12th DQWG MEETING
The Hague, 13-15 June 2017

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_V2.0.0_June-2015, 211n2492_Text_for_ISO_19115-2, ISO-TC211_N3521_ISO-FDIS_19157_Geographic_informati
Related projects:	S-100 development
Date	06 June 2017

Introduction / background

The DQWG was tasked by the S-100WG to review S-100 part 4C, Metadata – Data Quality. This paper describes the result of this review. One item that should be discussed is maintaining IHO standards in line with ISO when certain ISO standards reach end-of-life cycle.
The first outcome of this review is that S-100 adopts some of the ISO 19157 standard but is not complete and sometimes uses different wording for its definitions.

S_100_V2.0.0_June-2015 part 4C

This chapter contains the following information:
4c-1: Scope
4c-2: References
4c-3: Content
4c-3.1 ISO 19138 Quality Measures and UML Classes
4c-3.2 Core Metadata
Appendix 4c-A Hydrographic Quality Metadata profile, UML diagrams
Appendix 4c-B Hydrographic Quality Metadata profile Data Dictionary
Appendix 4c-C Hydrographic Quality Metadata Attribute Definitions

4c-1: Scope

No comments on this paragraph.

4c-2: References

ISO 19138, Geographic information – Quality measures is now obsolete and replaced by ISO 19157, Geographic information – Data Quality

4c-3: Content

ISO 19138 should be replaced by ISO 19157.

4c-3.1 ISO 19138 Quality Measures and UML Classes
Paragraph 3.1 states: “Additional quality measures may be described in a register of quality measures as described in ISO 19138 Annex-B.”

ISO 19157 has the following Annexes:
Annex A (normative) Abstract Test Suites
Annex B (informative) Data quality concepts and their use
Annex C (normative) Data dictionary for data quality
Annex D (normative) List of standardized data quality measures
Annex E (informative) Evaluating and reporting data quality
Annex F (informative) Sampling methods for evaluating
Annex G (normative) Data quality basic measures
Annex H (informative) Management of data quality measures
Annex I (informative) Guidelines for the use of Quality Elements
Annex J (informative) Aggregation of data quality results

The reference to ISO 19138 Annex-B is no longer valid. ISO 19157 Annex-D seems to most suitable one to be used.

4c-3.2 Core Metadata
No comments.

Appendix 4c-A Hydrographic Quality Metadata profile, UML diagrams
The present diagram in S-100 does not reflect the UML diagrams in ISO 19157, ref figure 1 page 16 and figure 2 page 17. These diagrams are shown below:

