**DQWG13-10.1A**

**13th DQWG MEETING**

**Monaco, 15-19 Jan 2018**

**Paper for consideration by the Data Quality Working Group (DQWG)**

**Review S-100 section 4C and ISO and INSPIRE standards**

**Submitted by:** NL

**Executive summary: Review S-100 section 4C and ISO and INSPIRE standards**

**Related documents:** S\_100\_V3.0.0\_…2017,

 19157\_ISO-TC211\_N3521\_Geographic\_information.pdf

 INSPIRE Technical Specifications <check>

**Related projects:** HSSC-8 ref A: Review ISO 19113, Geographic Information- Quality Principles, ISO 19114, Geographic Information-Quality Evaluation Procedures, and ISO 19115, Geographic Information - Metadata and propose relevant enhancements and amendments for incorporation in S-100 “IHO Universal Hydrographic Data Model” (IHO Task 2.5.2)

**Date** 18 July 2017

**Introduction / background**

The International Organization for Standardization (ISO) - Technical Committee 211 (ISO TC211) deals with the development of standards and specifications for the geospatial domain. The International Hydrographic Organization (IHO) is a Class A liaison member of ISO/TC211 and participates in its standards development and maintenance Working Groups. The ISO/TC211 19100 series of standards and specifications have been used for the development of the IHO S-100 Universal Hydrographic Data Model.

The DQWG has been tasked by HSSC-8 to review the dependent ISO standards on Data Quality and proposed crosswalk mappings to be delivered to the S-100WG. This paper addresses ISO 19157 and S-100\_Edition\_3.0.0, part 4c - Metadata - Data Quality.

This subject was presented and discussed at DQWG meeting12. Outcome of the discussion was that the paper at meeting 12 needed further updating. This paper is the result of that discussion. DQWG has worked by correspondence to finalize this paper and to be delivered to S-100WG.

**Analysis/Discussion**

ISO 19157 was prepared by Technical Committee ISO/TC 211, Geographic information/Geomatics. This second edition cancels and replaces ISO/TS 19138:2006, ISO 19114:2003 and ISO 19113:2002, which have been technically revised.

S-100 — UNIVERSAL HYDROGRAPHIC DATA MODEL, Edition 3.0.0, … 2017 has been published by the IHO and is available on www.iho.int. The chapter that deals with data quality is described in S-100 — Part 4c, Metadata - Data quality on pages 167 to 188.

As ISO standard 19138 is now replaced by ISO standard 19157, all references in the S-100 documentation to ISO 19138 should be replaced by ISO 19157. The following chapters are impacted:

* Contents
* 4c-2 References
* 4c-3 Content
* 4c-3.1 ISO 19138 Quality and UML Classes
* Appendix 4c-A
* Appendix 4c-B
* Appendix 4c-C

|  |
| --- |
| Contents  |

*Existing* *text*: “4c-3.1 ISO19138 Quality Measures and UML Classes”

*Proposal*: “4c-3.1 ISO 19157 Geographic Information - Data Quality”

|  |
| --- |
| 4c-2 Reference  |

*Existing* *text*: “ISO 19138, Geographic Information - Quality Measures.”

*Proposal*: “ISO 19157, Geographic Information - Data Quality”

|  |
| --- |
| 4c-3 Content |

*Existing* *text*: “This document describes elements for quality measures as defined and described in ISO 19138.”

*Proposal*: “This document describes elements for quality measures as defined and described in ISO 19157.”

|  |
| --- |
| 4c-3.1 ISO19138 Quality Measures and UML Classes  |

*Existing* *text* header: “ISO 19138 Quality Measures and UML Classes”

*Proposal*: “ISO 19157, Geographic Information - Data Quality”

*Existing* text: “Full descriptions of these measures are contained in ISO 19138 Geographic Information Data Quality Measures.”

*Proposal*: “Full descriptions of these measures are contained in ISO 19157 Geographic Information - Data Quality.”

*Existing* text: “Additional quality measures may be described in a register of quality measures as described in ISO 19138 Annex B.”

