Survey to ZOC

Guidelines and recommendations for the population of CATZOC values from survey data

Data Quality Sub Working Group

In alphabetical order:

Andrew Talbota

Antonio Di Lietob,

Christos Kastrisiosc

Edward Handsd

Morgane Gaumete

Nicolas Davide

Nigel Robinsonf

Paulene Hockingeg

Rogier Broekmanh

Scott Youngbluti

Sean Legeerj

Thijs Ligteringenk

Samantha Lengol,

……

Affiliations:

a …..;

b ……;

c Center for Coastal & Ocean Mapping/Joint Hydrographic Center, University of New Hampshire, Durham, USA;

d;

e …….;

f …….;

g………..;

h …….;

i ……;

j …….;

SubWG Meetings:

1. April 23, 2021 (VTC)

2. September 21, 2021 (VTC)

# Related Works/Publications

The following standards/specifications are related to this Group’ work:

Data Quality

1. ISO/FDIS 19157 Geographic information — Data quality
2. SDTS (Spatial Data Transfer Standard) Guide for Technical Managers <https://www.mapi.gov.il/gisforum/tzevet_technology/stds-guide%20for%20technical%20managers.pdf>
3. Inspire D2.8.II.1 Data Specification on Elevation – Technical Guidelines <https://inspire.ec.europa.eu/id/document/tg/el>
4. EPA Guidance for Geospatial Data Quality Assurance Project Plans <https://www.epa.gov/sites/production/files/documents/guidance_geospatial_data_qapp.pdf>
5. FGDC (Federal Geographic Data Committee) Content Standards for Digital Geospatial Metadata

<https://www.fgdc.gov/standards/projects/metadata/base-metadata/v2_0698.pdfhttps://www.fgdc.gov/standards/projects/metadata/base-metadata/v2_0698.pdf>

1. AGI (Association for Geographic Information) National Transfer Format <https://www.geos.ed.ac.uk/~gisteac/proceedingsonline/Source%20Book%202004/SDI/National/UK/AGI/Guidelines%20for%20GI%20content%20and%20quality.htm>

Nautical Cartography

* <https://eng.gst.dk/media/2919223/behind-nautical-chart-vers-dec2018.pdf>

Quality Assessment

<https://hal.archives-ouvertes.fr/hal-02355473/document>

<https://icaci.org/files/documents/ICC_proceedings/ICC1995/PDF/Cap431.pdf>

<https://www.isprs.org/proceedings/XXXVII/congress/2_pdf/3_WG-II-3/10.pdf>

Data Quality Visualization

Beard\_etal\_1991\_NCGIA\_Visualization\_of\_SpatialData\_eScholarship UC item 6w1695bs

# 1. Questions:

* 1. What are the relevant components of spatial data quality?
* 2. What conceptual frameworks should be used to integrate data quality components with spatial data?
* 3. Why are they important for improving our understanding and use of spatial data?
* Is the framework set up with specific data types in mind?
* Is a global framework a good starting point and should there be a model that applies to all or several applications?
* what purposes data can be used for. Should quality information be offered in the form of a list of recommended uses and limitations.
* relevance is very important in the quality equation, --- we must look at relevance in terms of what people want to do with the data. ---Do we need one new ZOC for Exclusive?
* Do we attempt to measure error directly or by describing the compilation process (by examining the lineage). In terms of the measurement view the group discussed whether it was possible to test against the truth? – there are more than one realities ---see Beard et al p.20/178
* At what level of aggregation (primitive, object, object class, layer, tile, database) should the data quality information be stored and/or linked to the data? group agreed that data quality information should be stored and linked at all levels. The level at which quality information is stored may depend on the type of measurement. The notion of data quality depends on the primitives, but there is a need to model how quality information aggregates. A simple example of an aggregation issue is the following: A surveyor measures points or lines20 which have error. The question is then how to infer the error in the area of a polygon from errors in the point or line measurements. The group suggested that once a model of aggregation is known, it could be performed by the system to allow the quality information to be more easily maintained
* What ZOC level is CSB appropriate for? --- Data quality information should be supplied at the lowest levels to ensure that data are appropriate for certain applications, but very often information is not available on the precision or accuracy of the lowest level of measurement. Most data are now collected by someone else. In such cases we should try to predict the types and levels of quality information other users will need (CK-this is the case with CSB. We can predict based on users quality indicators or ratings/trusted users but we don’t verify their data somehow ---or they should be verified before departure and after departure to make sure they are within the requirements).

# Survey to ZOC

The DQWG should provide guidance on data quality aspects to hydrographic offices, in particular to ensure harmonized implementation (Terms of Reference art.3.b.iv). At DQWG16 (Feb2021, VTC) a dedicated subWG was created to draft guidelines and recommendations to Hydrographic Offices based on best practices to allocate CATZOC values (or S-101 Quality of Bathymetric Data values) from survey data qualified in application of the new Ed. 6.0 of S-44 – IHO Standards for Hydrographic Surveying.

This publication aims to standardize/define the components and structures of data quality measures for allocating ZOC on charts.

It describes the process from the first ping (data capture) to data storage (S-57, S-101, S-102), validation and finally descriptive quality indicator (CATZOC, QoBD) to provide the end user with meta-information to make an assessment if this dataset will fulfil his/her requirements (*fit for purpose / fitness for use*).

**Keywords:** Data Quality, Quality Criteria, Cartographic Generalisation, Validation, Fitness for Purpose, Nautical Chart,

# 1. Introduction (“Justification and Impacts”)

S-101 and S-102 are under development. Combining datasets of these types of data requires appropriate data quality elements and appropriate meta-quality information (CATZOC) to the Mariner in order to assess a safe route planning and execution of voyage.

Datasets provided by adjacent Producing Authorities may provide different depictions of the shape of the seabed and associated quality indicators. This work aims to provide tools to make an assessment of the self-consistency of datasets produced by one Producing Authority and may explain the differences to datasets produced by the adjacent Producing Authority.

Additionally, this work aims to help Producing Authorities to have more confidence in including CSB data into their nautical charts where they feel it is appropriate to do so.

*From Survey to CATZOC*, describes the process from the first ping (data capture) to data storage (S-57, S-101, S-102), validation and finally descriptive quality indicator (CATZOC, QoBD) to provide the end user with meta-information to make an assessment whether the dataset fulfills the *fit-for-purpose* (also *fitness-for-use*). The following items are discussed in this work:

* *Data capture*, and more specifically the associated accuracy and evaluation according to S-44 Ed 6.0.0; (Section 2)
* Data storage in S-57 format and S-101 format, with a focus on soundings, depth contours, and depth areas; (Section 3).
* Data storage in S-102 format and associated uncertainty values; (Section 3)
* *Data generalization* (Section 4)
* *Data quality* measures and recommended target results (validation); (Section 5 and 6)
* Assigning appropriate CATZOC (S-57) and Quality of Bathymetric data (S-101) values, (page 10)
* Added value of CSB data. (page 10)

The objective of this Sub Working Group is the development of a specification for ….

that Standard is to provide principles for describing the quality for geographic data and concepts for handling quality information for geographic data, and a consistent and standard manner to determine and report a dataset’s quality information. It aims also to provide guidelines for evaluation procedures of quantitative quality information for geographic data.

Working with data quality includes:

* understanding of the concepts of data quality related to geographic data. Annex B is a description of data quality concepts used to establish the components for describing the quality of geographic data;
* defining data quality conformance levels in data product specifications or based on user requirements (establishment of data product specifications is described in ISO 19131:2007);
* specifying quality aspects in application schemas;
* evaluating data quality;
* reporting data quality.

# 2. Data Capture

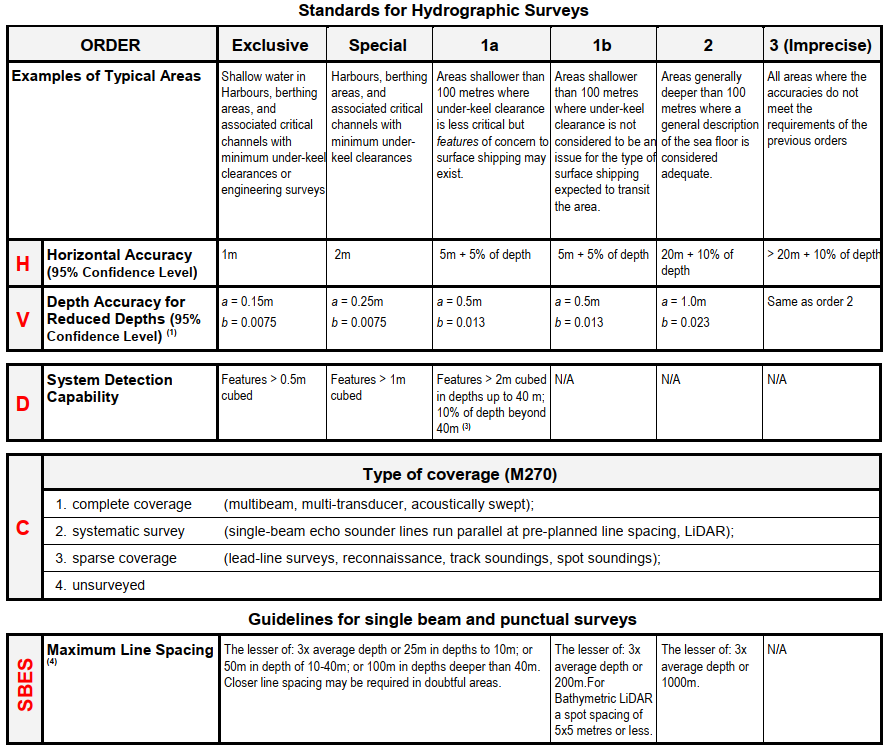
The elevation of a terrain surface, whether land based or bathymetric in nature, is one of the most important descriptors of the Earth‘s morphology. The main purpose of a Digital Elevation Model is to provide an elevation property with reference to a specified origin (vertical reference or datum). This property may be height (when the value is measured opposite to the gravity field of the Earth) or depth (when the value is measured in the direction of the gravity field). Therefore, they share the basic modelling concepts. Integrated land-sea models may be provided using either a height or depth property referenced to a known vertical reference. When an elevation property describes the bare surface of the land or sea floor, the related model is called Digital Terrain Model (DTM). When an elevation property includes the heights of the objects present on the surface (e.g., vegetation, man-made objects) the related model is referred as Digital Surface Model (DSM).[[1]](#footnote-2)

Capturing the elevation of a terrain surface on land is nowadays mostly done from an airplane using a LIDAR sensor. Capturing the depth of the seabed is nowadays done from ship (multibeam, singlebeam), airplane (LIDAR) or satellite (SDB)[[2]](#footnote-3). When measuring from ship, a distinction can be made between 1) official and officially sponsored surveys, 2) unofficial surveys, 3) passage surveys; see also IHO publication S-4, B-611 Credibility of sources.

Official (and officially sponsored) surveys prepared specifically for nautical charting should be validated by competent surveyors. It must be ensured, as far as possible, that any errors and uncertainties arising from the method of surveying are understood and that the survey remains acceptable for use; see IHO publication S-44.[[3]](#footnote-4) IHO publication S-44 has the classification of safety of navigation surveys shown in Table 1. [[4]](#footnote-5)

Table 1: S-44 Classifications

|  |  |  |
| --- | --- | --- |
| **Order** | **Characteristics** | **Intended usage** |
| **2** | General depiction of the bottom is considered adequate. Evenly distributed bathymetric coverage of 5% | Areas deeper than 200m. Existence of features that are large enough to impact on surface navigation and yet still remain undetected is considered to be unlikely. |
| **1b** | General depiction of the bottom is considered adequate. Evenly distributed bathymetric coverage of 5% | Only recommended where underkeel clearance is considered not to be an issue. An example would be an area where the bottom characteristics are such that the likelihood of there being a feature on the bottom that will endanger the type of surface vessel expected to navigate the area is low. |
| **1a** | 100% feature search, bathymetric coverage less than or equal to 100% is appropriate as long as the least depths over all significant features are obtained and the bathymetry provides an adequate depiction of the nature of the bottom topography. | Coastal waters, harbours, berthing areas, fairways and channels.  Underkeel clearance becomes less critical as depth increases, so the size of the feature to be detected increases with depth in areas where the water depth is greater than 40 meters. |
| **Special** | 100% feature search and 100% bathymetric coverage. Size of the features to be detected is more demanding than order 1a. | Areas where underkeel clearance is critical. Examples are: berthing areas, harbours and critical areas of fairways and shipping channels. |
| **Exclusive** | 200% feature search and 200% bathymetric coverage. Size of the features to be detected is more demanding than special order. | Shallow water areas (harbours, berthing areas and critical areas of fairways and channels) where there is an exceptional and optimal use of the water column and where specific critical areas with minimum underkeel clearance and bottom characteristics are potentially hazardous to vessels. |



## 2.1 Horizontal and vertical positioning and its associated uncertainty

Positioning is a fundamental part for every survey operation. The hydrographer must consider the geodetic reference frame, horizontal and vertical reference systems, their connections to other systems in use (e.g., land survey datums), as well as the uncertainty inherent within associated measurements. In this standard[[5]](#footnote-6), position and its uncertainty refer to the horizontal component of the sounding or feature, while the depth and its uncertainty refers to the vertical component of the same sounding or feature.[[6]](#footnote-7) Annex B provides background information on the realizations of geodetic reference frame, horizontal and vertical coordinate reference systems.

## 2.2 Horizontal Reference System

If horizontal positions are referenced to a local datum, the name and epoch of the datum should be specified and the datum should be tied to a realisation of a global (e.g., ITRF2018, WGS84(G1762)) or a regional (e.g., ETRS89, NAD83) reference frame and their later iterations. Transformations between reference frames/epochs should be taken into account, especially for surveys with low uncertainty[[7]](#footnote-8) (very accurate GNSS positioning).

## 2.3 Vertical Reference System

If the vertical component of the positions is referenced to a local vertical datum, the name and epoch of the datum should be specified. The vertical component of the positions (e.g., depths, drying heights) should be referenced to a vertical reference frame that is suitable for the data type and intended use. This vertical reference frame may be based on tidal observations (e.g., LAT, MWL, etc.), on a physical model (i.e., geoid) or a reference ellipsoid.[[8]](#footnote-9)

## 2.4 Chart and Land Survey Vertical Datum Connections

In order for bathymetric data to be correctly and fully utilised, chart and land survey vertical datum connections or relationships must be clearly determined and described. The IHO Resolution on Datums and Benchmarks, Resolution 3/1919, as amended, resolves practices which, where applicable, should be followed in the determination of these vertical datum connections. This essential resolution 3/1919, as amended, is available in the IHO Publication M-3, Resolutions of the International Hydrographic Organization, which is downloadable from the IHO homepage [www.iho.int](http://www.iho.int). Examples of DTM/DSM on land and sea are provided in Annex D.

## 2.5 Technologies

MBES --

SDB --

# 3. Data Storage

The DTM/DSM can be stored in a vector model (S-57, S-101) or grid model (S-102). The vector model consists of land elevation and bathymetry elements in the form of spot elevations, contour lines, and depth areas. Grid representation is based on a coverage geometry, indicating elevation values at the points of a rectified grid.

## 3.1 Heights and elevations (on land)

If it is required to encode the altitude of natural features (for example hills, coastlines, slopes), with the exception of trees, it must be done using the attribute ELEVAT (Figure 1 - a).

For artificial features (for example landmarks, buildings) or trees:

* If it is required to encode the altitude of the ground level at the base of the object, it must be done using ELEVAT (Figure 1 - b).
* If it is required to encode the altitude of the highest point of the object, it must be done using the attribute HEIGHT (Figure 1 - c).
* If it is required to encode the height of the object above ground level, it must be done using the attribute VERLEN (Figure 1 - d).

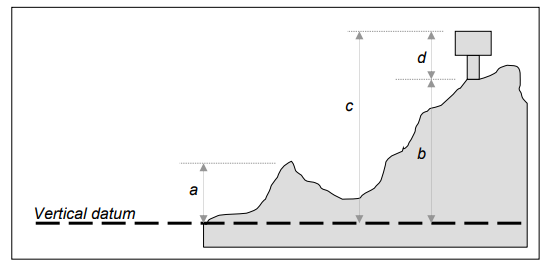


Figure 1 – heights and elevations (source IHO Publication S-57 UoC)

## 3.2 S-57 Geo-Objects

The storage of a DTM/DSM in S-57 is described in S-57 Appendix B.1: ENC Product Specification, Annex A: Use of the Object Catalogue for ENC, Edition 4.2.0 – April 2020.

### 3.2.1 Soundings

Geo Object: Sounding (SOUNDG)

Attributes: EXPSOU (exposition of sounding)

NOBJNM OBJNAM (object name)

QUASOU (quality of sounding measurement)

SOUACC (sounding accuracy) – see the use of the meta object M\_QUAL

STATUS

TECSOU (technique of sounding measurement) – only for lower reliability soundings.