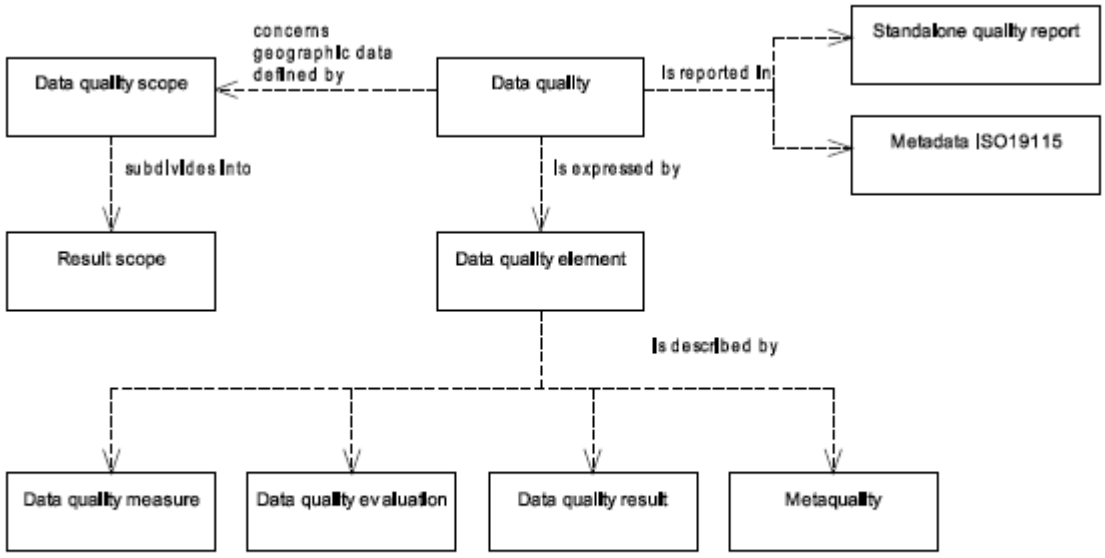

Figure 1 — Conceptual model of quality for geographic data

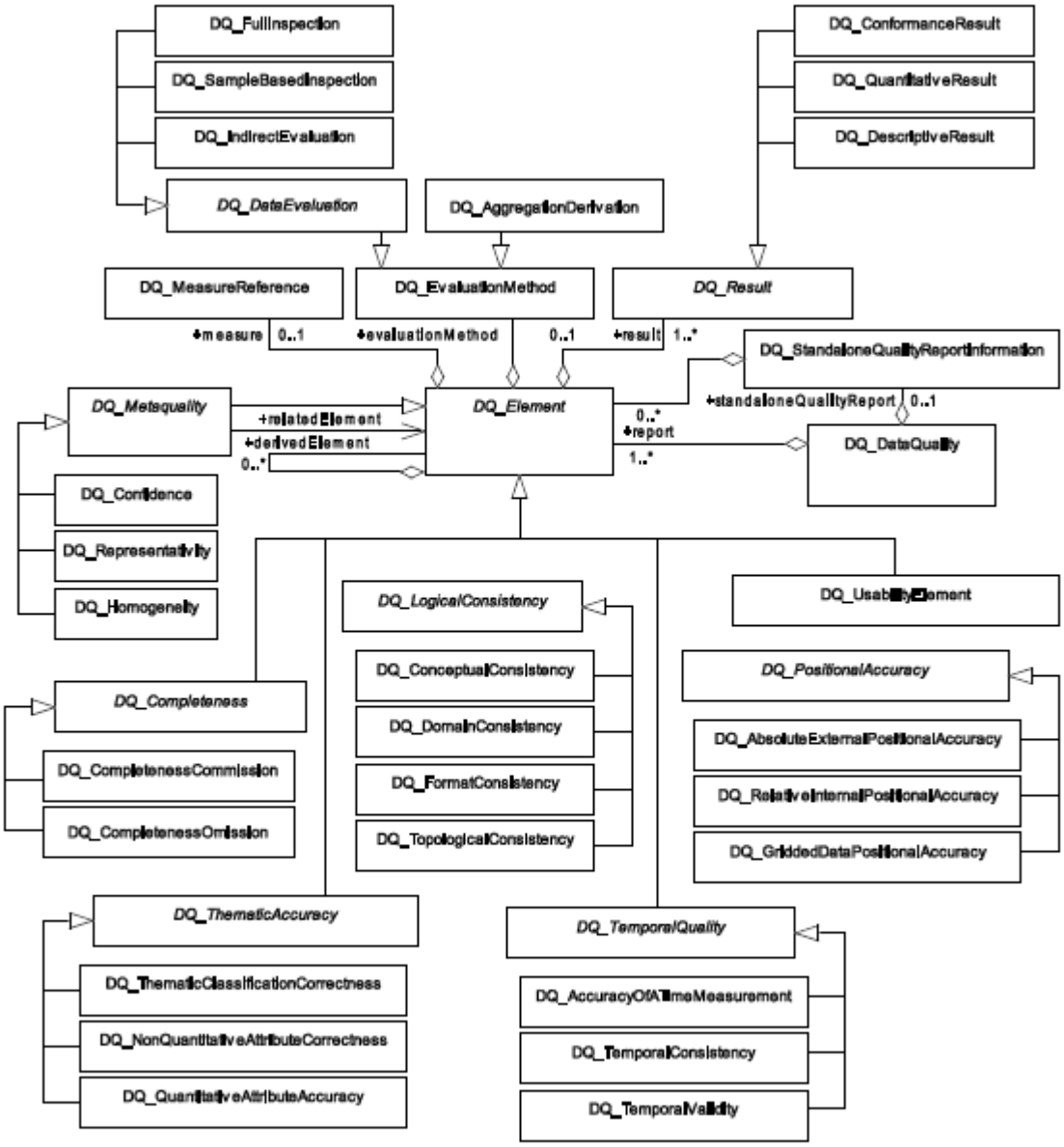

Figure 2 — Overview of the components of data quality

Appendix 4c-B Hydrographic Quality Metadata profile Data Dictionary
Text of S100 is compared with ISO-TC211_N3521_ISO-FDIS_19157_Geographic_Information, Annex C – Data dictionary for data quality

ISO LineNo is not present in ISO 19157

Name/role name: . A label assigned to a class or class attribute. Class names are unique within the entire data dictionary. 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.*

Definition: This is the class or class attribute description*

Obligation/condition: 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).

Maximum occurrence: 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).

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.

Domain: 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.*

All above are taken from ISO 19157. This items marked * are different in S-100.

The table in S-100 page 140 (part 4c- Metadata – Data Quality page 5) should be renewed entirely in line with ISO 19157.