*Proposal*: “Additional quality measures may be described in a register of quality measures as described in ISO 19157 Annex D.”

|  |
| --- |
| Appendix 4c-A |

*Existing* *figure*: “Figure 4c-A-2 — Data Quality Measure Registry UML (from ISO 19138)”

*Proposal*: “Figure 2 on page 7 of document 19157\_ISO-TC211\_N3521\_Geographic\_Information.pdf”

|  |
| --- |
| Appendix 4c-B Hydrographic Quality Metadata profile Data Dictionary |

*Existing* *text*:

1) The first column “ISO LineNo.” refers to the line numbers in the ISO 19115 Standard, however as this profile does not use all the 19115 elements, line numbers may not always be contiguous.”

*Proposal*:

1) The first column “ISO LineNo.” refers to the line numbers in the ISO 19115 Standard, however as this profile does not use all the 19115 elements, line numbers may not always be contiguous.”

*Existing* *text*:

2) Name/role name is a label assigned to a metadata entity or to a metadata element. Further columns could give the name or meaning in other languages.

*Proposal*:

2) Name/role name is a label assigned to a class or class attribute. Class attribute names are unique within a class. Role names are used to identify abstract model associations and are preceded by “Role name:” to distinguish them from other class attributes. Further columns could give the name or meaning in other languages.

*Motivation: ISO19157 text*:

A label assigned to class or class attribute. Class names start with an upper case letter. Spaces do not appear in a class name. Instead, multiple words are concatenated, with each new subword starting with a capital letter (example: XnnnYmmm). Class names are unique within the entire data dictionary of this International Standard. Class attribute names are unique within a class, not the entire data dictionary of this International Standard. Class attribute names are made unique, within an application, by the combination of the class name and class attribute names. Role names are used to identify abstract model associations and are preceded by “Role name:” to distinguish them from other class attributes. Names and role names may be in a language other than that used in this International Standard.

*Existing* *text*:

3) Definition column provides a description of the metadata entity/element.

*Proposal*:

3) Definition is the class or class attribute description.

*Motivation: ISO19157 text:*

This is the class or class attribute description

*Existing* *text*:

4) The obligation descriptor provides an indication of whether a metadata entity or metadata element shall always be documented or will only sometimes be documented. This descriptor may have the following values: M (mandatory), C (conditional), or O (optional).

*Proposal*:

4) The obligation descriptor provides an indication whether a class or class attribute shall always be documented in the dataset or sometimes be documented (i.e. contains value(s)). This descriptor may have the following values: M (mandatory), C (conditional), or O (optional).

*Motivation: ISO19157 text:*

This is a descriptor indicating whether a class or class attribute shall always be documented in the dataset or sometimes be documented (i.e. contains value(s)). This descriptor may have the following values: M (mandatory), C (conditional), or O (optional).

*Existing* *text*:

5) The Occurrence column specifies the maximum number of instances the metadata entity or the metadata element may have. Single occurrences are shown by “1”; repeating occurrences are represented by “N”. Fixed number occurrences other than one are allowed, and will be represented by the corresponding number (that is “2”, “3”...etc.).

*Proposal*:

5) The Occurrence column specifies the maximum number of instances the class, class attribute or association may have. Single occurrences are shown by “1”; repeating occurrences are represented by “N”. Fixed number occurrences other than one are allowed, and will be represented by the corresponding number (i.e. “2”, “3”…etc).

*Motivation: ISO19157 text:*

Specifies the maximum number of instances the class, class attribute or association may have. Single occurrences are shown by “1”; repeating occurrences are represented by “N”. Fixed number occurrences other than one are allowed, and will be represented by the corresponding number (i.e. “2”, “3”...etc).

*Existing* *text*:

6) Data type specifies a set of distinct values for representing the metadata elements; for example, integer, real, string, DateTime, and Boolean. The data type attribute is also used to define metadata entities, stereotypes, and metadata associations.

*Proposal*:

6) Data type specifies a set of distinct values for representing the class attributes; for example, integer, real, string, DateTime, and Boolean. The data type column is also used to define classes, stereotypes, and class associations.

*Motivation: ISO19157 text:*

Specifies a set of distinct values for representing the class attributes; for example, integer, real, string, DateTime, and Boolean. The data type column is also used to define classes, stereotypes, and class associations.