~~VERDAT~~ (vertical datum)

INFORM

NINFOM

SORDAT (source date) – the production date of the source, e.g. the date of measurement

### 3.2.2 Depth contours

Geo object: Depth contour (DEPCNT)

Attributes: VALDCO – value of depth contour

~~VERDAT~~ – INFORM NINFOM

### 3.2.3 Depth Areas

Geo Object: Depth Area (DEPARE)

Attributes: DRVAL1 (depth range value 1) – shoalest

DRVAL2 (depth range value 2) – deepest

QUASOU (quality of sounding measurement)

SOUACC

~~VERDAT~~

~~INFORM~~

NINFOM

## 3.3 S-101 Geo-Objects

The storage of a DTM/DSM in S-101 is described in the S-101 Annex A – Data Classification and Encoding Guide (DCEG) Edition 1.0.1 March 2021.

### 3.3.1 Sounding

S-101 Geo Feature: Sounding (SOUNDG)

Primitives: Pointset

S-101 Attribute: feature name – multiplicity 0,\*

display name – multiplicity 0,\*

language – multiplicity 0,\*

name (OBJNAM, NOBJNM) – multiplicity 1,1

quality of vertical measurement (QUASOU) – multiplicity 0,\*

reported date (SORDAT) – multiplicity 0,1

status (STATUS) - multiplicity 0,1 (only one value possible: existence doubtful)

technique of vertical measurement (TECSOU) – multiplicity 0,\*

vertical uncertainty – multiplicity 0,1 (multiplicity is proposed to be 1,1 by S-101PT)

uncertainty fixed (SOUACC) – multiplicity 1,1

uncertainty variable – multiplicity 0,1

scale minimum (SCAMIN) – multiplicity 0,1

### 3.3.2 Depth contour

S-101 Geo Feature: Depth Contour (DEPCNT)

Primitives: Curve

S-101 Attribute: value of depth contour (VALDCO) – multiplicity 1,1

scale minimum (SCAMIN) – multiplicity 0,1

### 3.3.3 Depth Areas

S-101 Geo Feature: Depth Area (DEPARE)

Primitives: Surface

S-101 Attribute: depth range maximum value (DRVAL2) – multiplicity 1,1

depth range minimum value (DRVAL1) - multiplicity 1,1

## 3.4 S-102

The storage of a DTM/DSM in S-102 is described in the S-102 Bathymetric Surface Product Specification (Edition 2.0.0 October 2019). The S-102 Feature Catalogue layout is shown in Table 2.

Table 2: layout of the S-102 Feature Catalogue

|  |  |
| --- | --- |
| classification | unclassified |
| Simple Attribute |  |
| name | Depth |
| definition | the vertical distance from a given water level to the bottom |
| code | depth |
| alias | DEPTH |
| value Type | real |
|  |  |
| Simple Attribute |  |
| name | uncertainty |
| definition | The interval (about a given value) that will contain the true value of the measurement at a specific confidence level |
| code | uncertainty |
| remarks | Represents a +/- value defining the possible range of associated depth expressed a positive number |
| value Type | real |
|  |  |
| Feature Type |  |
| name | Bathymetry Coverage |
| definition | A set of value items required to define a dataset representing an depth calculation and its associated uncertainty. |
| code | BathymetryCoverage |
| multiplicity | 1..1 (always 1) |
| attribute ref | depth |
| multiplicity | 1..1 (always 1) |
| attribute ref | uncertainty |
|  |  |
| feature Use Type | geographic |
| permitted Primitives | coverage |

# 4. Data Generalization

Cartographic Generalization is the selection and simplified representation of detail appropriate to the scale and/or the purpose of a map (ICA 1967). Data quality and generalization for map production are closely related. The quality of the source data is one of the controls of generalization, along with map purpose, map scale, and graphic limits (see Robinson *et al.,* 1995), and, as such, influences map generalization process. Generalization relies on the quality of the source data. The more reliable and precise the data, the more detail exists in the data that needs to be generalized. When the data is sparse and poor less generalization is required, but, at the same time, our knowledge about the studied phenomenon is more limited. On the other hand, the generalization process impacts the quality of the output data. Generalization needs to deliver generalised data fit for a purpose. The quality of the generalised data is an evaluation of how closely the data fulfils the specified requirements (Regnauld, 2015). The output of the generalization must not give the users the false impression of completeness and quality of the data better than the reality.

The compilation of bathymetry on nautical charts is one of the most complicated and time-consuming generalization processes. The charted bathymetry is derived from a more detailed (source) dataset, either the survey data or a larger scale chart, with cartographic generalization.

## 4.1 Cartographic Constraints

The generalization process is a continuous compromise among the legibility, topology, morphology, and safety constraints as they are often incompatible with each other. The constraints are the requirements that the outcome of the generalization must fulfil. In nautical cartography:

1. *Safety* (or *functional*): depth information on the chart must not appear deeper than the source data at any location. This is the most important constraint in nautical cartography that must be respected at all times.
2. *Legibility*: information should be shown in a clear and efficient way as an overdose of information slows down the chart reading process. Besides avoiding over-plot of symbols, legibility depends on the perceptibility threshold of map features, which is a function of scale, the minimum space between objects, their minimum sizes, and the minimum width of an area.
3. *Topology* (or *spatial*): the topology of the depicted chart elements must be correct, spatial relationships and relative distances between objects must be maintained e.g,, soundings must not be displaced, no gaps between skin of the earth features must exist, depth curves must not cross.
4. *Morphology* (or *shape*): morphological details of the seafloor (slope, roughness) must be maintained as much as possible.

With the above four constraints we describe the general requirement of the final chart to ensure that it is fit for the intended purpose. Constraints are used to describe the requirements that must be fulfilled by the final map or the individual tasks. They also act as the motors of generalization as the violation of them indicates the need for generalization and triggers one or more *generalization operators*. The problem with this *Constraint-based Modelling* (see Ruas and Plazanet, 1996) is that (i) it relies on the ability to formalize rules that define the condition-actions pairs, (ii) generalization of a map requires many rules to be defined in the rule base, especially if rules are set up for the relationships between map objects, and (iii) often operators have different goals and effects. Hence, when an operator / generalization algorithm is applied to solve a particular cartographic conflict, it may create other conflicts that have to be solved by subsequent generalization tasks.

## 4.2 Generalization Operators

add a short description of operators with examples in nautical cartography, e.g., simplification 🡪 depth contours ??

## 4.3 Graphic Limits (or Perceptibility Thresholds)

Essential to cartographic generalization is the minimum perceivable size of symbols and minimum separation between them. Many of the generalization operators are triggered from rules developed based on these limits. Visibility and separation limits depend on the physiology of eye and factors such as the level of human vision, the contrast and the colors used, the light conditions, and the map viewing distance.

Each map symbol, as any other physical object, subtends an angle at the eye of the observer, called *visual angle* (Figure 2). For instance, a 1-cm object viewed at 57 cm (that is approximately the distance at which we view our computer monitor) has a visual angle of, approximately 1 degree. Another approximation is that a thumbnail held at arm’s length subtends about 1 degree of visual angle, whereas the thumb subtends about two degrees.

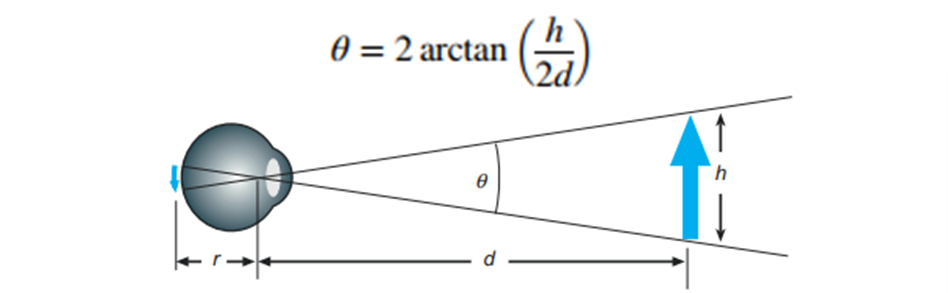


Figure 2: The visual angle of an object measured from the optical center of the eye (Ware, 2012).

Human eye has limits on its ability to see distinct details which is what the experts in human vision call *visual acuity*. It is a measure of the visual system’s ability to see distinctly the details of an object, such as a point, a line, a square, or a letter. The visual acuities are calculated in minutes and seconds of arc. Table 3 shows a list of the minimum acuities for different objects.

Table 3: Spatial resolution of human eye according to different sources and for different objects (Desjardins, 2014).



Using the equation in Figure X, one can translate the angles to lengths for the viewing distance. However, research (e.g., Mank, 2019) has showed that minimum symbol sizes in map context are larger than the theoretical minimum legible sizes such as those in Table X. For minimum sizes on maps cartographers usually point to those defined in a specification of the Swiss Society of Cartography (Figure 3).

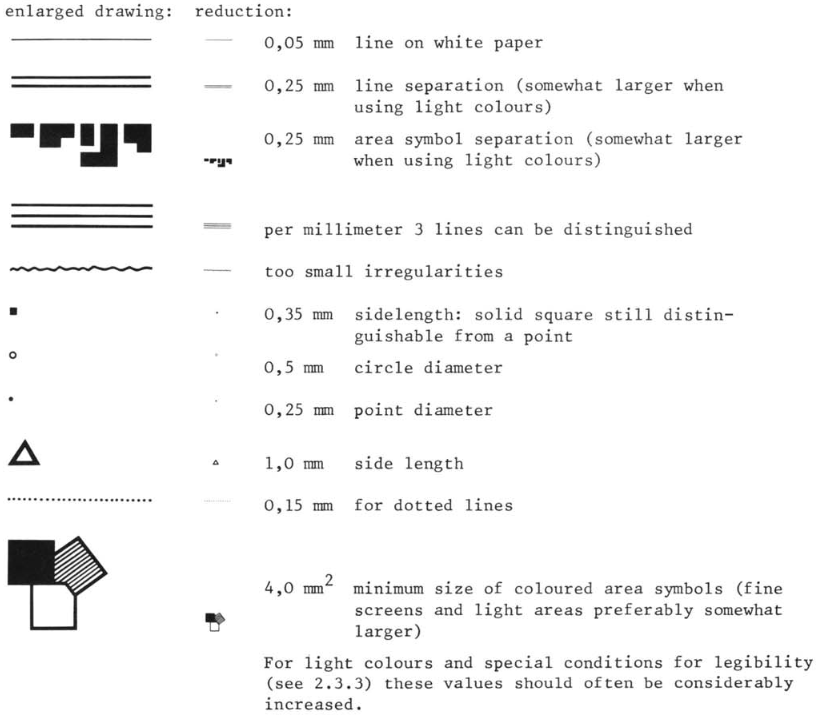


Figure 3. Minimum sizes for maps (Rytz et al., 1980).

Those minimal symbol sizes, established by the Swiss Society of cartography, are a useful guideline for high-frequency and high-contrast patterns. For lighter colors, it is recommended to use significantly larger symbols. Another source for the minimum symbol sizes on paper and digital maps is shown in the following Table 4.

Table 4. Minimum symbol sizes for paper and digital maps.

A screenshot of a cell phone

Description automatically generated

IHO S-4 and S-52 publications provide a number of guidelines /rules on symbol sizes and generalization of features that can be used for the evaluation of the quality of the final product.

Below are a few size examples in nautical cartography for paper charts:

* Shorelines: Man-made shoreline and structures (such as piers and breakwaters) are shown with an 0.15 mm lineweight, natural shoreline is symbolized by a solid black line of 0.25 mm, apparent shoreline and minimum sized islets are delineated with a 0.15 mm solid black line.
* Rivers and streams are shown with a 0.25-mm lineweight.
* Minimum size of isolated depths: ellipses of 3mm height and 2mm width (for depths of one digit, e.g., 8m), circle of 3mm diameter (for two-digit depths, e.g., 10m)
* font height = 2mm, font width = 1 mm (paper)
* Piers: shall not be charted if they are less than 0.8 mm in their greatest dimension.
* If the centerline separation of two adjacent parallel piers of different lengths is less than 0.3 mm at chart scale, the shorter of the two piers should not be charted unless it is identified as "essential".

Considering that ENCs are viewed from larger distances than paper maps / charts, the minimum sizes are to be adjusted accordingly. S-52 defines viewing distance to 70 cm, however, it also states that text should be readable from 1m. Therefore, font sizes and symbol sizes in ENCs are larger than those on paper charts (e.g., Figure 4).

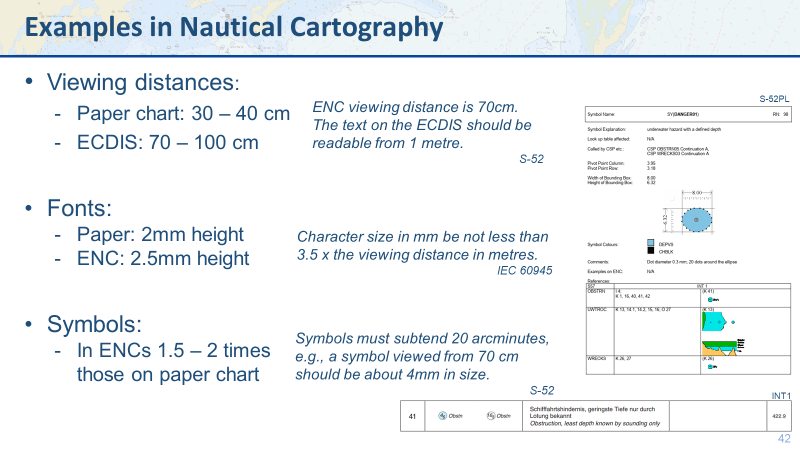


Figure 4: Examples for the concept of symbol sizes in nautical cartography (Kastrisios, 2020)

A few sizes examples for ENCs:

* Point density of line features must be higher than 0.3 mm.
* font height = 2.5mm, font width = 1.25 mm
* ……

These standards also provide guidelines for the generalization of features through scales (see, e.g., Section 4.4).

## 4.4 Compilation Guidelines

IHO Publications S-4, S-57, S-101 provide guidance on many compilation tasks. ………...

### 4.4.1 Depth discontinuities between surveys[[9]](#footnote-10)

Depth discontinuities between adjoining or overlapping source bathymetric surveys may be caused by:

* Surveys in areas of continually changing depth (see clause 5.7) conducted with a significant time gap between the surveys; or
* Adjoining areas having significant differences in the quality of bathymetric data (see clause 2.2.3.1).

It may not be possible to safely resolve significant depth discontinuity by interpolating approximate depth contours, which may compromise the ability for the compiler to adequately encode complete, non-overlapping Group 1 coverage of the area of the ENC cell covered by data. Where it is required to indicate these significant depth discontinuities, it should be done by encoding a “very narrow” UNSARE object (see clause 5.8.1).

The “very narrow area” should be at least 0.3mm in width at ENC compilation scale (see clause 2.2.6).

Remarks:

* An indication of the purpose of the UNSARE may be done by population of the attribute INFORM, for example Discontinuity between surveys.
* In order to provide an indication to the mariner of the more reliable encoded bathymetry in an area of continually changing depth, the attribute CATZOC should be downgraded for the M\_QUAL object (see clause 2.2.3.1) corresponding to the less reliable (or older) data.

### 4.4.2 Geometry of depth areas

**S-57**

Where areas are not closed on the source, it may be necessary to close these areas using edges without associated line objects. This is mandatory at the boundary of a cell (see Figure 5). As many depth areas as possible must be created using encoded depth contours.

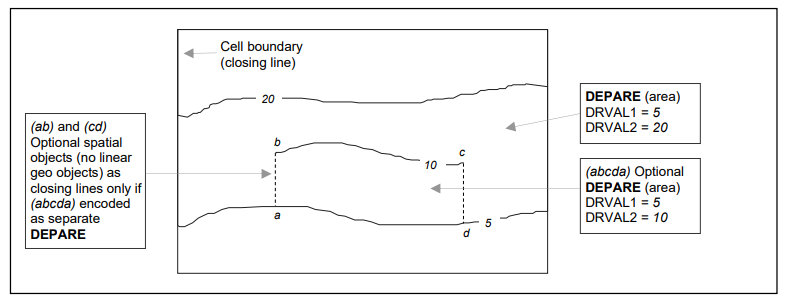


Figure 5: - Geometry of depth areas (source S-57 UoC)

**S-101**

Where surfaces are not closed on the source, it may be necessary to close these surfaces using edges without associated curve features. This is mandatory at the boundary of a dataset (see Figure 6). In Figure 6 below, the annotation “**min**” equates to the attribute **depth range minimum value** and the annotation “**max**” equates to the attribute **depth range maximum value**.

