IHO

International Hydrographic Organization

4. CROSS CHECK MATRIX

NOTES:

a) DQ chapters of the 8 published S-1xx PSs are not in a harmonized way.

b) Some S-1xx PSs like S-122 and S-123 confuse data quality and data validation.

c) DQ chapters of some S-1xx PSs do not conform to the S-97.

d) Except S-127, No S-1xx PSs implement the table C-7-1 – Recommended data quality measures of S-97.

IHO

International Hydrographic Organization

5. RECOMMENDATIONS

It is recommended to:

a) Develop a template by combining the DQ chapters of S-102 and S-127 for use in all S-1xx PSs.
b) Prepare suggestions to S-1xx PSs respectively.
c) Continue the cross check of DQ chapters of newly released S-1xx PSs such as S-104 Ed 1.0.0 and the new Edition of the published S-1xx PSs such as S-102 Ed 2.1.0.
d) Carry out the cross check of data validation of the published S-1xx PSs

6. ACTIONS REQUESTED FROM THE SUBWG

The DQWG is requested to:

a. Note the information provided;
b. Approve the recommendations above.