*Existing* *text*:

7) Domain - for an entity, the domain indicates the line numbers covered by that entity.

*Proposal*:

7) Domain - for a class (shaded rows), the domain indicates the line numbers covered by class attributes and associations for that class.

*Motivation: ISO 19157:*

For a class (shaded rows), the domain indicates the line numbers covered by class attributes and associations for that class.

For a class attribute or association, the domain specifies the values allowed or the use of free text. “Free text” indicates that no restrictions are placed on the content of the field. Integer-based codes shall be used to represent values for domains containing codelists.

|  |
| --- |
| Appendix 4c-C Hydrographic Quality Metadata Attribute Definitions |

These definitions are taken from ISO 19138 which is now replaced by ISO 19157 Annex D (normative) List of standardized data quality measures. The measures listed agree with ISO 19157. However the order in which they are listed deviate from ISO19157 and is also not logical. ISO19157 categories the data quality measures as follows:

* Completeness
* Logical Consistency
* Positional Accuracy
* Temporal Quality
* Thematic Accuracy
* Aggregation Measures

The items in S-100 Appendix 4c-C have been drafted in this order.

Below is a proposal for the text of Appendix 4c-C:

**Completeness**

**DQ\_CompletenessCommission**

Excess data present in a data set. [Per ISO 19115]

**Public Attributes:**

excessItem[0..1] : Boolean

This data quality measure indicates that an item is incorrectly present in the data. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that the item is in excess.

**numberOfExcessItems[0..1] : Integer**

This data quality measure indicates the number of items in the dataset, that should not have been present in the dataset. [Adapted from ISO 19157]

This is an INTEGER count of the number of excess items.

**rateOfExcessItems[0..1] : Real**

This data quality measure indicates the number of excess items in the dataset in relation to the number of items that should have been present. [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 5 measured values and 4 valid values then the ratio is 5/4 and the reported rate = 1.25.

**numberOfDuplicateFeatureInstances[0..1] : Integer**

This data quality measure indicates the total number of exact duplications of feature instances within the data. This is a count of all items in the data that are incorrectly extracted with duplicate geometries. [Adapted from ISO 19157]

This is an integer representing the error count.

**DQ\_CompletenessOmission**

This data absent from a data set. [Per ISO 19115]

**Public Attributes:**

**missingItem[0..1] : Boolean**

This data quality measure is an indicator that shows that a specific item is missing in the data. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that an item is missing.

**numberOfMissingItems[0..1] : Integer**

This data quality measure indicates the count of all items that should have been in the dataset and are missing. [Adapted from ISO 19157]

This is an INTEGER count of the number of missing items.

**rateOfMissingItems[0..1] : Real**

This data quality measure indicates the number of missing items in the dataset in relation to the number of items that should have been present. [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 3 measured values and 5 values are required the ratio is 3/.5 and the reported rate = 0.6.

**Logical Consistency**

**DQ\_LogicalConsistency**

**DQ\_ConceptualConsistancy**

Adherence to the rules of a conceptual schema. [Per ISO 19115]

Public Attributes:

**conceptualSchemaNonCompliance[0..1] : Boolean**

This data quality measure is an indication that an item is not compliant to the rules of the relevant conceptual schema. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that an item is not compliant with the rules of the conceptual schema.

**conceptualSchemaCompliance[0..1] : Boolean**

This data quality measure is an indication that an item complies with the rules of the relevant conceptual schema. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that an item is in compliance with the rules of the conceptual schema.

**numberOfNonCompliantItems[0..1] : Integer**

This data quality measure is a count of all items in the dataset that are noncompliant to the rules of the conceptual schema. If the conceptual schema explicitly or implicitly describes rules, these rules have to be followed. Violations against such rules, for example; can be invalid placement of features within a defined tolerance, duplication of features and invalid overlap of features. [Adapted from ISO 19157]

This is an integer count.

**numberOfInvalidSurfaceOverlaps[0..1] : Integer**

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. When reporting this data quality measure the types of feature classes corresponding to the illegal overlapping surfaces have to be reported as well. [Adapted from ISO 19157]

The allowable topological levels are described in the IHO/DGIWG joint profile of ISO 19107 Geographic Information Spatial Schema. Which particular topological structure may be used with a specific dataset is defined in the Product Specification for that type of data product, for example "Chain Node Topology" for IHO S-101.