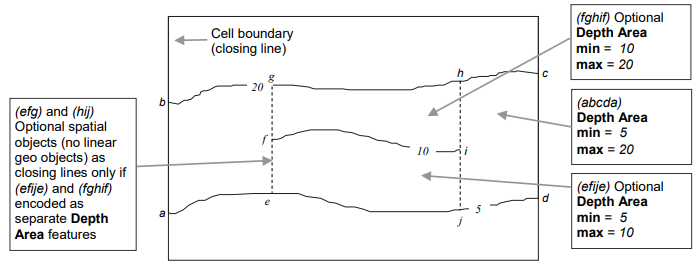


Figure 6: geometry of depth areas (source S-101 DCEG)

For short isolated sections of Depth Contour features such as (fi), it is up to the producing authority whether to encode the small areas (efije and fghif) as separate Depth Area features, or to encode only the curve (fi) as a floating Depth Contour feature within a single Depth Area (abcda) having attributes depth range minimum value = 5 and depth range maximum value = 20. NOTE: In Figure 3, if the optional Depth Area features are encoded, the depth area (abcda) will be split into two separate Depth Area features (abgea) and (jhcdj), both having depth range minimum value = 5 and depth range maximum value = 20.

**Relation between surveys, depth areas and Quality of Bathymetric Data**:

A single survey will measure the seafloor in a certain area relative to Chart Datum. Resulting from this single survey, one or more depth areas will be created. Depth areas have two attributes (depth range minimum value, depth range maximum value). A collection of surveys will measure a wider area of the seafloor, resulting in probably more depth areas, depending on the slope of the seabed. This is called an association (*semantic relationship between two or more classifiers that specifies connections among their instances*).

In practice many depth areas share the same Quality of Bathymetric Data value. In that case, we speak of an aggregation (*special form of association that specifies a whole-part relationship between the aggregate (whole) and a component part*). If a survey is carried out to a specific depth which is not bounded by a depth contour line, than the resulting Depth Area will have two separate Quality of Bathymetric Data values.

ENCs should form a seamless coverage in the navigable waters of the producer’s area of responsibility. However, it is often impractical to do so for all ECDIS display scales, and therefore S-101 ENCs declare a scale range, which dictate between what scales the data can be used.

When assigning meaningful Quality of Bathymetric Data information in an ENC, the Hydrographic Office should take into account the maximum display scale and minimum display scale. [[10]](#footnote-11)

|  |  |
| --- | --- |
| maximum display scale | minimum display scale |
| 700,000 | > 700,000 |
| 350,000 | 700,000 |
| 180,000 | 350,000 |
| 90,000 | 180,000 |
| 45,000 | 90,000 |
| 22,000 | 45,000 |
| 12,000 | 22,000 |
| 8,000 | 12,000 |
| 4,000 | 8,000 |
| 3,000 | 4,000 |
| 2,000 | 3,000 |
| 1,000 | 2,000 |

The largest chart scale available will have the maximum level of detail (LoD) and distinctiveness of areas of Quality of Bathymetric Data. The Hydrographic Office should assign values of Quality of Bathymetric Data at the highest level of detail possible. Then lower scale charts will inherit these values. Adjacent areas of Quality of Bathymetric Data will merge together (aggregation process) in smaller scale charts. The lowest value of two (or more) merging areas should be the aggregated value in the smaller scale chart.

## 4.5 Validation Checks

While managing the quality during generalization is a good idea, it is not always possible or practical. If we were to track all possible problems after each application of each algorithm, the performance would quickly become unacceptable. Similarly, checking the conformance to standard for the input data for each algorithm would result in the same hit on performances. Not performing systematic full checks opens the door to problems though, as a slightly odd result from one algorithm ends up as input to the next which can then produce unexpected results. It makes sense to have an initial pre-process that checks that the source data is of suitable quality, and a post-process that checks that the result meets the requirements. These are completed by targeted local checks performed within generalization steps. (Regnauld, 2015)

IHO Special Publication S-58 (IHO, 2018) contains a list of validation checks that aim to ensure that ENCs are compliant with S-57. Software that performs S-58 validation (e.g., SevenCs Analyzer, ESRI ArcGIS for Maritime, Teledyne CARIS S-57 Composer, C-MAP dKart Inspector) provide reports for errors among the Group 1 and 2 objects of the ENC in question as well as errors with objects in the adjoining ENCs. The classification of errors is in three categories according to the risk they pose for the safety of navigation (i.e., warnings, errors, and critical errors). In detail, a warning is “an error which may be duplication or an inconsistency which will not noticeably degrade the usability of an ENC in ECDIS”, an error “may degrade the quality of the ENC through appearance or usability but which will not pose a significant danger when used to support navigation”, whereas a critical error “would make an ENC unusable in ECDIS through not loading; or causing an ECDIS to crash; or presenting data which is unsafe for navigation” (IHO, 2018).

The list of validation checks in S-58 is not exhaustive; new validation checks are added and others are upgraded/ downgraded from one category to another as a result of the lessons learned from real-world situations and research in the field.

…….

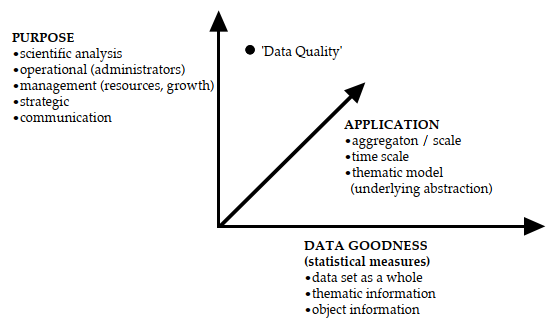
## 4.XX Temporary Notes

FROM:

Beard etal. (1991)

One of the most commonly cited components of data quality is error. Commonly  
recognized errors include those associated with data collection (source error) and  
the processing of data (process error). Process errors have proven difficult to  
analyze in many cases, for example in studies of digitizing error, or in modeling  
error associated with soil mapping (Fisher 1991). In statistics, the concept of  
Least Squares Error has been applied to determine reliability (or what is called  
ÔconfidenceÕ) in hypothesis testing. A third error component (use error) defined  
by Beard (1989) is associated with the appropriate application of data or data  
products.

The role of lineage was described as tracking and providing information about quality but not serving as a quality measure itself. Objectives, goals, and limitations of the data belong under lineage and the lineage report should state or reference all assumptions, definitions, and procedures applied in data collection. ----compilation as lineage and important factor of DQ. Do we attempt to measure error directly or by describing the process (by examining the lineage). In terms of the measurement view the group discussed whether it was possible to test against the truth? This creates the problem that we must deal with ambiguity in the truth and the fact that reality is a social construct. One option is to test a database against a replicate (db') rather than truth.



The error

in raw data and error which occurs during the manipulation of data cannot

always be controlled. For most data there is no way to obtain error measures, and

a particular problem with spatial data is that we typically do not have

replications. One assessment of quality can be made through knowledge of

expertise (of the data collector for example), but this is not easily measured or

documented.

Data quality is not merely a data collection issue. It is part of the whole processing scheme. For example, selecting the wrong map scheme or wrong statistics will effect quality. Data quality should be managed through processing to the generation of final products. The idea with respect to quality information is to try to come up with a model of quality that resembles how quality behaves during processing and product generation.

# 5. Data Quality Components

Information on the quality of geographic data allows a data producer to evaluate how well a dataset meets the criteria set forth in its product specification and assists data users in evaluating a product’s ability to satisfy the requirements for their particular application. For the purpose of this evaluation, clearly defined procedures are used in a consistent manner. AGI Guidelines for geographic information content and quality (AGI, n.d.) provide guidance to those who work with geographic information to ensure that it is fit for its intended purpose. Many parts of the information on data quality in this section comes from ISO 19157 (2013). Working with data quality includes:

* understanding of the concepts of data quality related to geographic data. This section provides a description of data quality concepts used to establish the components for describing the quality of geographic data;
* defining data quality conformance levels in data product specifications or based on user requirements. Establishment of data product specifications is described in ISO 19131:2007;
* specifying quality aspects in application schemas;
* evaluating data quality;
* reporting data quality

~~These data quality measures provide descriptors of the quality of geographic data through comparison with the universe of discourse. The use of incompatible measures makes data quality comparisons impossible to perform. This International Standard standardizes the components and structures of data quality measures and defines commonly used data quality measures.~~

A data quality evaluation can be applied to dataset series, a dataset, or a subset of data within a dataset, sharing common characteristics so that its quality can be evaluated. Data quality is described using data quality elements. Data quality elements and their descriptors are used to describe how well a dataset meets the criteria set forth in its data product specification or user requirements and provide quantitative quality information. When data quality information describes data that have been created without a detailed data product specification or with a data product specification that lacks quantitative measures and descriptors, the data element may be evaluated in a non-quantitative subjective way as a descriptive result for each element. Some quality related information is provided by purpose, usage and lineage. Quantitative data quality elements may have associated quality (termed *metaquality*). To facilitate comparisons, it is essential that the results of the quality reports are expressed in a comparable way and that there is a common understanding of the data quality measures that have been used. Figure 7 provides an overview of the model of data quality.

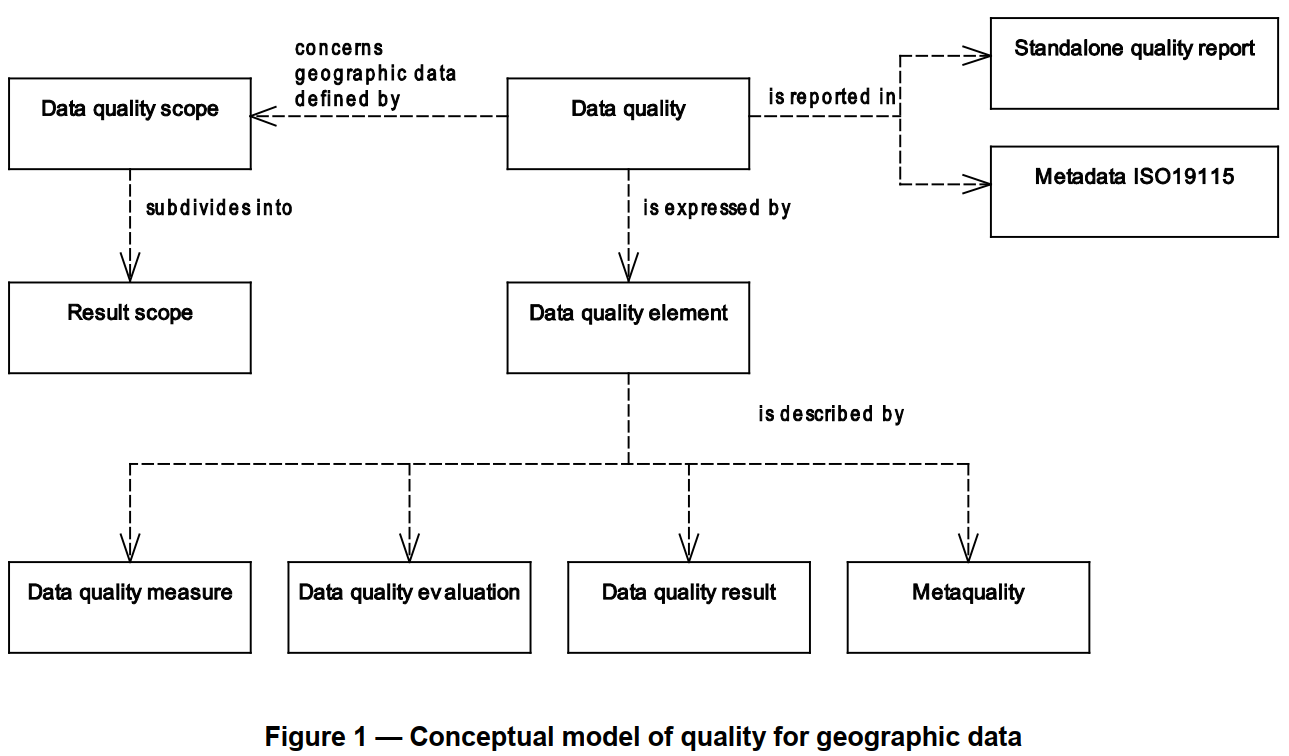


Figure 7: Conceptual Model of quality for geographic data

## 5.1 Quality Unit

When describing the quality of geographic data, different quality elements and different subsets of the data may be considered. In order to describe these, *data quality units* are used. A data quality unit is the combination of a *scope* and data quality *elements*. The *scope* of the data quality unit(s) specifies the extent, spatial and/or temporal, and/or common characteristic(s) that identify the data on which data quality is to be evaluated. One data quality scope shall be specified for each data quality unit. One data quality report (metadata or standalone quality report – see Section 6.3) may encompass several data quality units, since scopes are often different for individual data quality elements. These different scopes may be, for example, spatially separate, overlapping or even sharing the same extents.

The following are examples of what defines a data quality scope (see also MD\_Scope in ISO 19115-1):

a) a dataset series;

b) a dataset;

c) a subset of data defined by one or more of the following characteristics:

1) types of items (sets of feature types, feature attributes, feature operations or feature relationships);

2) specific items (sets of feature instances, attribute values or instances of feature relationships);

3) geographic extent;

4) temporal extent (the time frame of reference and accuracy of the time frame).

## 5.2 Quality Elements

Data quality shall be described using the *data quality elements*. Data quality elements evaluate the difference between the dataset and the *universe of discourse* (i.e., the perfect dataset that corresponds to the data product specification). A data quality element is a component describing a certain aspect of the quality of geographic data. Six different categories are defined, i.e., Completeness, Logical Accuracy, Positional Consistency, Thematic Accuracy, Temporal Quality, and Usability (Figure 8). In detail:

**Completeness** is the presence or absence of features and along with positional accuracy and thematic accuracy (explained bellow), are those that are used to describe how the dataset relates to the universe of discourse. It consists of:

* *commission* (excess data present in a dataset), and
* *omission* (data absent from a dataset).

**Logical consistency** is the degree of adherence to logical rules of data structure, attribution, and relationships (data structure can be conceptual, logical, or physical). Is the only one among the quality elements that can be fully evaluated without ground truth knowledge It consists of:

* *conceptual* *consistency* – adherence to rules of the conceptual schema.
* *domain* *consistency* – adherence of values to the value domains.
* *format* *consistency* – degree to which data is stored in accordance with the physical structure of the dataset.
* *topological consistency* – correctness of the explicitly encoded topological characteristics of a dataset.

The **positional accuracy** is the accuracy of the position of features within a spatial reference system. It consists of:

* *absolute* or *external accuracy* – closeness of reported coordinate values to values accepted as or being true.
* *relative* or *internal accuracy* – closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true.
* *gridded data positional accuracy* – closeness of gridded data spatial position values to values accepted as or being true. The accuracy of gridded data may be described using the same data quality measures as for the horizontal positional uncertainty. The band values in rasters may be described using the quantitative attribute accuracy of thematic accuracy.

**Thematic accuracy** is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. It consists of:

* *classification correctness* – comparison of the classes assigned to features or their attributes to a universe of discourse (e.g., ground truth or reference data);
* *non-quantitative attribute correctness* – measure of whether a non-quantitative attribute is correct or incorrect.
* *quantitative attribute accuracy* – closeness of the value of a quantitative attribute to a value accepted as or known to be true.

**Temporal quality** is defined as the quality of the temporal attributes and temporal relationships of features. Temporal quality consists of a mix of data quality elements that partly is dependent upon logical rules (comparable to logical consistency) and partly needs ground truth knowledge to be evaluated (in similar way as *completeness* and the *accuracy* elements):

* *accuracy of a time measurement* – closeness of reported time measurements to values accepted as or known to be true.
* *temporal consistency* – correctness of the order of events in the dataset.
* *temporal validity* – validity of data with respect to time. The temporal validity may be treated with the same data quality measures as for other domain specific attribute values.

**Usability element** is used for a quality evaluation based on user requirements which cannot be covered by the other five data quality categories. It may also be used to provide an aggregation result where results from several data quality categories are *aggregated* (for example, overall conformity to one specification). That is because an evaluation based on a single data quality element is often not sufficient. The quality of a dataset may be represented by one or more *aggregated data quality result* (ADQR). The ADQR combines quality results from data quality evaluations based on different data quality elements or different data quality scopes. A dataset may be deemed to be of an acceptable aggregate quality even though one or more individual data quality results fails acceptance. Aggregation should therefore only be used when compelling reasons exist. The meaning of the aggregate data quality result should always be made clear.  
As the ADQR may be difficult to fully understand, the meaning of the aggregate data quality result should be understood before drawing conclusions based on aggregate data quality results for the quality of the dataset. Examples of methods that may be used for producing an ADQR are given in Section J.2 to J.4 of ISO (2013), e.g., 100% Pass/Fail, Weighted Pass/Fail, and Max/Min Value.

