DRAFT DATA QUALITY CHECKLIST
drafted: 17 January 2018, Monaco for DQWG-13
Author: R. Broekman (DQWG-Chair)

This list is derived from ISO 19157. Definitions and descriptions are a direct copy of ISO-19157.
There are 20 different data quality measures defined that can be used for validation of S-1xx Product Specifications (PS).
There is a recommendation for the target result of Positional Accuracy for depth contour lines and gridded bathymetry.

Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Applicable to spatial representation types
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 S-1xx
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 S-1xx
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 S-1xx
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	PS with geometric surfaces.
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 S-1xx

Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Applicable to spatial representation types
Logical Consistency / Format Consistency	Degree to which data is stored in accordance with the physical structure of the data set, as described by the scope	physicalStructureConflictsNumber / This data quality measure is a count of all items in the dataset that are stored in conflict with the physical structure of the dataset.	dataset/dataset series	All S-1xx
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	PS with curves.
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	PS with curves.
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	PS with curves.
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	PS with geometric surfaces.
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	PS with curves / geometric surfaces.

Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Applicable to spatial representation types
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	PS with curves / geometric surfaces.
Positional Accuracy / Absolute or External Accuracy	Closeness of reported coordinative values to values accepted as or being true.	Root Mean Square Error / Standard deviation, where the true value is not estimated from the observations but known a priori.	spatial object / spatial object type	PS with objects that have coordinative values associated.
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	PS with objects that have a vertical coordinative values associated.
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	PS with objects that have a horizontal coordinative values associated.
Positional Accuracy / Gridded Data Position Accuracy	Closeness of reported coordinative values to values accepted as or being true.	Root mean square error of planimetry / Radius of a circle around the given point, in which the true value lies with probability P.	spatial object / spatial object type	PS with objects that have a gridded coordinative values associated.
Temporal Quality / Temporal Consistency	Consistency with time.	Correctness of ordered events or sequences, if reported.	dataset/dataset series/ spatial object type	PS with objects that have a time value associated.

Data quality element and sub element	Definition	DQ measure / description	Evaluation scope	Applicable to spatial representation types
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 S-1xx PS.
Aggregation Measures / AggregationMeasures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationPassed / This data quality measure is a boolean indicating that all requirements in the referred data product specification are fulfilled.	dataset/dataset series/ spatial object type	PS that a require a complete pass of all elements of a dataset/ dataset series/spatial object types
Aggregation Measures / AggregationMeasures	In a data product specification, several requirements are set up for a product to conform to the specification.	DataProductSpecificationFailRate / This data quality measure is a number indicating the number of data product specification requirements that are not fulfilled by the current product/dataset in relation to the total number of data product specification requirements.	dataset/dataset series/ spatial object type	PS that a require a complete pass of all elements of a dataset/ dataset series/spatial object types

Recommendations for Positional Accuracy / Absolute or External Accuracy:
Maximum RMSE (horizontal) = E / 10000
Maximum RMSE (vertical) = Vint / 6

Recommendation for Positional Accuracy / Gridded Data Position Accuracy:
Maximum RMSE (horizontal) = GSD / 6
Maximum RMSE (vertical) = GSD / 3

Where:
E = Denominator of intended scale of mapping
Vint = normal contour line interval
GSD = Ground Sampling Distance