This is an error count.

**nonComplianceRate[0..1] : Real**

This data quality measure indicates the number of items in the dataset that are noncompliant to the rules of the conceptual schema in relation to the total number of these items that are expected to be in the dataset. [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 5 items that are non compliant and there are 100 of the items in the dataset then the ratio is 5/100 and the reported rate = 0.05.

**complianceRate[0..1] : Real**

This data quality measure indicates the number of items in the dataset that are in compliance with the rules of the conceptual schema in relation to the total number of these items that are expected to be in the dataset. [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 95 items that are compliant and there are 100 of the items in the dataset then the ratio is 95/100 and the reported rate = 0.95.

**DQ\_DomainConsistancy**

Adherence of the values to the value domains. [Per ISO 19115]

**Public Attributes:**

valueDomainNonConformance[0..1] : Boolean

This data quality measure is an indication that an item is not in conformance with its value domain. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that an item is not in conformance with its value domain.

**valueDomainConformance [0..1] : Boolean**

This data quality measure is an indication that an item is conforming to its value domain. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that an item conforming to its value domain.

**numberOfNonconformantItems[0..1] : Integer**

This data quality measure is a count of all items in the dataset that are not in conformance with their value domain. [Adapted from ISO 19157]

This is an integer count.

**valueDomainConformanceRate[0..1] : Real**

This data quality measure indicates the number of items in the dataset that are in conformance with their value domain in relation to the total number of items in the dataset. [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 95 items that are in conformance and there are 100 of the items in the dataset then the ratio is 95/100 and the reported rate = 0.95.

**valueDomainNonConformanceRate[0..1] : Real**

This data quality measure indicates the number of items in the dataset that are not in conformance with their value domain in relation to the total number of items in the dataset. [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 5 items that are NOT in conformance and there are 100 of the items in the dataset then the ratio is 5/100 and the reported rate = 0.05.

**DQ\_FormatConsistancy**

Degree to which data is stored in accordance with the physical structure of the data set. [Per ISO 19115]

**Public Attributes:**

**physicalStructureConflicts[0..1] : Boolean**

This data quality measure is an indication that items are stored in conflict with the physical structure of the dataset. [Adapted from ISO 19157]

This is a BOOLEAN where TRUE indicates physical structure conflict

**physicalStructureConflictsNumber[0..1] : Integer**

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. [Adapted from ISO 19157]

This is an integer count.

**physicalStructureConflictRate[0..1] : Real**

This data quality measure indicates the number of items in the dataset that are stored in conflict with the physical structure of the dataset divided by the total number of items. [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 3 items that are in conflict and there are 100 of the items in the dataset then the ratio is 3/100 and the reported rate = 0.03.

**DQ\_TopologicalConsistency**

Measures of the topological consistency of geometric representations of features. [Adapted from ISO 19157]

Note: in ISO 19115, this is “Correctness of the explicitly encoded topological characteristics of a dataset”, but ISO 19157 states that the measures “will not serve as measures of the consistency of explicit descriptions of topology using the topological objects specified in ISO 19107”, and S-100 does not explicitly encode geometry.

**Public Attributes:**

**numberOfFaultyPointCurveConnections[0..1] : Integer**

This data quality measure is a count of the number of faulty point-curve connections in the dataset. A point curve connection exists where different curves touch. These curves have an intrinsic topological relationship that has to reflect the true constellation. For example, two point-curve connections exist when there should only be one. [Adapted from ISO 19157]

This is an integer count.

**rateOfFaultyPointCurveConnections[0..1] : Real**

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. [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 2 items that are faulty link-node connections and there are 100 of the connections in the dataset then the ratio is 2/100 and the reported rate = 0.02.

**numberOfMissingConnectionsUndershoots[0..1] : Integer**

This data quality measure is a count of items in the dataset within the parameter tolerance that are mismatched due to undershoots. [Adapted from ISO 19157]

This is an integer count.

**numberOfMissingConnectionsOvershoots[0..1] : Integer**

This data quality measure is a count of items in the dataset within the parameter tolerance that are mismatched due to overshoots. [Adapted from ISO 19157]

This is an integer count.