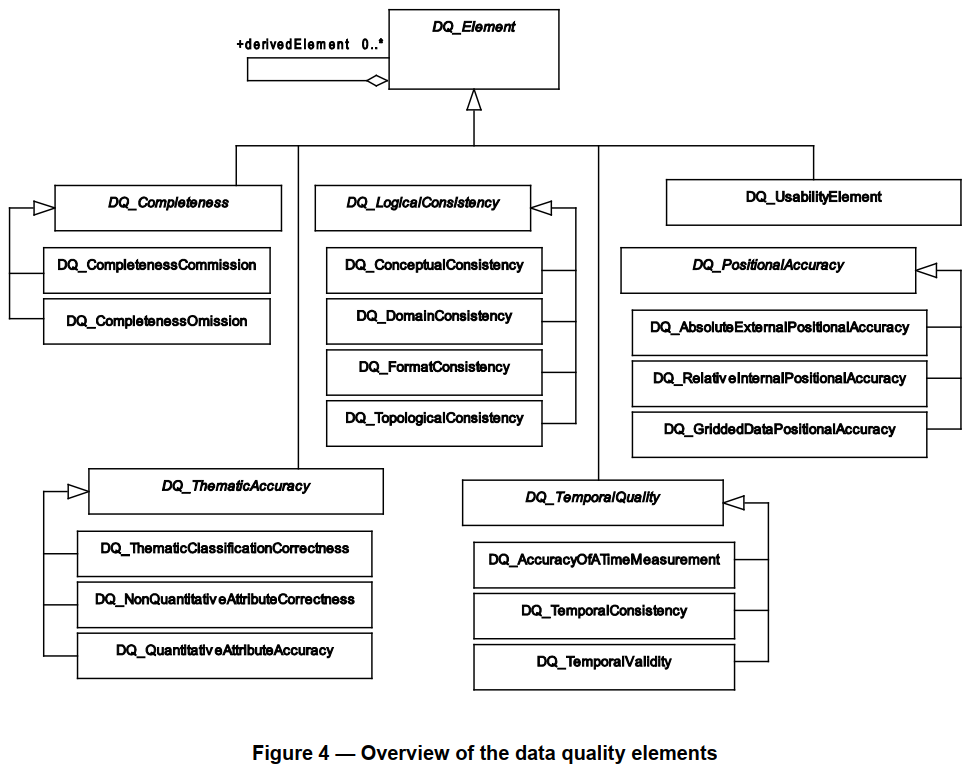


Figure 8: Overview of the data quality elements

Table 5 shows the data quality elements relevant to depth.

Table 5: Data quality elements for depth

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| DQ element | DQ sub-element | Definition | Evaluation scope | Application to spatial representation types | |
| Vector | Grid |
| Completeness | Commission | excess data present in the dataset. | dataset / dataset series | X |  |
| Omission | data absent from the dataset. | dataset / dataset series / spatial object type | X | X |
| Logical consistency | Conceptual consistency | adherence to rules of the conceptual schema | spatial object / spatial object type | X | X |
| Domain consistency | adherence of values to the value domains | spatial object / spatial object type | X |  |
| Format consistency | degree to which data is stored in accordance with the physical structure of the dataset, as described by the scope | dataset / dataset series | X | X |
| Topological consistency | correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope | spatial object type / dataset series / dataset | X |  |
| Positional accuracy | Absolute (external) accuracy | closeness of reported coordinate values to values accepted as or being true | spatial object / spatial object type / dataset series / dataset | horizontal component |  |
| vertical component | vertical component |
| Positional accuracy | Gridded data position accuracy | closeness of gridded data position values to values accepted as or being true | spatial object / spatial object type / dataset series / dataset |  | horizontal component |
| Thematic Accuracy |  |  |  |  |  |
| Temporal Quality |  |  |  |  |  |
| Usability |  |  |  |  |  |

## 5.3 Quality Descriptors

Data quality elements and their descriptors are used to evaluate how well a dataset meets the criteria set forth in its data product specification or user requirements and provide quantitative quality information. An evaluation of a data quality element is described by the type of evaluation (*measure),* the procedure used to evaluate the measure (*method), and* the output of the evaluation (*result)*. In detail:

### 5.3.1 Measure

Measure is the type of evaluation. A data quality element should refer to one measure only. An example of measure can be the percentage of the values of an attribute which are correct. A single data quality measure might be insufficient for fully evaluating the quality of the data specified by a data quality scope and providing a measure of quality for all possible utilizations of a dataset. A combination of data quality measures can give useful information. Multiple data quality measures may be reported for the data specified by a data quality scope.

To facilitate dataset comparisons, it is necessary that the results in the data quality reports are expressed in a comparable way and that there is a common understanding of the data quality measures that have been used. In order to make evaluations and data quality reports (metadata or a standalone quality report) from different sources comparable, standardized data quality measures shall be used when possible.

### 5.3.2 Method

The evaluation method is the procedure used to evaluate the measure. Data quality evaluation method should be included for each applied data quality measure. Different evaluations are often used for the various data quality elements.

Based on the guidelines of ISO 19157 and its predecessors, the ISO 19113 and ISO 19114, a number of studies have tried to develop methods for assessing data quality and its adherence to the decided measures (see Section 7).

### 5.3.3 Result

The result is the output of the evaluation method used to evaluate the specific type of evaluation of the quality element. Quality frequently differs between various parts of the dataset for which quality is evaluated. Therefore, several evaluations may be applied for the same data quality element to more completely and in more detail describe quantitative data quality. To avoid repeating the measure and evaluation procedure descriptions in several instances of data quality element, several results with individual result scopes can be used. For example, A dataset may contain features of identical type whose positions have been established with separate methods yielding different positional accuracies. The same quality evaluation method and the same measure are however applied for the whole dataset, and provide different results depending on the data acquisition method. In this case, it may be desirable to have several results with individual result scopes (the area covered by each data acquisition method) and one data quality scope (the dataset). A result can be *quantitative*, *conformance*, *descriptive*, or *coverage* result:

* **Quantitative** result **–** the outcome in numerical form such as statistics, percentages, etc.
* **Conformance** result **–** the outcome of comparing the value or set of values obtained from applying a measure to the data specified by a data quality scope with a specified acceptable conformance quality level.
* **Descriptive** result **–** thesubjective evaluation of an element, expressed with a textual statement, when a quantitative result is not possible (e.g., with thematic and geoscientific observations).
* **Coverage** result **–**the result of a data quality evaluation, organized as a coverage. This is documented in ISO 19115-2:2009.

Below are some measure and result examples for the quality elements of Section 5.2.

**Completeness :**

**Commission**: Number OR rate of excess items, i.e., items that should not have been in the dataset and are present, e.g., “2” as a result for “two points (or lines or polygons) collected on top of each other”; the presence of buildings under 5 m2 are reported as commission if only buildings that are bigger than 5 m2 should be included in the dataset.

**Omission**: Number OR rate of missing items, i.e., items that should have been in the dataset and are missing, e.g., “2” as a result for “10 houses are in the dataset although only 12 exist within the universe of discourse”; the absence of buildings under 5 m2 are reported as omission if all buildings should be included in the dataset.

**Logical Consistency:**

**Conceptual**: number of items not compliant with the rules of the relevant conceptual schema. If the conceptual schema explicitly or implicitly describes rules, these rules shall be followed, e.g., invalid placement of features within a defined tolerance, duplication of features and invalid overlap of features, e.g., “invalid placement of airport inside lake” or “two lights with identical attribution and within search tolerance ε”.

**Domain**: value domain conformance. Example: True as result for an item that is conforming to its value domain.

**Format**: physical structure conflicts number. Example: 5 as a result for features type encoded with more than 3 characters when the requirement in the specification is 3.

**Topological**: missing connections (e.g., Figure 9 for undershoot, Figure 10 for overshoot), self-intersect (Figure 11), self-overlap (Figure 12).

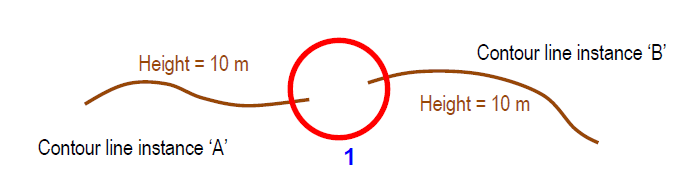


Figure 9: Example of missing connections due to undershoots.

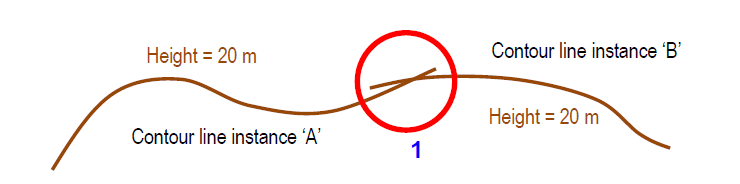


Figure 10: Example of rate of missing connections due to overshoots.

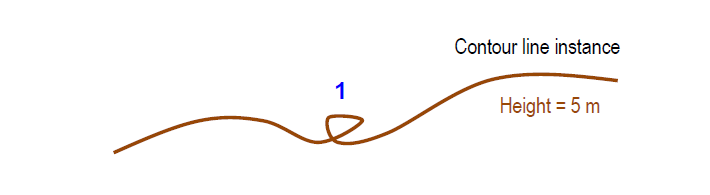


Figure 11: Example of rate of invalid self-intersect errors.

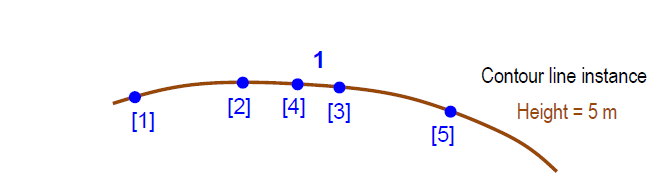


Figure 12: Example of rate of invalid self-overlap errors (• vertices [digitized order]).

**Positional Consistency:**

**Absolute** or **external**: horizontal accuracy at 95 % significance level, e.g., 5 m;

**Relative** or **internal** accuracy – vertical error. That is the evaluation of the random errors of one relief feature to another in the same dataset or on the same map/chart. It is a function of the random errors in the two elevations with respect to a common vertical datum, e.g., 1.1 m;

**Gridded** **data** positional accuracy – the band values in rasters must be within 95%.

**Thematic Accuracy:**

**Classification correctness** – number of incorrectly classified features, e.g., 12 incorrectly classified features.

**Non-quantitative** **attribute** correctness – rate of incorrect attribute values, e.g., 12%.

**Quantitative attribute** accuracy – attribute value uncertainty at 90 % significance level.

**Temporal Quality:**

**Accuracy of a time measurement** – time accuracy at 68 % significance level, the start date (DATSTA) field cannot contain a value in the future.

**Temporal consistency** – chronological error, i.e., the indication that an event is incorrectly ordered against the other events (e.g., True – 2 events in the dataset are not ordered correctly), the end date (DATEND) shall be the same as or after start date (DATSTA).

**Temporal validity** –date value shall be in a specific format, e.g., “CCYYMMDD” (20210423).

## 5.4 Metaquality

Metaquality provides quality information about quality evaluation. Metaquality elements are a set of quantitative and qualitative statements about a quality evaluation and its result. The knowledge about the quality and the suitability of the evaluation method, the measure applied, and the given result may be of the same importance as the result itself. Metaquality may be described using *confidence*, *representativity*, and *homogeneity*:

* **Confidence** – trustworthiness of a data quality result.
* **Representativity** – degree to which the sample used has produced a result which is representative of the data within the data quality scope.
* **Homogeneity** – expected or tested uniformity of the results obtained for a data quality evaluation.

A metaquality element is described by the same descriptors as for the quality element (measure, evaluation method and result, with the addition of *related quality element* (i.e., the element on which the metaquality element applies)*.*

The main difficulty is to aggregate these measures in a way that can help decide if a set of data is fit for a particular use. While the criteria related to clean geometries (no double points, well-formed polygons, no self-intersection, etc.) and topology (no overshoots or undershoots at junctions, no overlaps, no double points; well-formed polygons, no self-intersection, etc.) can fairly easily be checked and sometimes automatically fixed, the others are more difficult to check, and would require a reference dataset. For example, looking at a single dataset, it is often impossible to detect that a feature is missing or misclassified. Exceptions exist, when semantic inconsistencies can be detected, for example, when a section of river has been misclassified as a road. The combination of the fact that a section of road is isolated from the rest of the road network and the fact that this section also connects two dead end river nodes could lead to the conclusion that this section of road has been misclassified and should be reclassified as a river. Writing such rules is extremely time consuming though, as combinations of potential misclassifications and the context in which they occur are almost endless. Such rules are worth writing once we have identified that a particular type of misclassification occurs frequently. Some simpler types of checks can still be done though, like checking that all the features have a valid height attribute if one is expected. (Regnauld, 2015)

# 6. Data Quality Evaluation

Datasets are continually being created, updated and merged with the result that the quality or a component of the quality of a dataset may change. The quality of a dataset can be affected by three conditions:

* when any quantity of data is deleted from, modified or added to a dataset,
* when a dataset’s data product specification is modified or new user specified data quality requirements are identified,
* when the real world has changed.

The first condition, a modification to a dataset, may occur frequently. Many datasets are not static. There is an increase in the interchange of information, the use of datasets for multiple purposes and an accompanying update and refinement of datasets to meet multiple purposes. If the reported quality of a dataset is likely to change with modifications of the dataset, the quality of this dataset should be reassessed and updated as required when changes occur. Complete knowledge of all applicable data quality elements should be available when a dataset is created. Only the data producer’s usage (assuming the data producer actually uses the dataset) of a dataset can initially be reported. There is a reliance on data users to report uses of a dataset that differ from its intended purpose so that continual updates to this particular data quality overview element can be made to reflect occurring, unforeseen uses.

The second condition, a modification to a dataset’s data product specification, is most likely to occur before initial dataset construction and prior to the release of quality information. It is conceivable, however, that as a dataset is used, its data product specification is updated so that future modifications to the dataset will better meet the actual needs. As the data product specification changes, the quality of the current dataset also changes. The quality information for a dataset should always reflect the current dataset given its current data product specification.

The third condition, a change of the real world, occurs continuously. Changes may be caused by natural phenomena such as movements in the earth’s crust or erosion, but it is most often a result of human activity. Changes are often very rapid and dramatic. For this reason, the date of data collection is equally important as the date of quality evaluation when judging the quality of a dataset. In some cases, when known, even the rate of change is of interest. The update frequency of the dataset may also be of interest in some cases. However, it might not be possible to create a new data quality report every time the real-world changes.

The quality evaluation process is a sequence of steps followed to produce a quality evaluation result. When the geographic data evaluated is heterogeneous with different quality for different parts, tests should be applied to suitable parts of the data.

## 6.1 Process

Figure 13 illustrates a possible workflow for evaluating data quality and Table 6 the process steps (see ISO 19157 Section E.3 for a study case).

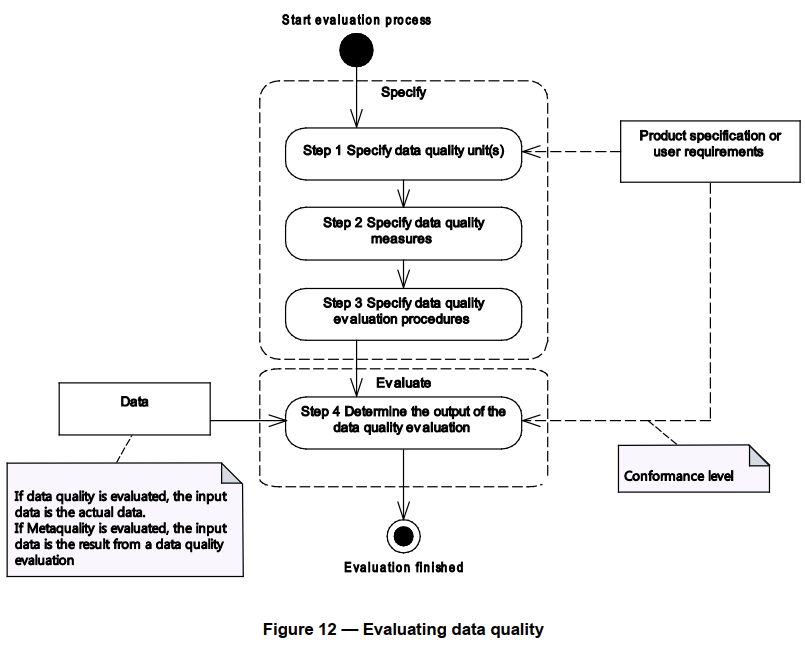


Figure 13. Data quality evaluation workflow (ISO, 2013)

Table 6. Evaluation process steps (ISO, 2013)

|  |  |  |
| --- | --- | --- |
| Step | Action | Description |
| 1 | Specify data quality unit(s) | A data quality unit is composed of a scope and quality element. All data quality elements relevant to the data for which quality is to be described should be used. |
| 2 | Specify data quality measures | If applicablea a measure should be specified for each data quality element. |
| 3 | Specify data quality evaluation procedures | A data quality evaluation procedure consists of applying one or more evaluation methods |
| 4 | Determine the output of the data quality evaluation | A result is the output of applying the evaluation |

A data quality evaluation procedure comprises one or more data quality evaluation methods. Data quality evaluation methods can be divided into two main classes, *direct* and *indirect*:

* **Direct** evaluation method is a method of evaluating the quality of a dataset based on inspection of the items within the dataset.
* **Indirect** evaluation methods infer or estimate data quality using information on the data such as lineage. An indirect evaluation method is a method of evaluating the quality of a dataset based on external knowledge or experience of the data product and can be subjective. This external knowledge may include, but is not limited to, one or more non-quantitative quality information usage, lineage, and purpose (see ISO 19115-1:2013) or other data quality reports on the dataset or data used to produce the dataset. Data quality may be estimated, for example from knowledge about the source, tools and methods used for the capturing of the data and evaluated against procedures and specifications worked out for this product. Indirectly evaluated data quality may also be based on experience alone.

