S-130PT [**S-130PT4-3.7_Rev1**]
Meeting 4 24 OCT 2022
Agenda Item 3.7

Draft Data Quality chapter of S-130

X Data quality

X.1 Introduction to data quality

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

For S-130 the following Data Quality Elements have been included:

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

X.2 Completeness

X.2.1 Commission

Commission is applicable for S-130.

S-130 products must be tested with Commission checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. It is allowable to publish the data with a quality statement which indicates non-conformance.

In terms of Commission, S-130 products shall at least populate numberOfExcessItems that indicates the number of items that should not have been present in the dataset, and numberOfDuplicateFeatureInstances that indicates the total number of exact duplications of feature instances within the data.

X.2.2 Omission

Omission is applicable for S-130.

S-130 products must be tested with Omission checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. It is allowable to publish the data with a quality statement which indicates non-conformance.

In terms of Omission, S-130 products shall at least populate numberOfMissingItems that is the total number of missing items.

X.3 Logical Consistency

X.3.1 Conceptual Consistency

Conceptual Consistency is applicable for S-130 and follows the guidelines from S-100 Part 1.

S-130 products must be tested with Conceptual Consistency checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. Data should only be published if it passes the test.

In terms of Conceptual Consistency, S-130 products shall at least populate numberOfInvalidSurfaceOverlaps that is the total number of erroneous overlaps within the data.

X.3.2 Domain Consistency

Domain Consistency is applicable for S-130 and follows the guidelines from S-100 Part 5.

S-130 products must be tested with Domain Consistency checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. It is allowable to publish the data with a quality statement which indicates non-conformance.

In terms of Domain Consistency, S-130 products shall at least populate numberOfNonconformantItems that is a count of all items in the dataset that are not in conformance with their value domain.

X.3.3 Format Consistency

Format Consistency is applicable for S-130 and follows the guidelines from S-100 Part 10b.

S-130 products must be tested with Format Consistency checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. Data should only be published if it passes the test.

In terms of Format Consistency, S-130 products shall at least populate physicalStructureConflicts that is a count of all items in the dataset that are stored in conflict with the physical structure of the dataset.

X.3.4 Topological Consistency

Topological Consistency is applicable for S-130 and follows the guidelines from S-100 Part 7.

S-130 products must be tested with Topological Consistency checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. Data should only be published if it passes the test.

In terms of Topological Consistency, S-130 products shall at least populate rateOfFaultyPointCurveConnections that is the number of faulty link-node connections in relation to the number of supposed link-node connections, numberOfMissingConnectionsUndershoots that is a count of items in the dataset within the parameter tolerance that are mismatched due to undershoots, numberOfMissingConnectionsOvershoots that is a count of items in the dataset within the parameter tolerance that are mismatched due to overshoots, numberOfInvalidSlivers that is a count of all items in the dataset that are invalid sliver surfaces, numberOfInvalidSelfIntersects that is a count of all items in the dataset that illegally intersect with themselves, and numberOfInvalidSelfOverlap that is all items in the dataset that illegally self-overlap.

X.4 Positional Uncertainty and Accuracy

X.4.1 Vertical Position Accuracy

Vertical Position Accuracy isn’t applicable for S-130.

X.4.2 Horizontal Position Accuracy

Horizontal Position Accuracy is applicable for S-130 and follows the guidelines from S-100 Part 4c.

S-130 products must be tested with Horizontal Position Accuracy checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. it is allowable to publish

the data with a quality statement which indicates non-conformance.

In terms of Horizontal Position Accuracy, S-130 products shall at least populate linearMapAccuracy2Sigma that is the half length of the interval defined by an upper and lower limit in which the true value lies with probability 95%.

X.4.3 Gridded Data Positional Accuracy

Gridded Data Position Accuracy isn't applicable for S-130.

X.5 Thematic Accuracy

X.5.1 Thematic Classification Correctness

Thematic Classification Correctness is applicable for S-130 and follows the guidelines from S-100 Part 4c.

S-130 products must be tested with Thematic Classification Correctness checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. Data should only be published if it passes the test.

In terms of Thematic Classification Correctness, S-130 products shall at least populate miscalculationRate that is the number of incorrectly classified features in relation to the number of features that are supposed to be there.

X.5.2 Non-Quantitative Attribute Accuracy

Non-Quantitative Attribute Accuracy is applicable for S-130 and follows the guidelines from S-100 Part 4c.

S-130 products must be tested with Non-Quantitative Attribute Accuracy checks prior to release by the data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. it is allowable to publish the data with a quality statement which indicates non-conformance.

The accuracy of non-quantitative attributes can be correct or incorrect. S-130 products shall at least populate numberOfIncorrectAttributeValues that is a count of all attribute values where the value is incorrect.

X.5.3 Quantitative Attribute Accuracy

Quantitative Attribute Accuracy is applicable for S-130 and follows the guidelines from S-100 Part 4c.

S-130 products must be tested with Quantitative Attribute Accuracy checks prior to release by the

data producer. The data producer must review the check results and address any issues to ensure sufficient quality of the data products. The checks are listed in Annex X. it is allowable to publish the data with a quality statement which indicates non-conformance.

The accuracy of quantitative attributes can be measured in terms of uncertainty intervals. S-130 products shall at least populate attributeValueUncertainty2Sigma that is half the length of the interval defined by an upper and lower limit in which the true value for the quantitative attribute lies with a probability of 90%.

X.6 Temporal Quality

X.6.1 Temporal Consistency

Temporal Consistency isn't applicable for S-130.

X.6.2 Temporal Validity

Temporal Validity isn't applicable for S-130.

X.6.3 Temporal Accuracy

Temporal Accuracy isn't applicable for S-130.