**numberOfInvalidSlivers[0..1] : Integer**

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. [Adapted from ISO 19157]

This is an integer count.

**numberOfInvalidSelfIntersects[0..1] : Integer**

This data quality measure is a count of all items in the dataset that illegally intersect with themselves. [Adapted from ISO 19157]

This is an integer count.

**numberOfInvalidSelfOverlaps[0..1] : Integer**

This data quality measure is a count of all items in the dataset that illegally self-overlap. [Adapted from ISO 19157]

This is an integer count.

**Positional accuracy**

**DQ\_AbsoluteExternalPositionalAccuracy**

Closeness of reported coordinative values to values accepted as or being true. [Per ISO 19115]

**Public Attributes:**

**meanValuePositionalUncertainties[0..1] : Real**

Mean value of the positional uncertainties for a set of positions where the positional uncertainties are defined as the distance between a measured position and what is considered as the corresponding true position. [Adapted from ISO 19157]

**BiasOfPositions[0..1] : Real**

This data quality measure is a bias of the positions for a set of positions where the positional uncertainties are defined as the deviation between a measured position and what is considered as the corresponding true position. [Adapted from ISO 19157]

**meanExcludingOutliers[0..1] : Real**

Mean value of the positional uncertainties, excluding outliers. For a set of points where the distance does not exceed a defined threshold, the arithmetical average of distances between their measured positions and what is considered as the corresponding true positions. [Adapted from ISO 19157]

**numberOfPositionalUncertaintiesAboveThreshold[0..1] : Integer**

Number of positional uncertainties above a given threshold for a set of positions. The errors are defined as the distance between a measured position and what is considered as the corresponding true position. [Adapted from ISO 19157]

**rateOfPositionalErrorsAboveThreshold[0..1] : Real**

Number of positional uncertainties above a given threshold for a set of positions in relation to the total number of measured positions. The errors are defined as the distance between the measured position and what is considered as the corresponding true position. [Adapted from ISO 19157]

**covarianceMatrix[0..1] : Real Matrix**

Symmetrical square matrix with variances of point coordinates on the main diagonal and covariances between these coordinates as off diagonal elements. [Adapted from ISO 19157]

**Vertical position uncertainties**

**linearErrorProbable[0..1] : Real**

Half length of the interval defined by an upper and lower limit in which the true value lies with probability 50%. [Adapted from ISO 19157]

**standardLinearError[0..1] : Real**

Half length of the interval defined by an upper and lower limit in which the true value lies with probability 68.3%. [Adapted from ISO 19157].

**linearMapAccuracy90%Level[0..1] : Real**

Half length of the interval defined by an upper and lower limit in which the true value lies with probability 90%. [Adapted from ISO 19157].

**linearMapAccuracy2Sigma[0..1] : Real**

Half length of the interval defined by an upper and lower limit in which the true value lies with probability 95%. [Adapted from ISO 19157].

**linearMapAccuracy3Sigma[0..1] : Real**

Half length of the interval defined by an upper and lower limit in which the true value lies with probability 99%. [Adapted from ISO 19157].

**nearCertainityLinearError[0..1] : Real**

Half length of the interval defined by an upper and lower limit in which the true value lies with probability 99.8%. [Adapted from ISO 19157].

**RMSError[0..1] : Real**

Standard deviation where the true value is not estimated from the observations but known apriori. [Adapted from ISO 19157].

**AbsoluteLineairError90%Level[0..1] : Real**

Absolute vertical accuracy of the data’s coordinates, expressed in terms of linear error at 90 % probability given that a bias is present [Adapted from ISO 19157]

**Horizontal position uncertainties**

**circularStandardDeviation[0..1] : Real**

Radius describing a circle in which the true point location lies with the probability of 39.4%. [Adapted from ISO 19157].

**circularErrorProbable[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 50 %

[Adapted from ISO 19157].

**circularMapAccuracyStandard[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 90%. [Adapted from ISO 19157].

**circularError95[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 95%. [Adapted from ISO 19157].

**circularNearCertaintyError[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 99.8%. [Adapted from ISO 19157].

**RMSErrorPlanimetry[0..1] : Real**

Radius of a circle around a given point, in which the true value lies with probability P. [Adapted from ISO 19157].