Direct evaluation methods should be used in preference to indirect evaluations. The direct evaluation methods are further sub classified, by the source of the information needed to perform the evaluation, in *internal* and *external*.

* **Internal** **direct** data quality evaluation uses only data that resides in the dataset being evaluated.
* **External direct** quality evaluation requires reference data external to the dataset being tested.

For both external and internal evaluation methods, either *full inspection* or *sampling* may be used:

* **Full inspection** tests every item in the population specified by the data quality scope. It is most appropriate for small populations or for tests that can be accomplished by automated means.
* **Sampling** means that tests are performed on subsets of the geographic data defined by the data quality scope. For detailed information on sampling strategies, the reader is referred to ISO 19157 Section F.4.

Quality evaluation procedures may be used in different phases of a product’s life cycle, for instance:

* **Development of a data product specification or user requirements** – When developing a data product specification or defining user requirements, quality evaluation procedures may be used to facilitate the establishment of conformance quality levels that should be met by the final product. A data product specification or user requirements may include conformance quality levels for the data and quality evaluation procedures to be applied during production and updating.
* **Quality control during dataset creation** – At the production stage, the producer may apply quality evaluation procedures, either explicitly established or not contained in the data product specification, as part of the process of quality control. The description of the applied quality evaluation procedures, when used for production quality control, may be reported as lineage metadata including, but not necessarily limited to, the quality evaluation procedures applied, conformance quality levels established and the results.
* **Inspection for conformance to a data product specification** – On completion of the production, a quality evaluation process may be used to produce and report data quality results. These results may be used to determine whether a dataset conforms to its data product specification or not. If the dataset passes inspection (composed of a set of quality evaluation procedures), the dataset is considered to be ready for use. The outcome of the inspection will be either acceptance or rejection of the dataset. If the dataset is rejected, then, after the data have been corrected, a new inspection will be required before the product can be deemed to be in conformance with the data product specification.
* **Evaluation of dataset conformance to user requirements** – Quality evaluation procedures may be used to establish if a dataset meets the conformance quality levels specified in user requirements. Indirect as well as direct methods may be used in analyses of dataset conformance to user requirements.
* **Quality control during dataset update** – Quality evaluation procedures are applied to dataset update operations, both to the items being used for update and to benchmark the quality of the dataset after an update has occurred.

## 6.2 Order

When evaluating geographic data, one individual error may influence several data quality elements. The usual order that is followed is:

1. Format consistency: The very first to be evaluated is the readability (or interpretability) of the data to decide whether it is possible to decode/read/understand the data or not. Not interpretable data should be reported and ignored in the further evaluation. The result of the format consistency should describe which parts of the data are not readable.
2. Logical consistency: Decide if the rules set up for the dataset are followed. Parts of the dataset not conforming to the rules should be ignored in the further evaluation.
3. Completeness: The next step in the evaluation is the feature existence aspect covered by  
   completeness. To evaluate this, the features in the actual dataset and the ground truth data are compared, and commissions and omissions reported.
4. Accuracy (positional, thematic, and temporal aspects): The last step in the evaluation covers the accuracy aspect, measuring the deviation between actual and ground truth feature properties. These measurements can be based only on parts of the dataset present in both the actual dataset and the universe of discourse.

## 6.3 Reporting

Data quality reporting can aid discovery and encourage use of the dataset, demonstrate the compliance to a data product specification or to user requirements, permit downstream judgements about the quality of information derived from the data set, and permit optimal decision making. Data quality shall be reported as *metadata*. In order to provide more details than reported as metadata, a *standalone quality report* may additionally be created, however, it shall not replace the metadata. These two mechanisms complement each other by allowing the reporting of data quality evaluation with different levels of detail:

* **Metadata**: The metadata aims at providing short, synthetic, and generally structured information to enable metadata interoperability and web services usage (e.g., M\_QUAL).
* **Standalone quality report:** The standalone quality report may be used to provide fully detailed information about the data quality evaluation (e.g., survey report). It should contain sufficient information to meaningfully describe the relevant aspects of data quality and their results. This may take the form of references to supporting documentation such as a data product specification or measure catalogue. It is important to present the quality information in a succinct, easily understood and easily retrievable way.

For example, in the case of aggregation of different quality results, the standalone quality report will provide full information on the original results (with evaluation procedures and measures applied), the aggregated result and the aggregation method whereas the metadata may describe only the aggregated result with a reference to the original results described in the standalone quality report.

# 7. Evaluation Techniques

## 7.1 General

To assess data quality, several methods have been used in the literature.

**Completeness:**

Completeness refers to the relationship between the objects in the database and the ‘abstract universe’ of all such objects. Selection criteria, definitions, and other mapping rules used to create the database are important determinants of completeness. This requires a precise description of the abstract universe since the relationship between the database and the abstract universe cannot be ascertained if the objects in the universe cannot be described. The abstract universe can be defined in terms of a desired degree of abstraction and generalisation (i.e., a concrete description or specification for the database). This leads to the realisation that there are in fact two different types of completeness. *Data completeness* is a measurable error of omission observed between the database and the specification. Data completeness is used to assess data quality, which is application independent. Even highly generalised databases can be complete if they contain all of the objects described in the specification. *Model completeness* refers to the agreement between the database specification and the abstract universe that is required for a particular database application (Brassel *et al.,* 1995). Model completeness is application-dependent and therefore an aspect of fitness-for-use. It is also a component of ‘semantic accuracy’ (Salgé 1995). The  
definitions of completeness given above are examples of ‘feature or entity completeness’. In  
addition we can identify ‘attribute completeness’ as the degree to which all relevant attributes of a  
feature have been encoded. A final type of completeness is ‘value completeness’ which refers to  
the degree to which values are present for all attributes (Brassel et al 1995). (Veregin, 2005)

Specifically to evaluate the completeness of crowd-sourced data, researchers have used grid-based length comparison against authoritative data (see e.g., Haklay, 2010; Ludwig *et al*., 2011; Zielstra & Zipf, 2010; Ciepluch *et al*., 2011; Forghani & Delavar, 2014), comparison of number of features (Girres & Touya, 2010; Jackson *et al*., 2013), comparison of total length or total area (Girres and Touya, 2010; Kounadi, 2009; Koukoletsos *et al*., 2011; Fan *et al*., 2014; Kalantari and La, 2015; Arsanjania and Vaz, 2015), completeness measure (Mashhadi *et al*., 2014) and completeness index (Arsanjani *et al*., 2015). Antoniou & Skopeliti (2015).

**Logical Consistency:**

Logical consistency measures the consistency of different database objects with other objects of the same theme (intra-theme consistency) or objects of other themes (inter-theme consistency) (Girres and Touya, 2010). In addition to this, hierarchical ordering and outliers spotting are used to check administrative data integrity (Ali and Schmid, 2014), mathematical techniques determine topological consistency (Corcoran *et al*., 2010) and spatial similarity in multi-representation are used to assess topological relationships (Hashemi *et al*., 2015). Moreover, a number of techniques deal with semantic similarity between the tags (Vandecasteele and Devillers, 2015) and the identification of entities with inappropriate classification (Ali *et al*., 2014). The improvement of the semantic quality can be also achieved by using ontologies such as OSMonto in data tagging (Codescu *et al*., 2011) or a tag recommendation system as OSMantic (Vandecasteele and Devillers, 2015), which automatically suggests relevant tags to contributors during the editing process. Antoniou & Skopeliti (2015).

**Positional Accuracy:**

Measurement of spatial accuracy depends on dimensionality. Metrics are well defined for point  
entities, but widely accepted metrics for lines and areas have yet to be developed. For points, error is usually defined as the discrepancy (normally Euclidean distance) between the encoded location  
and the location as defined in the specification. Various metrics have been developed to summarise spatial error for sets of points. One such metric is mean error, which tends to zero when ‘bias’ is absent. Bias refers to a systematic pattern of error (e.g., error arising from map misregistration). When bias is absent error is said to be random. Another common metric is root mean squared error (RMSE), which is computed as the square root of the mean of the squared errors (see Beard and Buttenfield, 2005). RMSE is commonly used to document vertical accuracy for DEMs. RMSE is a measure of the magnitude of error but it does not incorporate bias since the squaring eliminates the direction of the error. For lines and areas the situation is more complex since there is no simple statistical measure of error that can be adopted from statistics. Errors in lines arise from the errors in the points that define those lines. However, as these points are not randomly selected the errors present at points cannot be regarded as somehow typical of errors present in the line (Goodchild 1991b). Error is usually defined for lines using some variant of the *epsilon band*. The epsilon band is defined as a zone of uncertainty around an encoded line within which there is a certain probability of observing the ‘actual’ line. As yet there is no agreement as to the shape of the zone and the distribution of error within it. Early models assumed that the zone was a uniform ‘sausage’ within which the distribution of error was uniform (Blakemore 1983; Chrisman 1982). More recent studies show that both the distribution and the band itself might be non-uniform in shape (Caspary and Scheuring 1993; Honeycutt 1986) (Figure 14). (Veregin, 2005)

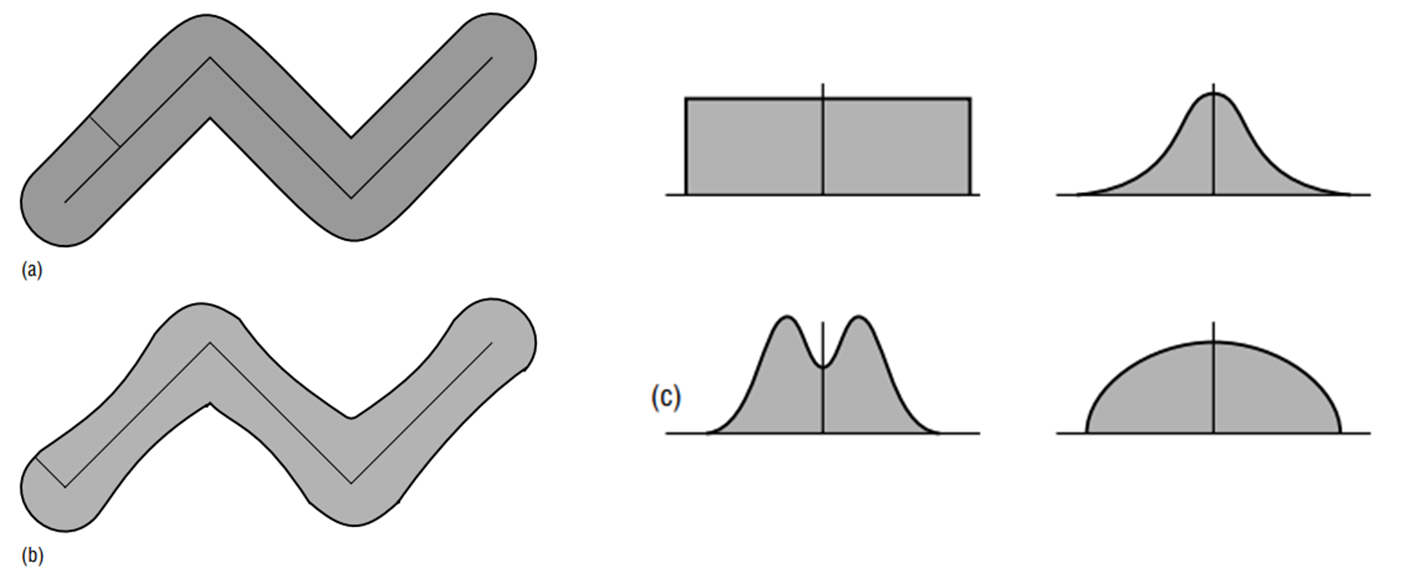


Figure 14: (a) Early models of the epsilon band show a uniform ‘sausage’ of width epsilon, f, surrounding the encoded line; (b) more recent studies suggest that the band may be nonuniform in width; and (c) four of the many possible distributions of error around the encoded line (Veregin, 2005).

Specifically for crowd-sourced data, positional accuracy has been assessed using: the buffer zone methodology as proposed by Goodchild and Hunter (1997) (see for example Haklay, 2010; Kouadi, 2009; Koukoletsos et al., 2011; Arsanjani et al., 2013), the distance between corresponding intersections of a road network (Antoniou, 2011), the Euclidean distance for point features (Girres and Touya, 2010; Stark, 2011; Jackson et al., 2013; Mashhadi et al., 2014), the average Euclidean distance for linear features (Girres and Touya, 2010; Fan et al., 2014), the Hausdorff distance for linear features (Girres and Touya, 2010), the surface distance, granularity and compactness for area features (Girres and Touya, 2010), the shape similarity (turning function) (Fan et al., 2014; Kalantari and La, 2015), x and y error distance (Stark, 2011), the grid based minimum bounding geometry and the directional distribution (Standard Deviational Ellipse) (Forghani and Delavar, 2014), the number of vertices, the mean vertex distance and distances between polygons centroids (Kalantari and La, 2015) and spatial similarity in multirepresentation considering directional and metric distance relationships (Hashemi et al., 2015)). Antoniou & Skopeliti (2015).

**Temporal Accuracy:**

Temporal accuracy has not received much attention in the literature. Temporal accuracy is often equated with ‘currentness’ (Thapa and Bossler 1992). The two concepts are quite distinct; temporal accuracy refers to the agreement between encoded and ‘actual’ temporal coordinates, where currentness is an application-specific measure of temporal accuracy. A value is current if it is correct in spite of any possible time-related changes in value. Thus, currentness refers to the degree to which a database is up to date (Redman 1992). To equate temporal accuracy with currentness is to state, in effect, that to be temporally accurate a database must be up to date. Clearly this is not the case since a database can achieve a high level of temporal accuracy without being current. (Veregin, 2005)

Temporal accuracy has been measured as the time difference between the time of the data capturing and the time of the data uploading (Antoniou *et al*., 2010).

**Thematic Accuracy:**

Metrics of thematic accuracy (or ‘attribute accuracy’) vary with measurement scale. For quantitative attributes, metrics are similar to those used to measure spatial accuracy for point features (e.g., RMSE). Quantitative attributes can be conceived as statistical surfaces for which accuracy can be measured in much the same way as for elevation. For categorical data most of the research into data quality has come from the field of classification accuracy assessment in  
remote sensing. This work was carried out initially to devise methods to assess the accuracy of classification procedures. Accuracy assessment is based on the selection of a sample of point locations, and a comparison of the land cover classes assigned to these locations by the classification procedure with the classes observed at these locations on a reference source (usually ‘ground truth’). A cross tabulation of the results (the ‘classification error matrix’) permits accuracy assessment (Aronoff 1985; Genderen and Lock 1977). Various metrics summarising the information in the error matrix have been developed, such as proportion correctly classified (e.g., Stark, 2011; Kounadi, 2009; Girres and Touya, 2010; Fan *et al*., 2014), kappa (e.g., Arsanjani and Vaz, 2015; Arsanjani *et al*., 2015), user’s and producer’s accuracies (Arsanjani and Vaz, 2015), etc.). These metrics are useful for assessing overall thematic accuracy. The classification error matrix contains additional information on the frequency of various types of  
misclassification, e.g., which pairs of classes tend most often to be confused. In addition, the matrix permits assessment of errors of omission (omission of a location from its ‘actual’ class) and errors of commission (assignment of a location to an incorrect class). (Veregin, 2005)

In respect with thematic accuracy for crowdsourced data, besides the above, researchers have proposed the use of the percentage (%) of specific values existing in tags (Girres and Touya, 2010, Antoniou, 2011), the Levenstein distance (Girres and Touya, 2010; Mashhadi *et al*., 2014; Kalantari and La, 2015), and the number of features with specific attributes (Fan *et al*., 2014; Arsanjani *et al*., 2013). Antoniou & Skopeliti (2015).