Appendix 4c-C: Hydrographic Quality Metadata Attribute Definitions
The following classes have been defined in S-100:

Class	Description
DQ_AbsoluteExternalPositionalAccuracy	Closeness of reported coordinative values to values accepted as or being true
DQ_AccuracyOfATimeMeasurement	Correctness of the temporal references of an item (reporting of error in time measurement)
DQ_CompletenessCommission	Excess data present in a data set
DQ_CompletenessOmission	This data absent from a data set
DQ_ConceptualConsistancy	Adherence to the rules of a conceptual schema
DQ_DomainConsistancy	Adherence of the values to the value domains
DQ_FormatConsistancy	Degree to which data is stored in accordance with the physical structure of the data set.
DQ_GriddedDataPositionalAccuracy	Closeness of gridded data position values to values to values accepted as or being true.
DQ_NonQuantitativeAttributeAccuracy	Correctness of non-quantitative attribute
DQ_QuantitativeAttributeAccuracy	Accuracy of a quantitative attribute

DQ_RelativeInternalPositionalAccuracy	Closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true.
DQ_TemporalConsistancy	Correctness of ordered events or sequences, if reported.
DQ_TemporalValidity	Validity of data with respect to time
DQ_ThematicClassificationCorrectness	Comparison of the classes assigned to features or their attributes to a universe of discourse
DQ_TopologicalConsistency	Measures of the topological consistency of geometric representations of features.

ISO-TC211_N3521_ISO-FDIS_19157_Geographic_Information, Annex D – List of standardized data quality measures reports:

Group	**Class**	**Measure**
Completeness	Commission	Excess item
		Number of excess items
		Rate of excess items
		Number of duplicate features instances
	Omission	Missing item
		Number of missing items
		Rate of missing items
Logical consistency	Conceptual consistency	Conceptual schema non-compliance
		Conceptual schema compliance
		Number of items not compliant with the rules of the conceptual schema
		Number of invalid overlaps or surfaces
		Non compliance rate with respect to the rules of the conceptual schema
		Compliance rate with the rules of the conceptual schema
	Domain consistency	Value domain non-compliance
		Value domain conformance
		Number of items not in conformance with their value domain
		Value domain conformance rate
		Value domain non-conformance rate
	Format consistency	Physical structure conflicts
		Physical structure conflicts number
		Physical structure conflict rate
	Topological consistency	Number of faulty point curve corrections
		Rate of faulty point curve corrections
		Number of missing connections due to undershoots
		Number of missing connections due to overshoots
		Number of invalid sliver
		Number of invalid self-intersect errors
		Number of invalid self-overlap errors
Positional accuracy	Absolute or external accuracy	Mean value of positional uncertainties
		Bias of positions
		Mean value of positional uncertainties excluding outliers
		Number of positional uncertainties above a given threshold

		Rate of positional errors above a given threshold
		Covariance matrix
	Vertical positional uncertainties	Linear error probable
		Standard linear error
		Linear map accuracy at 90% significance level
		Linear map accuracy at 95% significance level
		Linear map accuracy at 99% significance level
		Near certainty linear error
		Root mean square error
		Absolute linear error at 90% significance level of biased vertical data (NATO)
		Absolute linear error at 90% significance level of biased vertical data
	Horizontal positional uncertainties	Circular standard deviation
		Circular error probable
		Circular map accuracy standard
		Circular error at 95% significance level
		Circular near certainty error
		Root mean square error of planimetry
		Absolute circular error at 90% significance level of biased data
		Absolute circular error at 90% significance level of biased data (NATO)
		Uncertainty ellipse
		Confidence ellipse
	Relative or internal accuracy	Relative vertical error
		Relative horizontal error
	Gridded data position accuracy	Use “horizontal positional uncertainties”
Temporal quality	Accuracy of a time measurement	Time accuracy at 68,3% significance level
		Time accuracy at 50% significance level
		Time accuracy at 90% significance level
		Time accuracy at 95% significance level
		Time accuracy at 99% significance level
		Time accuracy at 99,8% significance level
	Temporal consistency	Chronological error
	Temporal validity	See “domain consistency”
Thematic accuracy	Classification correctness	Number of incorrectly classified features
		Misclassification rate
		Misclassification matrix
		Relative misclassification matrix
		Kappa coefficient
	Non-quantative attribute correctness	Number of incorrect attribute values
		Rate of correct attribute values
		Rate of incorrect attribute values
	Quantative attribute accuracy	Attribute value uncertainty at 68,3% level
		Attribute value uncertainty at 50% level
		Attribute value uncertainty at 90% level
		Attribute value uncertainty at 95% level
		Attribute value uncertainty at 99% level
		Attribute value uncertainty at 99,8% level

Aggregation measures		Data product specification passed
		Data product specification fail count
		Data product specification pass count
		Data product specification fail rate
		Data product specification pass rate

Conclusion: S-100 has 15 quality measures defined. In ISO 19157 dictionary, 83 measures are listed. So 18% of possible measures are used. Are the remaining measures used in different context than S-100? This may need to be reconsidered.

INSPIRE standards have directly copied ISO standards and their definitions into their schema.

The DQWG is invited to:

1. Note this report
2. Discuss its content
3. Draw a conclusion to the S-100 WG.