**ACE\_CE90[0..1] : Real**

The absolute horizontal accuracy of the data's coordinates expressed in terms of circular error at 90% probability given that a bias is present, per the equation in table D.50 in ISO 19157. [Adapted from ISO 19157].

**uncertaintyEllipse[0..1] : Record**

A 2D ellipse with the two main axes indicating the direction and magnitude of the highest and lowest uncertainty of a 2D point. The data values are a record of real numbers corresponding to "phi" the bearing of the major semi-axis, and "a" and "b" the length of the two axes, per the equations in Table D.52 of ISO 19157. [Adapted from ISO 19157].

**confidenceEllipse[0..1] : Record**

A 2D ellipse with the two main axes indicating the direction and magnitude of the highest and lowest uncertainty of a 2D point. The data values are a record of real numbers corresponding to "phi" the bearing of the major semi-axis, and "a" and "b" the length of the two axes, per the equations in Table D.53 of ISO 19157 and a significance level parameter. [Adapted from ISO 19157].

**Gridded data position accuracy**

**circularStandardDeviation[0..1] : Real**

Radius describing a circle in which the true point location lies with the probability of 39.4%. [Adapted from ISO 19157].

**circularErrorProbable[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 50 %

[Adapted from ISO 19157].

**circularMapAccuracyStandard[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 90%. [Adapted from ISO 19157].

**circularError95[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 95%. [Adapted from ISO 19157].

**circularNearCertaintyError[0..1] : Real**

Radius describing a circle, in which the true point location lies with the probability of 99.8%. [Adapted from ISO 19157].

**RMSErrorPlanimetry[0..1] : Real**

Radius of a circle around a given point, in which the true value lies with probability P. [Adapted from ISO 19157].

**ACE\_CE90[0..1] : Real**

The absolute horizontal accuracy of the data's coordinates expressed in terms of circular error at 90% probability given that a bias is present, per the equation in table D.50 in ISO 19157. [Adapted from ISO 19157].

**uncertaintyEllipse[0..1] : Record**

A 2D ellipse with the two main axes indicating the direction and magnitude of the highest and lowest uncertainty of a 2D point. The data values are a record of real numbers corresponding to "phi" the bearing of the major semi-axis, and "a" and "b" the length of the two axes, per the equations in Table D.52 of ISO 19157. [Adapted from ISO 19157].

**confidenceEllipse[0..1] : Record**

A 2D ellipse with the two main axes indicating the direction and magnitude of the highest and lowest uncertainty of a 2D point. The data values are a record of real numbers corresponding to "phi" the bearing of the major semi-axis, and "a" and "b" the length of the two axes, per the equations in Table D.53 of ISO 19157 and a significance level parameter. [Adapted from ISO 19157].

**Temporal Quality**

**DQ\_AccuracyOfATimeMeasurement**

Correctness of the temporal references of an item (reporting of error in time measurement). [Per ISO 19115]

Public Attributes:

**attributeValueUncertaintyMean[0..1] : Real**

This data quality measure indicates the attribute value of uncertainty where 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 50%. [Adapted from ISO 19157]

**attributeValueUncertainty1Sigma[0..1] : Real**

This data quality measure indicates the attribute value of uncertainty where 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 68.3%. [Adapted from ISO 19157]

**attributeValueUncertainty90[0..1] : Real**

This data quality measure indicates the attribute value of uncertainty where 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%. [Adapted from ISO 19157]

**attributeValueUncertainty2Sigma[0..1] : Real**

This data quality measure indicates the attribute value of uncertainty where 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 95%. [Adapted from ISO 19157]

**attributeValueUncertainty3Sigma[0..1] : Real**

This data quality measure indicates the attribute value of uncertainty where 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 99%. [Adapted from ISO 19157]

**attributeValueUncertainty4Sigma[0..1] : Real**

This data quality measure indicates the attribute value of uncertainty where 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 99.8%. [Adapted from ISO 19157]

**DQ\_TemporalConsistancy**

Correctness of ordered events or sequences, if reported. [Per ISO 19115]

**Public Attributes:**

**chronologicalError[0..1] : Boolean**

This is data quality measure indicates that an event is incorrectly ordered against the other events

[Adapted from ISO 19157]

This is a BOOLEAN where TRUE indicates true indicates that the event is incorrectly ordered.