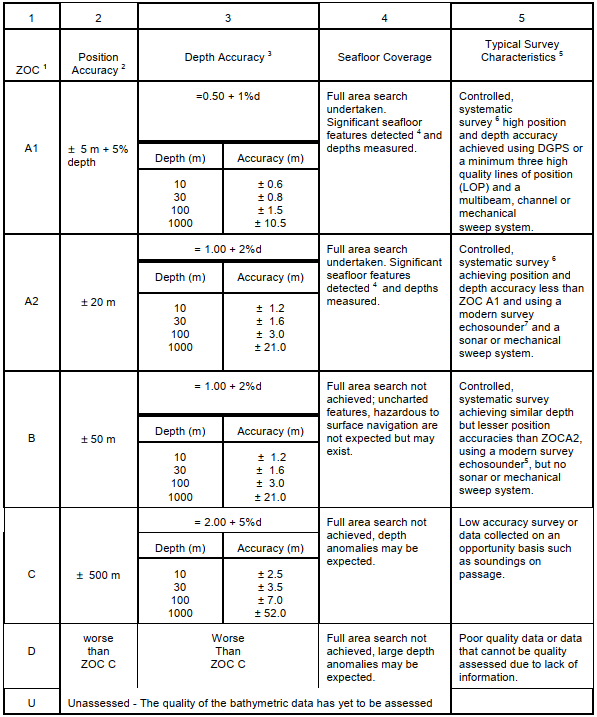
**Usability:**

Specifically for crowd-sourced data researchers claim that usability is the one most adequate data quality element since it is related to the term fitness for use for VGI (see for example Grira *et al*., 2010; Hacklay *et al*., 2010; Barron *et al*., 2014). Antoniou & Skopeliti (2015).

## 7.2 CATZOC

S-57 Annex A paragraph 2.2 Data quality description provides the existing guidance of how HOs should populate information about quality, reliability and accuracy of bathymetric data. The meta object M\_QUAL for an assessment of the quality of bathymetric data is mandatory for areas containing depth data or bathymetry. More detailed information about CATZOC can be found in IHO Publication S-67.

The guidance of how to assign appropriate CATZOC values, based on the quality of the data capture process, described in the quality report according to S-44, the connection to the accuracy and quality of the depth contours, soundings and completeness of seabed coverage, taking into account the intended scale of mapping, is further to be discussed by the DQWG subWG.



## 7.3 S-44 – CATZOC comparison

S-44 Table 1 lists the minimum standards for Hydrographic Surveys. It is a mixture of Bathymetric Data and non-Bathymetric Data in the water/at the surface and of non-Bathymetric Data connected to land (coastline/topography significant to navigation). S-57 and S-101 meta object M\_QUAL (meta\_quality) defines areas within which a uniform assessment exists for the quality of bathymetric data. The differences between S-57 and S-101 is that in S-101 the attribute *Category of temporal variation* has been included and that Data assessment can be assigned a value of assessed (Oceanic). This *Category of temporal variation* attribute will by default be set to value 5: *unlikely to change* and the Hydrographic Office is recommended to set this value for each area to the appropriate level when upgrading to S-101.

S-44 and S-57 share the following concepts:

1. Horizontal accuracy (position)
2. Vertical accuracy (depth)
3. Completeness (full seafloor coverage)
4. Isolated dangers (feature detection)

The S-57 M\_QUAL has a mandatory attribute CATZOC (=Category Zone of Confidence). There is a one-to-one or many-to-one relation between S-44 assigned values of surveys and S-57 assigned values of CATZOC. This means that a single survey can translate directly into a single value of CATZOC or that an adjacent set of surveys translate into a single value of CATZOC. In theory a single survey can be separated into more than one CATZOC value but this is very unlikely to happen. To relate both concepts, a cross-table is presented for each of the four sharing concepts:

Table 1: cross reference on horizontal accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ZOC value | Special Order | 1a | 1b | 2 |
| A1 | ± 5m + 0.05\* depth | 2 m | ± 5m + 0.05\* depth | ± 5m + 0.05\* depth | 20 m + 0.1\* depth |
| A2 | ± 20m | 2 m | ± 5m + 0.05\* depth | ± 5m + 0.05\* depth | 20 m + 0.1\* depth |
| B | ± 50m | 2 m | ± 5m + 0.05\* depth | ± 5m + 0.05\* depth | 20 m + 0.1\* depth |
| C | ± 500m | 2 m | ± 5m + 0.05\* depth | ± 5m + 0.05\* depth | 20 m + 0.1\* depth |
| D | > 500m | 2 m | ± 5m + 0.05\* depth | ± 5m + 0.05\* depth | 20 m + 0.1\* depth |

Table 2: cross reference on vertical accuracy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ZOC value | Special Order | 1a | 1b | 2 |
| A1 | ± 0.5m + 0.01\* depth | √((0.252 + (0.0075\*depth)2) | √((0.52 + (0.013\*depth)2) | √((0.52 + (0.013\*depth)2) | √((1.02 + (0.023\*depth)2) |
| A2 | ± 1.0m + 0.02\* depth | √((0.252 + (0.0075\*depth)2) | √((0.52 + (0.013\*depth)2) | √((0.52 + (0.013\*depth)2) | √((1.02 + (0.023\*depth)2) |
| B | ± 1.0m + 0.02\* depth | √((0.252 + (0.0075\*depth)2) | √((0.52 + (0.013\*depth)2) | √((0.52 + (0.013\*depth)2) | √((1.02 + (0.023\*depth)2) |
| C | ± 2.0m + 0.05\* depth | √((0.252 + (0.0075\*depth)2) | √((0.52 + (0.013\*depth)2) | √((0.52 + (0.013\*depth)2) | √((1.02 + (0.023\*depth)2) |
| D | > 2.0m + 0.05\* depth | √((0.252 + (0.0075\*depth)2) | √((0.52 + (0.013\*depth)2) | √((0.52 + (0.013\*depth)2) | √((1.02 + (0.023\*depth)2) |

Table 3: cross reference on completeness

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ZOC value | Special Order | 1a | 1b | 2 |
| A1 | YES | YES | YES | NO | NO |
| A2 | YES | YES | YES | NO | NO |
| B | NO | YES | YES | NO | NO |
| C | NO | YES | YES | NO | NO |
| D | NO | YES | YES | NO | NO |

Table 4: cross reference on isolated dangers

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | ZOC value | Special Order | 1a | 1b | 2 |
| A1 | detected (2 meter, 10% of depth (>40m)) | cubic features > 1 meter | cubic features > 2 meter, 10% of depth (>40m) | NA | NA |
| A2 | detected (2 meter, 10% of depth (>40m)) | cubic features > 1 meter | cubic features > 2 meter, 10% of depth (>40m) | NA | NA |
| B | not expected but may exist | cubic features > 1 meter | cubic features > 2 meter, 10% of depth (>40m) | NA | NA |
| C | unknown, depth anomalies may be expected | cubic features > 1 meter | cubic features > 2 meter, 10% of depth (>40m) | NA | NA |
| D | unknown, large depth anomalies may be expected | cubic features > 1 meter | cubic features > 2 meter, 10% of depth (>40m) | NA | NA |

When assigning a CATZOC value, HO’s are recommended to follow the guideline (paper DQWG14-06B) developed by the DQWG. This consists of stages in the following order:

1. Data assessment
2. Category of temporal variation (S-101 only)
3. Significant features detected
4. Least depth of significant features known
5. Full seafloor coverage achieved
6. Depth accuracy
7. Positional accuracy

The result is then computed by going through these 7 stages and indicating a valid (green) or fault (red) outcome:

Check 1: Data assessment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ZOC | Special Order | Order 1a | Order 1b | Order 2 | unknown | Oceanic |
| A1 | valid | valid | valid | valid | fault | fault |
| A2 | valid | valid | valid | valid | fault | fault |
| B | valid | valid | valid | valid | fault | fault |
| C | valid | valid | valid | valid | fault | fault |
| D | valid | valid | valid | valid | fault | fault |
| U | fault | fault | fault | fault | valid | fault |
| Oceanic | fault | fault | fault | fault | fault | valid |

If a CATZOC value is given U=unassessed or Oceanic, then no further checks are required.

Check 2: Category of temporal variation

This is regardless of the S-44 classification of the survey and will be further explained in a different paper. In S-57 and when upgrading to S-101, the default value of this attribute is “unlikely to change” and thus not affecting the outcome of this checking process. HO’s are however requested to assign the correct value to this attribute when making the upgrade to S-101.

Check 3: Significant features detected

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ZOC | Special Order | Order 1a | Order 1b | Order 2 |
| A1 | valid | valid | fault | fault |
| A2 | valid | valid | fault | fault |
| B | valid | valid | valid | valid |
| C | valid | valid | valid | valid |
| D | valid | valid | valid | valid |

Check 4: Least depth of significant features known:

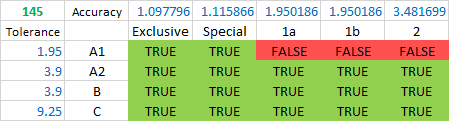
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ZOC | Special Order | Order 1a | Order 1b | Order 2 |
| A1 | valid | valid | fault | fault |
| A2 | valid | valid | fault | fault |
| B | valid | valid | valid | valid |
| C | valid | valid | valid | valid |
| D | valid | valid | valid | valid |

Check 5: Full seafloor coverage achieved

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ZOC | Special Order | Order 1a | Order 1b | Order 2 |
| A1 | valid | valid | fault | fault |
| A2 | valid | valid | fault | fault |
| B | valid | valid | valid | valid |
| C | valid | valid | valid | valid |
| D | valid | valid | valid | valid |

Check 6: Depth accuracy

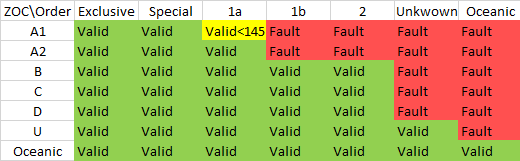
----for 145m this if fails:

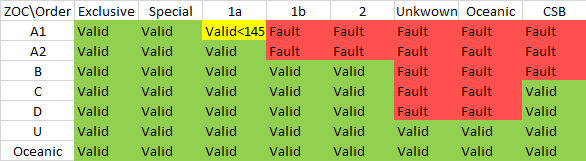


Check 7: Positional accuracy

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ZOC | Special Order | Order 1a | Order 1b | Order 2 |
| A1 | valid | valid | valid | fault |
| A2 | valid | valid | valid | fault |
| B | valid | valid | valid | valid |
| C | valid | valid | valid | valid |
| D | valid | valid | valid | valid |

When combining these 7 steps we get the following result:





NOTE: valid does not mean appropriate. For example, a special order survey has the appropriate CATZOC level of A1. Values of A2, B, C and D are valid but does not justify the high quality of the original survey.

## 7.4 Added Value of CSB Data

Data capture by official (hydrographic) surveys are done to capture the shape of the seabed and process this data into an official nautical chart. The captured data is validated against acquisition standards (S-44), then Product Specification (S-57) and User requirement (guaranteed use in certified ECDIS equipment by trained and certified user).

In order to assess the usefulness of CSB data to the nautical chart, metaquality is recommended to be used.

Metaquality elements are a set of quantitative and qualitative statements about a quality evaluation and its results. The knowledge about the quality and the suitability of the evaluation method, the measure applied and the given result may be of the same importance as the result itself.

Metaquality may be described using the following elements:

* confidence – trustworthiness of a data quality result.
* representativity – degree to which the sample used has produced a result which is representative of the data within the data quality scope
* homogeneity – expected or tested uniformity of the results obtained for a data quality evaluation

NOTE quantitative figures for confidence can be obtained by statistical parameters such as standard deviation or a confidence interval on a given confidence level.

## 7.4 HO National Methodologies

Below is a summary of the methodologies for assigning CATZOC of the National Hydrographic Offices of Australia, Brazil, Finland, France, Italy, Japan, Netherlands, Norway, United Kingdom, and USA.

Methods:

* **Data Source**
  + **AU**: ZOC (A2 max for LIDAR, B for scientific research and environmental MBES surveys, C max for opportunity soundings, D for soundings with little or no metadata and/or on unknown datums)
  + **BR**: according to ZOC table
  + **FI**: ZOC with modifications. In detail:
    - A: Full sea floor ensonification or sweep. A1 and A are combined to A.
    - B: Full seafloor coverage not achieved; depth anomalies may exist.
    - C: Full seafloor coverage not achieved; depth anomalies may be expected.
    - D: Not in use
    - U: Data unassessed.
  + **FR**: ZOC (for data from other HO can’t be better than B, D for reconnaissance survey, D for aerial photography, and from foreign charts CATZOC is used as is if exists or C for charts scale>250K and D for charts of scale <250K)
  + **IT**: ZOC but if from paper assigned B or C
  + **JP**: ZOC
  + **NL**: ZOC. approaches to main ports are A1 but cartographer can downgrade to A2 if less frequently surveyed and, e.g., area is changeable.
  + **NO**: ZOC
  + **UK**: ZOC in principle, B for LIDAR, C for SDB, D for aerial photography.
  + **US**: ZOC but if from USACE normally B (with exception A1 or A2), C for SDB, D when no survey record.
* **Currentness** 
  + **BR**: only until 2014
  + **FR**: CATZOC is not downgraded due to the passage of time (mariners have to read CATZOC with M\_SREL. For older surveys:
    - B if survey is after 2003 but source is not SHOM
    - since 1992: A2 if MBES, full seafloor, and sidescan; B if MBES + full sea floor
    - since 1980: A2 if scale>20K and sidescan
    - since 1970: B if scale >5K (P<20m) or scale>100K (P<100m), else C
    - 1935-1970: C
    - post-1970:
  + **IT**: No
  + **JP**:
    - 1968-2011: B, C, D (depending on line spacing)
    - pre-1968: D
  + **NL**: downgrading is considered when a resurvey is overtime (i.a.w. survey plan)
  + **NO**:
    - A2 for 1990 -early 2000s when d>30m and B when d<30m
    - B for surveys 1950-1990
    - C for pre-1950
  + **UK**: No
  + **US**:
    - B for surveys after 1940 with SBES, nearshore, and survey scale > 40K
    - C for 1940-1990 offshore; 1940-1990, scale <40K and chart scale >survey scale; 1920 to 1940 on known horizontal and vertical datums.
    - D for pre-1940 with unknown horizontal or vertical datums; all pre-1920
* **Chart Scale**
  + **IT**: No but D for Bands 1 and 2 (with the exception of UNSARE that is U)
* **Generalization**
  + **BR**: No
  + **FR**: Yes (could be downgraded due to generalization for safety purposes)
  + **IT**: No
* **Seabed change/mobility**:
  + **IT**: downgrades for instability of bathymetry
  + **NL**: D after extreme events
  + **UK**: D may be attributed after natural disaster
  + **US**: for any variation concerning vertical and horizontal accuracy

Other:

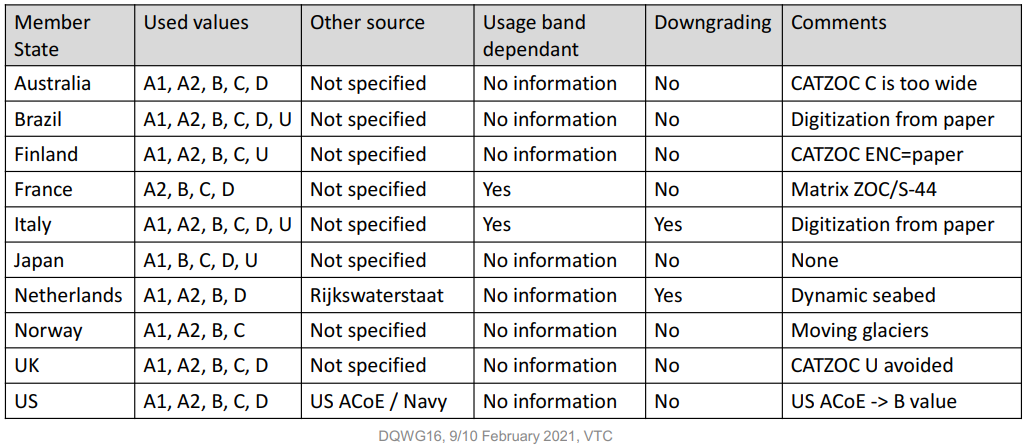
* UNSARE:
  + **IT**: U but may change policy (e.g., D when overlaps bathy features)
  + **JP**: U
  + **UK**: D
* DEPARE:
  + **IT**: Unsurveyed DEPARE 0-2 or 0-5m assigned D (when survey A1 or A2) or B,C. DEPARE at least 1cm side at scale
* DRGARE:
  + **IT**: Regularly maintained A1, otherwise D

Notes:

* AU: CATZOC C category is too wide. It covers old (but good for their day) hydrographic surveys which cannot be transformed accurately to modern datums, and also opportunity soundings such as passage sounding.
* NO: Svalbard: For general information about the quality of the charts around Svalbard, reference is made to The Norwegian Pilot guide, Volume 7 and the information given in each chart. The glacier fronts seawards are continually changing. In general, the glacier fronts are receding. Observations exist where the glaciers have receded several hundred meters during the last decades. For this reason, contour lines and terrain close to the glacier can deviate from contour lines on the chart. Surveys are in some areas of Svalbard incomplete. Large areas have not been surveyed using modern technology but include some very old bathymetry. In the ENCs these areas have mostly been given CATZOC D.
* UK: Historically, many UK ENCs had their CATZOC populated using survey source, dates, scale and sounding technique (taken from the corresponding chart SOURCES diagram).