X.7 Aggregation

Aggregation isn't applicable for S-130.

X.8 Quality measure elements

The data quality measures recommended in S-97 (Part C) and their applicability in S-130 are indicated in Table X.1 below. NA indicates the measure is not applicable. The application schema above has indicated how the data quality elements will be related to the data items, and the encoding description below will indicate how the quality elements will be encoded.

Table X.1 - IHO recommended quality elements and their relevance to S-130

No.	Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Scope in S-130
1	Completeness / Commission	Excess data present in a dataset, as described by the scope.	numberOfExcessItems / This data quality measure indicates the number of items in the dataset, that should not have been present in the dataset.	dataset/dataset series	All features and info types
2	Completeness / Commission	Excess data present in a dataset, as described by the scope.	numberOfDuplicateFeatureInstances / This data quality measure indicates the total number of exact duplications of feature instances within the data.	dataset/dataset series	All features and info types

3	Completeness / Omission	Data absent from the dataset, as described by the scope.	numberOfMissingItems / This data quality measure is an indicator that shows that a specific item is missing in the data.	dataset/dataset series/spatial object type	All features and info types
4	Logical Consistency / Conceptual Consistency	Adherence to the rules of a conceptual schema.	numberOfInvalidSurfaceOverlaps / This data quality measure is a count of the total number of erroneous overlaps within the data. Which surfaces may overlap and which must not is application dependent. Not all overlapping surfaces are necessarily erroneous.	spatial object / spatial object type	Features with surface geometry; spatial objects of type surface
5	Logical Consistency / Domain Consistency	Adherence of the values to the value domains.	numberOfNonconformantItems / This data quality measure is a count of all items in the dataset that are not in conformance with their value domain.	spatial object / spatial object type	All features and info types
6	Logical Consistency / Format Consistency	Degree to which data is stored in accordance with the physical structure of the data set, as described by the scope	physicalStructureConflicts / This data quality measure is a count of all items in the dataset that are stored in conflict with the physical structure of the dataset.	dataset/dataset series	All features and info types
7	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	rateOfFaultyPointCurveConnections / This data quality measure indicates the number of faulty link-node connections in relation to the number of supposed link-node connections. This data quality measure gives the erroneous point-curve connections in relation to the total number of point-curve connections.	spatial object / spatial object type	Features with curve geometry; spatial objects of curve types
8	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	numberOfMissingConnectionsUndershoots / This data quality measure is a count of items in the dataset within the parameter tolerance that are mismatched due to undershoots.	spatial object / spatial object type	Features with curve geometry; spatial objects of curve types

9	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	numberOfMissingConnectionsOvershoots / This data quality measure is a count of items in the dataset within the parameter tolerance that are mismatched due to overshoots.	spatial object / spatial object type	Features with curve geometry; spatial objects of curve types
10	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	numberOfInvalidSlivers / This data quality measure is a count of all items in the dataset that are invalid sliver surfaces. A sliver is an unintended area that occurs when adjacent surfaces are not digitized properly. The borders of the adjacent surfaces may unintentionally gap or overlap to cause a topological error.	dataset / dataset series	Features with surface geometry; spatial objects of type surface
11	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	numberOfInvalidSelfIntersects / This data quality measure is a count of all items in the dataset that illegally intersect with themselves.	spatial object / spatial object type	Features with surface geometry; spatial objects of type surface
12	Logical Consistency / Topological Consistency	Correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope.	numberOfInvalidSelfOverlap / This data quality measure is a count of all items in the dataset that illegally self-overlap.	spatial object / spatial object type	Features with surface geometry; spatial objects of type surface
13	Positional Accuracy / Vertical Position Accuracy	Closeness of reported coordinative values to values accepted as or being true.	linearMapAccuracy2Sigma / Half length of the interval defined by an upper and lower limit in which the true value lies with probability 95%.	spatial object / spatial object type	NA. S-130 does not include vertical measurements.
14	Positional Accuracy / Horizontal Position Accuracy	Closeness of reported coordinative values to values accepted as or being true.	linearMapAccuracy2Sigma / Half length of the interval defined by an upper and lower limit in which the true value lies with probability 95%.	spatial object / spatial object type	Objects that have a horizontal coordinate values associated.

15	Positional Accuracy / Gridded Data Position Accuracy	Closeness of reported coordinative values to values accepted as or being true.	RMSerrorofplanimetry / Radius of a circle around the given point, in which the true value lies with probability P.	spatial object / spatial object type	NA.
16	Temporal Quality / Temporal Consistency	Consistency with time.	Correctness of ordered events or sequences, if reported.	dataset/dataset series/spatial object type	NA.
17	Thematic Accuracy / ThematicClassificationCorrectness	Comparison of the classes assigned to features or their attributes to a universe of discourse.	miscalculationRate / This data quality measure indicates the number of incorrectly classified features in relation to the number of features that are supposed to be there. [Adapted from ISO 19157] This is a RATE which is a ratio, and is expressed as a REAL number representing the rational fraction corresponding to the numerator and denominator of the ratio. For example, if there are 1 items that are classified incorrectly and there are 100 of the items in the dataset then the ratio is 1/100 and the reported rate = 0.01.	dataset/dataset series/spatial object type	All features and info types
18	Aggregation Measures / Aggregation Measures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationPassed / This data quality measure is a boolean indicating that all requirements in the referred data product specification are fulfilled.	dataset/dataset series/spatial object type	NA
19	Aggregation Measures / Aggregation Measures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationFailRate / This data quality measure is a number indicating the number of data product specification requirements that are not fulfilled by the current product/dataset in relation to the total number of data product specification requirements.	dataset/dataset series/spatial object type	NA