**DQ\_TemporalValidity**

Validity of data with respect to time. [Per ISO 19115]

**Public Attributes:**

**valueDomainNonConformance[0..1] : Boolean**

This data quality measure is an indication that an item is not in conformance with its value domain. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that an item is not in conformance with its value domain.

**valueDomainConformance[0..1] : Boolean**

This data quality measure is an indication that an item is conforming to its value domain. [Adapted from ISO 19157]

This is a Boolean where TRUE indicates that an item is conforming to its value domain.

**numberOfNonConformantItems[0..1] : Integer**

This data quality measure is a count of all items in the dataset that are not in conformance with their value domain. [Adapted from ISO 19157]

This is an integer count.

**valueDomainConformanceRate[0..1] : Real**

This data quality measure indicates the number of items in the dataset that are in conformance with their value domain in relation to the total number of items in the dataset. [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.

**valueDomainNonConformanceRate[0..1] : Real**

This data quality measure indicates the number of items in the dataset that are not in conformance with their value domain in relation to the total number of items in the dataset. [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 5 items that are in conformance and there are 100 of the items in the dataset then the ratio is 5/100 and the reported rate = 0.05.

**Thematic Accuracy**

**DQ\_ThematicClassificationCorrectness**

Comparison of the classes assigned to features or their attributes to a universe of discourse. [Per ISO 19113]

For example, ground truth or reference dataset.

**Public Attributes:**

**numberOfIncorrectlyClassifiedItems[0..1] : Integer**

This data quality measure is a count of the number of incorrectly classified features. [Adapted from ISO 19157]

This is an integer count.

**miscalculationRate[0..1] : Real**

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.

**misclassificationMatrix[0..1] : Integer Matrix**

This data quality measure is a matrix of integer numbers that indicates the number of items of class (i) classified as class (j). The misclassification matrix is a quadratic matrix with n columns and n rows where n denotes the number of classes under consideration. MCM (i,j) = (# items of class (i) classified as class (j). The diagonal elements of the misclassified matrix contain the correctly classified items, and the off diagonal items contain the number of misclassified errors. [Adapted from ISO 19157]

**relativeMiscalculationMatrix[0..1] : Real Matrix**

This data quality measure is a matrix of real numbers that indicates the number of items of class (i) classified as class (j) divided by the number of items of class (i) \* 100 represented as a percentage. The misclassification matrix has n columns and n rows where n denotes the number of classes under consideration. RMCM (i,j) = (# items of class (i) classified as class (j) / number of items of class (i) ) \*100. [Adapted from ISO 19157]

**kappaCoefficient[0..1] : Real**

This data quality measure is real number coefficient to quantify the proportion of agreement of assignments to classes by removing misclassifications. [Adapted from ISO 19157].

**DQ\_QuantativeAttributeAccuracy**

The data quality measures for the data quality element quantitative attribute accuracy

[Adapted from ISO 19157]

**Public Attributes:**

**AttributeValueUncertainty1Sigma[0..1] : Real**

This data quality measure is a measure for the half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 68,3 % [Adapted from ISO 19157]

**AttributeValueUncertainty50%Level[0..1] : Real**

This data quality measure is a measure for the half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 50 % [Adapted from ISO 19157]

**AttributeValueUncertainty90%Level[0..1] : Real**

This data quality measure is a measure for the half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 90 % [Adapted from ISO 19157]

**AttributeValueUncertainty2Sigma[0..1] : Real**

This data quality measure is a measure for the half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 95 % [Adapted from ISO 19157]

**AttributeValueUncertainty3Sigma[0..1] : Real**

This data quality measure is a measure for the half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 99 % [Adapted from ISO 19157]

**AttributeValueUncertainty4Sigma[0..1] : Real**

This data quality measure is a measure for the half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 99,8 % [Adapted from ISO 19157]

**Aggregation Measures**

**DQ\_AggregationMeasures**

In a data product specification, several requirements are set up for a product to conform to the specification. [Adapted from ISO 19157].