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Member State | Used Values | Sources | Usage Band dependent | Downgrading with | | Time dependent | | |  |
| time | gener/tion | 1 | 5 |  |  |
| Australia |  |  |  |  |  |  |  |  |  |
| Brazil | All |  |  |  |  | A1 | A2 |  |  |
| Finland |  |  |  |  |  |  |  |  |  |
| France |  |  |  |  |  |  |  |  |  |
| Italy |  |  |  |  | Yes |  |  |  |  |
| Japan |  |  |  |  |  |  |  |  |  |
| Netherlands |  |  |  |  |  |  |  |  |  |
| Norway |  |  |  |  |  |  |  |  |  |
| UK |  |  |  |  |  |  |  |  |  |
| USA |  |  |  |  |  |  |  |  |  |

Table 7



**US NOAA OCS Hydrographic Health Model**

# 8. Conclusions and recommendations

* Depth data can be represented by a vector model (S-57/S-101) and/or a grid model (S-102). The Producing Authority of an ENC, who also provides depth data in S-102 format, should ensure consistency between the two datasets (maintain integrity and positional consistency).
* In order to ensure consistency, as a minimum, the same vertical datum and realization of that vertical datum should be used when datasets are provided by the same Producing Authority.
* Coherence of depth data at different levels of detail should be maintained.
* There is a possible relationship between the ground sampling distance of the data captured, the intended scale of mapping and normal contour line vertical interval. This is further to be investigated to ensure that in a chart not too many contour lines are depicted giving a false sense of accuracy.
* Depth data has an associated uncertainty (mainly in a vertical sense). When combining different datasets, these uncertainties provide a degree of confidence to which data is the most representative to depict the shape of the seabed and additional man-made structures (wrecks).
* Using the associated uncertainties presented as meta-quality data, CSB may have additional value to existing nautical charts, most likely in areas where no gridded data exists and where soundings and depth contours have a high degree of uncertainty (remote areas).

# References

AGI n.d. Guidelines for geographic information content and quality: For those who use, hold or need to acquire geographic information. Association for Geographic Information. London, UK. accessed on April 4th, 2021 at <https://www.geos.ed.ac.uk/~gisteac/proceedingsonline/Source%20Book%202004/SDI/National/UK/AGI/Guidelines%20for%20GI%20content%20and%20quality.htm>

Beard, M., Buttenfield, B. P, & Clapham, S. B. (1991). NCGIA Research Initiative 7 Visualization of Spatial Data Quality: Scientific Report for the Specialist Meeting (91-26). *UC Santa Barbara: National Center for Geographic Information and Analysis*. Retrieved from <https://escholarship.org/uc/item/6w1695bs>

ICA, 1967. Hundred Technical Terms in Cartography. International Association of Cartography.

INSPIRE D2.8.II.1 Data Specification on Elevation – Technical Guidelines.

IHO Publication S-4 Regulations for International (INT) Charts and Chart Specifications of the IHO (Nov 2018), item B-611.1

IHO (International Hydrographic Organization) (2018). ENC Validation Checks. Publication S-58. 6.1.0 ed. Monaco: International Hydrographic Organization.

International Organisation for Standardisation, 2005. *ISO19113 Geographic information - Quality principles*, Geneva: ISO.

International Organisation for Standardisation, 2013. *ISO19157:2013 Geographic information - Data quality*, Geneva: ISO.

Mank, A.K. (2019). Cartographic symbolization for high-resolution displays. Master *Thesis.* University of Wien.

Regnauld, N. (2015). Generalization and Data Quality. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-3/W3, 2015 ISPRS Geospatial Week 2015, 28 Sep – 03 Oct 2015, La Grande Motte, France.

Robinson, A. H., Morrison, J. L., Muehrcke, P. C., Kimerling, A. J., & Guptill, S. C. (1995). *Elements of cartography*. 6th ed. New York, N.Y: John Wiley & Sons.

Ruas A. and Plazanet C. (1996). “Strategies for automated generalization”, In Advances in GIS research II, Proceedings of the 7th International Symposium on Spatial Data Handling; Kraak, M.J., Molenaar, M., Eds.; Taylor & Francis: London, UK, 1996; pp. 319–336.

Rytz, A., Bantel, E., Hoinkes, C., Merkle, G., and Schelling, G. (1980) Cartographic Generalization: Topographic Maps. Cartographic Publication Series (English translation of publication No.1). Swiss Society of Cartography.

Veregin, H. (1995). “Data Quality Parameters”, In New Developments in Geographical Information Systems: Principles, Techniques, Management and Applications; P A Longley, M F Goodchild, D J Maguire, and D W Rhind., Eds.; pp. 177–189.

Ware, C. (2012). *Information Visualization: Perception for Design*: Morgan Kaufmann Publishers Inc.

…..

SDTS (Spatial Data Transfer Standard) Guide for Technical Managers <https://www.mapi.gov.il/gisforum/tzevet_technology/stds-guide%20for%20technical%20managers.pdf>

Inspire D2.8.II.1 Data Specification on Elevation – Technical Guidelines <https://inspire.ec.europa.eu/id/document/tg/el>

EPA Guidance for Geospatial Data Quality Assurance Project Plans <https://www.epa.gov/sites/production/files/documents/guidance_geospatial_data_qapp.pdf>

FGDC (Federal Geographic Data Committee) Content Standards for Digital Geospatial Metadata

<https://www.fgdc.gov/standards/projects/metadata/base-metadata/v2_0698.pdfhttps://www.fgdc.gov/standards/projects/metadata/base-metadata/v2_0698.pdf>

# ANNEX A: Acronyms and Abbreviations

ADQR Aggregated Data Quality Result.

CATZOC Category of Zone of Confidence

CD Chart Datum (also known as Sounding Datum)

CRS Coordinate Reference System

CSBWG Crowdsourced Bathymetry Working Group

DCEG S-101 Data Classification and Encoding Guide

DEM Digital Elevation Model

DQWG Data Quality Working Group

DSM Digital Surface Model

DTM Digital Terrain Model

EGM2008 Earth Gravity Model 2008 (realization 2008)

ETRS89 European Terrestrial Reference System 89 (realization 1989)

ITRF International Terrestrial Reference Frame

ITRS International Terrestrial Reference System

MSL Mean Sea Level

NAD83 North American Datum 83 (realization 1983)

S-4 Regulations for International (INT) Charts and Chart Specifications of the IHO

S-44 IHO Standards for Hydrographic Surveys

S-101 ENC Product Specification

S-101PT S-101 Project Team

S-102 Bathymetric Surface Product Specification

S-102PT S-102 Project Team

S-57 IHO Transfer Standard for Digital Hydrographic Data

S-67 Mariner’s Guide to Accuracy of Depth Information in Electronic Navigational Charts

SDB Satellite Derived Bathymetry

TWCWG Tides, Water Levels and Current Working Group

WGS 84 World Geodetic System 1984 (realization 1984)

# ANNEX B: Terms and Definitions

* **completeness**: completeness is defined as the presence and absence of features, their attributes, and relationships. [1]
* **chart datum**: vertical coordinate reference system which is used to refer and portray depth. measurements as property values. (INSPIRE D2.8.II.1) (also known as s**ounding datum**).
* **contour line**: linear spatial object composed of a set of adjoining locations characterized by having the same elevation property value. It describes, together with other contour lines present in the area, the local morphology of the Earth's surface. (INSPIRE D2.8.II.1)
* **depth:** Elevation property measured along a plumb line in a direction coincident to Earth’s gravity field. (INSPIRE D2.8.II.1)
* **depth area**: a water area whose depth is within a defined range of values.
* **depth contour**: a line connecting points of equal water depth (S-57)
* **fitness for use**:
* **height**: Elevation property measured along a plumb line in a direction opposite to Earth’s gravity field. (INSPIRE D2.8.II.1)
* **lineage**: lineage describes the history of a dataset and recount the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form. [1]
* **metaquality**: metaquality describes the quality of the data quality results in terms of defined characteristics. [1]
* **purpose**: purpose describes the rationale for creating a dataset and contains information about its intended use, which may not be the same as the actual use of the dataset. [1]
* **result scope**: a subset of the data quality scope. [1]
* **sounding datum**: see **chart datum**.
* **uncertainty**: estimate characterising the range of values within which the true value of a measurement is expected to lie as defined within a particular confidence level. It is expressed as a positive value.
* **universe of discourse**: view of the real or hypothetical world that includes everything of interest. [1]
* **usage**: usage describes the application(s) for which a dataset has been used, either by the data producer or by other data users. [1]

# ANNEX C: Realization of geodetic reference frame and horizontal and vertical coordinate reference systems

IHO Publication C-51 Manual on Technical Aspects on the UN Convention on the Law of the Sea (6th Edition), Chapter 2 – Geodesy and positioning, provides an in-depth reading to the establishment of modern space geodesy, geodetic datums, vertical datums and satellite positioning. For this paper, a summary with images is presented below.

The locations of points in three-dimensional space are most conveniently described by Cartesian coordinates: X,Y and Z. Since the start of the Space Age, such coordinate systems are typically “geocentric”, with the Z-axis aligned with either the Earth’s conventionally defined or instantaneous rotation axis. Because the Earth’s geocentre, or centre of mass, is located at one focus of a satellite’s orbital ellipse, this point is the natural origin of a coordinate system defined by satellite-based geodetic methods. The International Celestial Reference System (ICRS) forms the basis for describing celestial coordinates, and the International Terrestrial Reference System (ITRS) is the foundation for the definition of terrestrial coordinates to the highest possible accuracy. The definitions of these systems include the orientation and origin of their axes, scale, physical constants and models used in their realisation, e.g., the size, shape and orientation of the “reference ellipsoid” that approximates the geoid and the Earth’s gravity field model. The coordinate transformation between the ICRS and ITRS is described by a sequence of rotations that account for precession, nutation, Greenwich apparent sidereal time, and polar motion, which collectively account for variations in the orientation of the Earth’s rotation axis and its rotational speed.

While a reference system is a mathematical abstraction, its practical realisation through geodetic observations is known as a “reference frame”. The conventional realisation of the ITRS is the International Terrestrial Reference Frame (ITRF), which is a set of coordinates and linear velocities (the latter due mainly to crustal deformation and tectonic plate motion) of well-defined fundamental ground stations.

The solid surface of the Earth (including the sea floor) consists of a number of large tectonic plates (and many smaller ones whose boundaries are less well defined) that slide across the Lithosphere, in the process colliding with other plates. The speed of the plates may be as high as a decimetre or more per year, though typically tectonic plate motion is of the order of a few centimetres per year relative to a fixed coordinate framework. That framework is realised by the fixed axes of the ITRF.

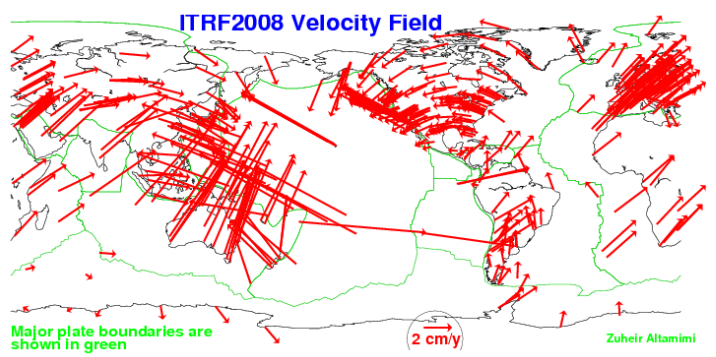


Figure 14: Global model of the Earth’s tectonic plates with estimated velocities (ITRF2008, 2012)

## The geoid

The word geoid is used to designate that special equipotential or geopotential surface which coincides with, but is not exactly equivalent to, the mean sea level (MSL) surface of the oceans in an average sense. It is that surface to which the oceans would conform over the entire Earth, if free to adjust to the combined effect of the Earth's mass attraction and the centrifugal force of the Earth's rotation, the forces of which are collectively referred to as the Earth’s gravity field. Although the above definition refers to sea level, conceptually the geoid extends under the continents and differs from a best fitting ellipsoid by vertical distances that are up to one hundred metres or so.

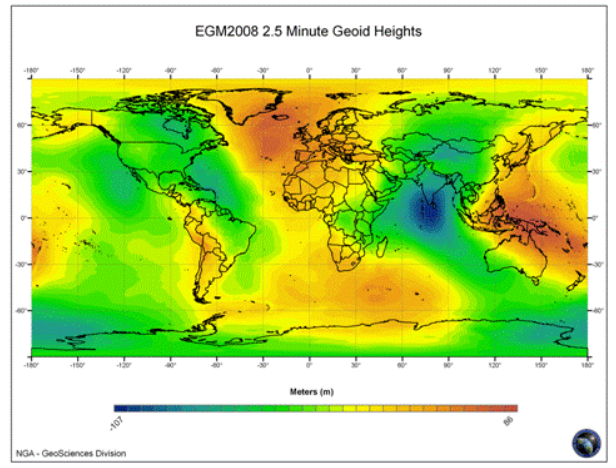


Figure 15: Earth Gravity Model EGM2008 geoid heights, derived from a combination of surface gravity data principally on land, satellite altimetry over ocean areas and an analysis of the observed orbit perturbations of many near-Earth satellites (EGM2008, 2012)

## The reference ellipsoid

The geoid is a very irregularly shaped surface (Figure 9) and therefore for geodetic and mapping purposes it has been necessary to use a simplified geometric shape – the ellipsoid which closely approximates the shape of the geoid for all calculations. The IAG recommends the use of the “GRS80” reference ellipsoid. However, the slightly different WGS84 ellipsoid is also commonly used. A reference ellipsoid with its centre at the geocentre best fits, in a geometric sense, the geoid globally – resulting in the maximum separation of these two surfaces being of the order of 100 metres (Figure 10). The geoid height is mathematically defined as so many metres above (+N) or below (-N) a given ellipsoid (Figure 11)

H = H + N

Where:

H – geodetic height (height above the ellipsoid)

H = height above MSL

N – geoidal height

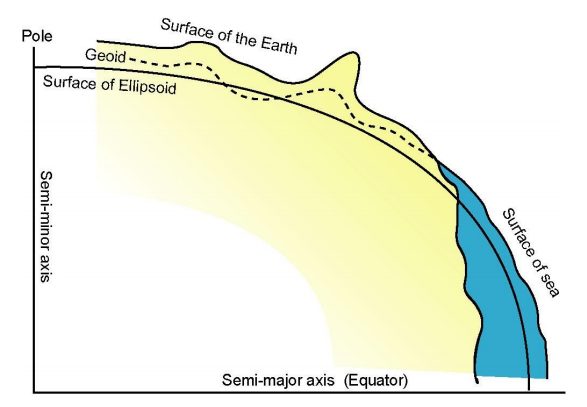


Figure 16: A meridian section of the Earth showing the various physical and mathematical surfaces used in geodesy.

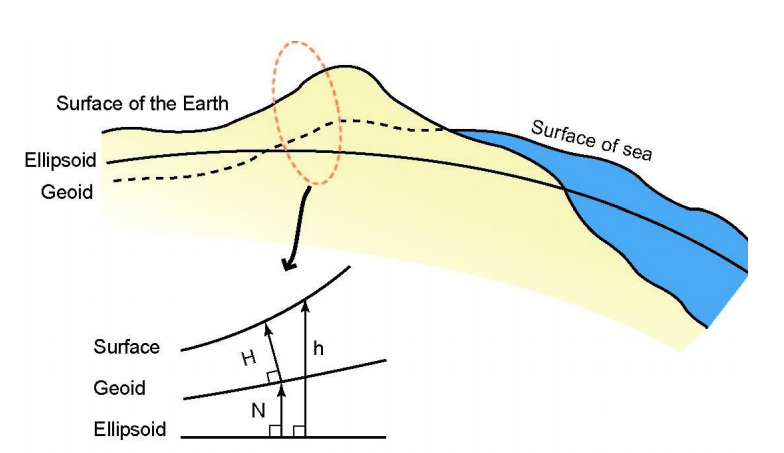


Figure 17: Geodetic height (h), orthometric height (H), geoid height (N) and their interrelations

## Transformation between geodetic datums

The parameters of the transformation between two geodetic datums are determined empirically from the coordinates of a set of identical points on both datums. These positions are always distorted due to the inevitable presence of both systematic and random errors, and hence the determination of transformation parameters must be done carefully. It is recommended that the transformation parameters relevant to a State’s datum be obtained from the appropriate national mapping or charting agency. There are also published 14-parameter transformation models that incorporate time-rate-of-change of the standard 7 similarity transformation parameters. With such a model, it is possible to both accommodate epoch year differences between datums, as well as origin, orientation and scale effects.