**Public Attributes:**

**DataProductSpecificationPassed[0..1] : Boolean**

This data quality measure is a boolean indicating that all requirements in the referred data product specification are fulfilled. [Adapted from ISO 19157].

This is a Boolean where TRUE indicates that all the requirements in the referred data product specification are fulfilled.

**DataProductSpecificationFailCount[0..1] : Integer**

This data quality measure is a number indicating the number of data product specification requirements that are not fulfilled by the current product/dataset. [Adapted from ISO 19157].

This is an integer count.

**DataProductSpecificationPassCount[0..1] : Integer**

This data quality measure is a number indicating the number of data product specification requirements that are fulfilled by the current product/dataset. [Adapted from ISO 19157].

This is an integer count.

**DataProductSpecificationFailRate[0..1] : Real**

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 . [Adapted from ISO 19157].

**DataProductSpecificationPassRate[0..1] : Real**

This data quality measure is a number indicating the number of data product specification requirements that are fulfilled by the current product/dataset in relation to the total number of data product specification requirements . [Adapted from ISO 19157].

**Observations:**

The existing text in Appendix 4c-C is almost correct but some public attributes are missing.

Uncertainty levels are wrongly described. 1Sigma = 68%, 2Sigma = 95%, 3Sigma = 99,8% 4Sigma = 99,99%. 90% and 99% are also used but these are not specific Sigma levels.

In Hydrography the 2Sigma Level (95%) is commonly used to indicate quality in terms of uncertainty levels.

Name/role name: Class names are unique within the entire data dictionary of this International Standard. Class attribute names are unique within a class, not the entire data dictionary of this International Standard. Class attribute names are made unique, within an application, by the combination of the class name and class attribute names.

In S101\_Data Classification and Encoding Guide\_Final 0.02 Clean the attribute DRVAL1 (depth range minimum value) appears in the following classes:

Quality of bathymetric data (primitive = surface)

Quality of survey (primitive = curve, surface)

Gate (primitive = point, curve, surface)

Berth (primitive = point, curve, surface)

Dry dock (primitive = surface)

Floating dock (primitive = point, curve, surface)

Dredged area (primitive = surface)

Swept area (primitive = surface)

Depth area (primitive = surface)

Submarine/land pipelines (primitive = curve)

Recommended track (primitive = curve)

Fairway (primitive = surface)

Recommended routes (primitive = curve)

Two-way part route (primitive = surface)

Deep water route centreline (primitive = curve)

Deep water route part (primitive = surface)

The IHO definition of depth range minimum value is: Depth range is the depth from a specified sounding datum as a depth interval bounded by the minimum (shoalest) and maximum (deepest) depth values. Depth range minimum value defines the minimum (shoalest) value of a depth range.

As can be seen from the example above, the same attribute name is used throughout different classes with the same definition and therefor the same value if on the same location. However when two different classes are stacked above each other, the same attribute cannot have the same value. This is an issue addressed in:

DQWG11-04.4A\_Guidance\_on\_stacked\_quality\_of\_bathymetric\_data

**Conclusions:**

The change from ISO19138 to ISO19157 has limited impact on S-100 Edition 3.0.0 at Part 4c “Metadata - data quality.” The order in which the measures and associated public attributes are listed may be re-ordered to be in line with ISO19157. Some public attributes apparent in ISO19157 are missing from S-100 Part 4c. This paper has drafted a proposed text in which these issues have been addressed.

INSPIRE standards have been verified and are based on ISO19138. There is no direct conflict between INSPIRE quality measures and the ones listed in S-100 part 4C as proposed.

**Recommendations:**

* S-100WG to adapt the proposed changes of Part 4C.
* To make the 2Sigma Level (95%) the preferred option to describe the quality in terms of uncertainty level.
* To note the difficulties that arise when using the same attribute name in different classes if the value of the attribute is not consistently the same.

**Actions required from DQWG:**

* Review and revise this paper
* Provide comments on this paper
* Agree on pre-delivery to S-100WG at HSSC-9
* To discuss the outcome of HSSC-9 and response from S-100WG at DQWG meeting 13