## Vertical datum

For land height systems, the geoid and MSL are assumed to coincide at the fundamental benchmark(s) or tide gauge(s) that define(s) a State’s geodetic height datum to which heights on land maps are referred (see Figure 12). A high water level, determined by some procedure that samples the high tide, may define the so-called hydrographic shoreline, where land mapping transitions to marine charting. In some countries a high water level marks the limit of land property that can be registered in a cadastre (register of rights of property owners).

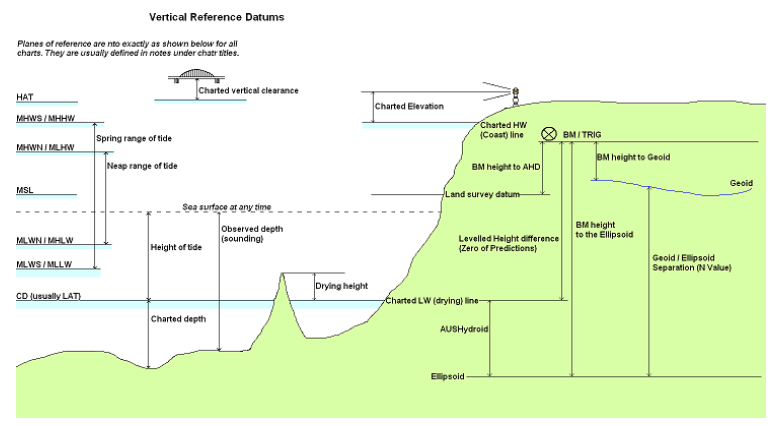


Figure 18: Vertical datums for mapping and charting (courtesy Australian Hydrographic Office)

## Chart Datum

To provide the mariner with a margin of safety in terms of depth measurements, all charted depths are referred to chart datum, which is equated to the datum of tidal predictions and defined by the IHO as a plane so low that the tide will not frequently fall below it (see Figure 12). Thus, unlike heights on land maps, which are normally referred to MSL as a proxy surface for the geoid, depths on charts are referred to a low water level. For the determination of chart datum, it is necessary to observe heights of points above the low water. Thus, the height of low water below MSL must be determined. This is done by analysing the records of tide gauges from the vicinity of the area of interest, which may require a specific expertise that is generally available within a State’s hydrographic survey authority or harbour port authorities.

Owing to the many varied tidal characteristics existing throughout the world, a precise, scientific definition for chart datum, which could be used universally, has not been agreed upon. Over the past 200 years, different countries have adopted different methods for computing chart datum, depending usually on the type of prevailing tide. In accordance with an IHO Resolution of 1926, chart datum should:

* Be so low that the water will but seldom fall below it.
* Not be so low as to cause the charted depths to be unrealistically deep.
* Vary only gradually from area to area and from chart to chart to adjoining chart, to avoid significant discontinuities.[[11]](#footnote-12)

In very basic terms, a chart datum can be defined as the mean of specific low waters over an extended period of time. The time period should ideally be 19 years or more, in order to include all the significant astronomical variations described above. Opinions vary, however, in terms of which low waters should be used to arrive at this mean value, and as a result different definitions are in use. For example, some countries define chart datum as the mean of all the lower low waters (MLLW) over a specified 19 year period. Others use a chart datum called lower low water large tides (LLWLT), which is defined as the average of the lowest low waters, one from each of 19 years of prediction. Yet others use the lowest low water spring tide (LLWST), which is the average of the lowest low water observations of spring tides, over a specified period. The most conservative use the lowest astronomical tide (LAT), which is the lowest level that can be predicted to occur under average meteorological conditions and under a combination of astronomical conditions (see Figure 12).

# ANNEX D: Examples of DTM/DSM on land and sea

The data models incorporated are aimed at describing the three-dimensional shape of the Earth‘s surface in terms of Elevation properties, either height or depth. Both properties are constrained to the physical vertical dimension, measured along the plumb line from a well defined surface, such as a geoid or a specific water level. The orientation of the positive axis is opposite to the Earth's gravity field in the case of the height property (upwards) and coincident to the Earth's gravity field in the case of the depth property (downwards). Hence, heights are positive above the surface taken as origin whereas depths are positive below it, as show in the next figures.

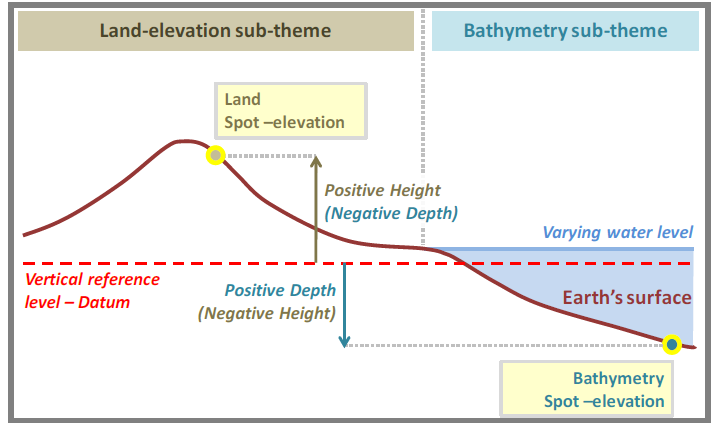


Figure 19: Measuring of elevation properties

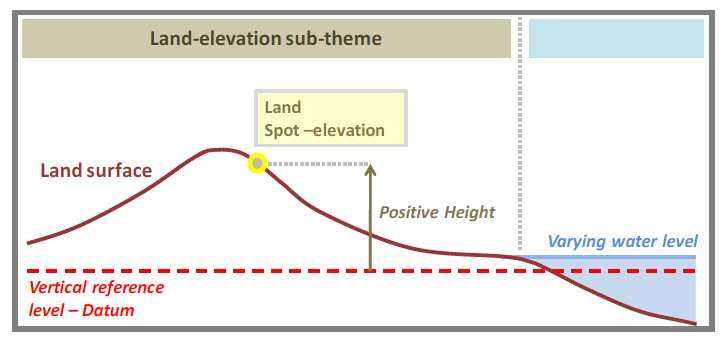


Figure 20: Description of land elevation

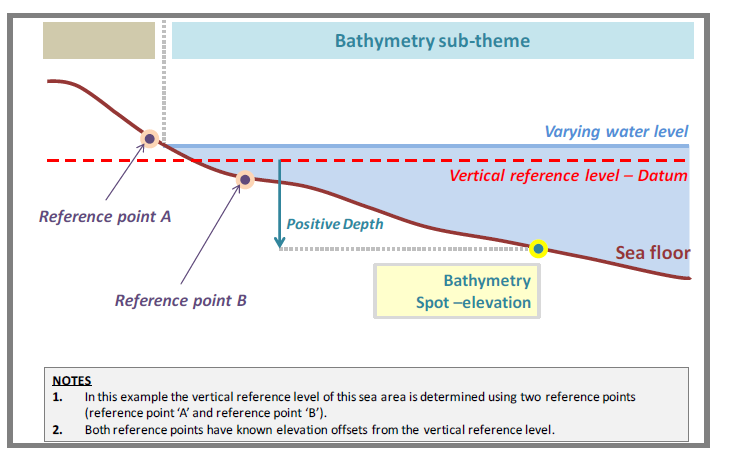


Figure 21: Description of the sea floor

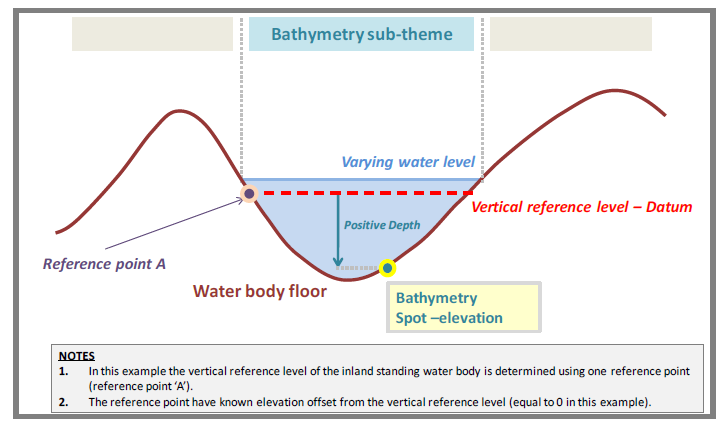


Figure 22: Description of the floor of an inland standing water body

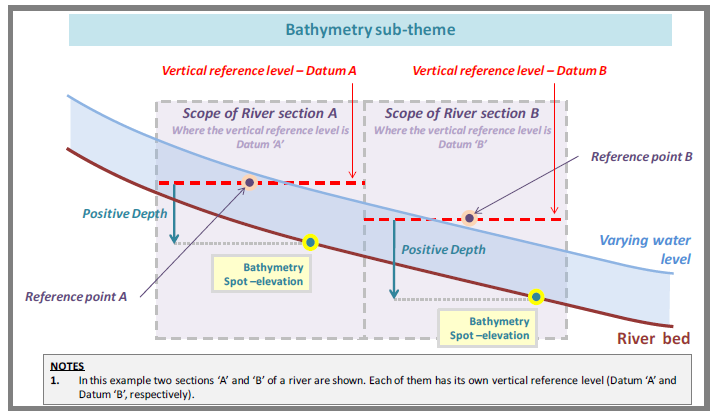


Figure 23: Description of the bed of a navigable river

# ANNEX E: ZOC Data Quality Measures

This Annex contains a list of data quality measures for point, linear, and polygonal objects (see ISO 19157 Annex D).

**NOTE to Sub-WG members:** AGI Guidelines for geographic information content and quality contains the following list of questions designed to help those who require geographic information to ensure that it is fit for its intended purpose. The list is primarily intended for people who are occasionally involved with geographic information. This WG is already familiar with many of the issues discussed (and already answered). However, it is temporarily included here as we may find a few that are worth reviewing/discussing for deciding the quality measures and results (e.g., the highlighted).  **1. Defining your task**1.1 What is your task? What question or questions are you trying to answer?  
1.2 What region are you interested in?  
1.3 Is your answer time dependent?  
1.4 How will you know if you have the right answer?  
1.5 How would you like the results presented?  
1.6 Has someone else already got the answer to your question or would they be interested in your answer?  
**2. Identifying the contents of the dataset that your task requires**2.1 For what region do you want your data?  
2.2 For what instant or time period do you want your data?  
2.3 What objects are required of your data?  
2.4 What attributes of these objects will your task's answer depend on?  
2.5 Do you need to know about every instance of these objects or a selected sample of them?  
**3. Specifying spatial attributes**3.1 How will you identify the location or size of the objects?  
3.2 Do you envisage objects as a point, line or area?  
3.3 Do you require cross references between one type of object and another?  
3.4 Is the reference system for your location important to you?  
3.5 To what resolution do you need to know location?  
3.6 Do you need to consider further the positional accuracy?  
3.7 Are the spatial relationships between objects important?  
**4. Specifying temporal attributes**4.1 In what way, if any, will time be important to the dataset?  
4.2 Is the resolution, or how the time and date are written, important?  
4.3 How do you want an interval specified?  
4.4 Are the temporal relationships between different objects or different attributes important?  
**5. Specifying thematic attributes**5.1 What other attributes of these objects are important to you?  
5.2 For any attributes you require, can you list all the possible values?  
5.3 For any attribute you require, what types of values do you want?  
5.4 What units of measurement, language etc. do you require?  
5.5 Can you detail any constraints to which you expect your attribute value to conform?  
**6. Identifying other issues**6.1 Do datasets exist that meet (or partially meet) your needs?  
6.2 Have you considered the most important elements of the data in your specification?  
6.3 What assurance do you require that the data is to your specification?  
6.4 Do you require the data to be kept up to date, and if so how and by whom?  
6.5 Could anyone else use this data now or in the future?  
6.6 Have you considered the functionality you will require of the software package to process this data?  
6.7 Will resources constrain your ability to acquire and use the data?  
6.8 How would you like to receive the data?  
6.9 Are there likely to be any restrictions in the use of this data?  
**7. What next?**7.1 How are you going to use your data specification?  
7.2 Do you need further advice?

Recommended minimum data quality results for depth (to be discussed and agreed upon) are shown in Table 8.

Table 8: recommended minimum data quality results[[12]](#footnote-13)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| nr | Data quality element and sub-element | Measure name(s) | Target result(s) | Condition |
| 1 | Completeness / Commission | Rate of excess items | 0 % | Data duplication can be checked and corrected automatically |
| 2 | Completeness / Omission | Rate of missing items | 0% |  |
| 3 | Logical consistency /  Conceptual consistency | Non-compliance rate with respect to the rules of the conceptual schema | 0% |  |
| 4 | Logical consistency /  Domain consistency | Value domain non-conformance rate | 0% | Quality controls and data editing can be performed automatically |
| 5 | Logical consistency /  Format consistency | Physical structure conflict rate | 0% |
| 6 | Logical consistency /  Topological consistency | Rate of missing connections due to undershoots | 0% |
| Rate of missing connections due to overshoots | 0% |
| Rate of invalid self-intersect errors | 0% |
| Rate of invalid self-overlap errors | 0% |
| 7 | Positional accuracy /  Absolute or external accuracy | Root mean square error of planimetry | Vector  *Horizontal (m)*  Max RMSEH = E/10000 |  |
| Root mean square error | Vector  *Vertical (m)*  Max RMSEV =Vint / 6  Note: Vint can be approximated by  E / 1000 | In low reliability areas the maximum error can be increased by 50% |
| Grid  *Horizontal (m)*  Max RMSEv = GSD / 3 |  |
| 8 | Positional accuracy /  Gridded data position accuracy | Root mean square error of planimetry | Grid  *Horizontal (m)*  Max RMSEH = GSD / 6 |  |

NOTE The following notation is used:

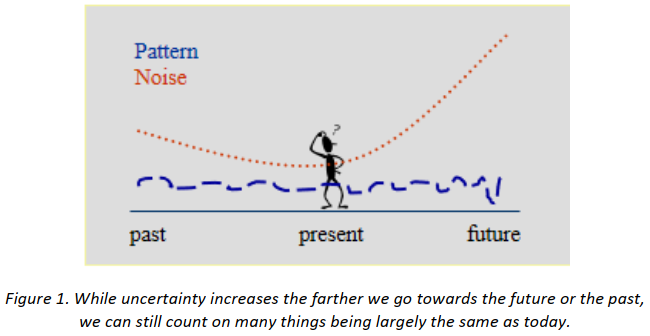
E Denominator of the intended scale of mapping. (Compilation scale)

Vint Normal contour line vertical interval.

GSD Ground sample distance

# Things to consider.

* S-44 – ZOC relation
* Data generalization
* CSB valid ZOCs
* National methodologies
* Depreciation over time



* Quality measures
* Object detection requirement?? 🡺 resolution

1. INSPIRE D2.8.II.1 Data Specification on Elevation – Technical Guidelines [↑](#footnote-ref-2)
2. Satellite Derived Bathymetry [↑](#footnote-ref-3)
3. IHO Publication S-4 Regulations for International (INT) Charts and Chart Specifications of the IHO (Nov 2018), item B-611.1 [↑](#footnote-ref-4)
4. IHO Publication S-44 Ed.6.0.0 Chapter 1 [↑](#footnote-ref-5)
5. IHO Publication S-44 Ed.6.0.0 Chapter 2 [↑](#footnote-ref-6)
6. IHO Publication S-44 Ed.6.0.0. par 2.1 [↑](#footnote-ref-7)
7. IHO Publication S-44 Ed.6.0.0 par. 2.3 [↑](#footnote-ref-8)
8. IHO Publication S-44 Ed.6.0.0 par 2.4 [↑](#footnote-ref-9)
9. IHO Publication S-57, UoC par 5.8.4 [↑](#footnote-ref-10)
10. S-101 DCEG par. 3.4 page 41 [↑](#footnote-ref-11)
11. IHO publication C-51, page 2-19 [↑](#footnote-ref-12)
12. INSPIRE D2.8.II.1\_v3.0 par 7.3 [↑](#footnote-ref-13)