

7 Data quality

This chapter includes a description of the data quality elements and sub-elements as well as the corresponding data quality measures that should be used to evaluate and document data quality for data sets related to the spatial data theme *Protected Sites* (section 7.1).

It may also define requirements or recommendations about the targeted data quality results applicable for data sets related to the spatial data theme *Protected Sites* (sections 7.2 and 7.3).

In particular, the data quality elements, sub-elements and measures specified in section 7.1 should be used for

- evaluating and documenting data quality properties and constraints of spatial objects, where such properties or constraints are defined as part of the application schema(s) (see section 5);
- evaluating and documenting data quality metadata elements of spatial data sets (see section 8); and/or
- specifying requirements or recommendations about the targeted data quality results applicable for data sets related to the spatial data theme *Protected Sites* (see sections 7.2 and 7.3).

The descriptions of the elements and measures are based on Annex D of ISO/DIS 19157 Geographic information – Data quality.

7.1 Data quality elements

Table 4 lists all data quality elements and sub-elements that are being used in this specification. Data quality information can be evaluated at level of spatial object, spatial object type, dataset or dataset series. The level at which the evaluation is performed is given in the “Evaluation Scope” column.

The measures to be used for each of the listed data quality sub-elements are defined in the following sub-sections.

Table 4 – Data quality elements used in the spatial data theme *Protected Sites*

Section	Data quality element	Data quality sub-element	Definition	Evaluation Scope
7.1.1	Completeness	Commission	excess data present in the dataset, as described by the scope	dataset series; dataset; spatial object type
7.1.2	Completeness	Omission	data absent from the dataset, as described by the scope	dataset series; dataset; spatial object type
7.1.3	Logical consistency	Conceptual consistency	adherence to rules of the conceptual schema	spatial object type; spatial object
7.1.4	Logical consistency	Domain consistency	adherence of values to the value domains	spatial object type; spatial object
7.1.5	Positional accuracy	Absolute or external accuracy	closeness of reported coordinate values to values accepted as or being true	dataset series; dataset; spatial object type

Recommendation 13 Where it is impossible to express the evaluation of a data quality element in a quantitative way, the evaluation of the element should be expressed with a textual statement as a data quality descriptive result.

7.1.1 Completeness – Commission

Recommendation 14 Commission should be evaluated and documented using *Rate of excess items* as specified in the table below.

Name	Rate of excess items
Alternative name	–
Data quality element	Completeness
Data quality sub-element	Commission
Data quality basic measure	Error rate
Definition	Number of excess items in the dataset in relation to the number of items that should have been present.
Description	–
Evaluation scope	spatial object type, data set, data set series
Reporting scope	data set
Parameter	–
Data quality value type	Real, percentage, ratio (example: 0,0189 ; 98,11% ; 11:582)
Data quality value structure	–
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	–
Measure identifier	3

7.1.2 Completeness – Omission

Recommendation 15 Omission should be evaluated and documented using *Rate of missing items* as specified in the tables below.

Name	Rate of missing items
Alternative name	–
Data quality element	Completeness
Data quality sub-element	Omission
Data quality basic measure	Error rate
Definition	Number of missing items in the dataset in relation to the number of items that should have been present.
Description	–
Evaluation scope	spatial object type, data set, data set series
Reporting scope	data set
Parameter	–
Data quality value type	Real, percentage, ratio (example: 0,0189 ; 98,11% ; 11:582)
Data quality value structure	–
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	–
Measure identifier	7

7.1.3 Logical consistency – Conceptual consistency

The Application Schema conformance class of the Abstract Test Suite in Annex I defines a number of tests to evaluate the conceptual consistency (tests A.1.1-A.1.9) of a data set.

Recommendation 16 For the tests on conceptual consistency, it is recommended to use the *Logical consistency – Conceptual consistency* data quality sub-element and the measure *Number of items not compliant with the rules of the conceptual schema* as specified in the table below.

Name	
Alternative name	-
Data quality element	logical consistency
Data quality sub-element	conceptual consistency
Data quality basic measure	error count
Definition	count of all items in the dataset that are not compliant with the rules of the conceptual schema
Description	If the conceptual schema explicitly or implicitly describes rules, these rules shall be followed. Violations against such rules can be, for example, invalid placement of features within a defined tolerance, duplication of features and invalid overlap of features.
Evaluation scope	spatial object / spatial object type
Reporting scope	data set
Parameter	-
Data quality value type	integer
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	
Measure identifier	10

7.1.4 Logical consistency – Domain consistency

The Application Schema conformance class of the Abstract Test Suite in Annex I defines a number of tests to evaluate the domain consistency (tests A1.10-A.1.12) of a data set.

Recommendation 17 For the tests on domain consistency, it is recommended to use the *Logical consistency – Domain consistency* data quality sub-element and the measure *Number of items not in conformance with their value domain* as specified in the table below.

Name	Number of items not in conformance with their value domain
Alternative name	-
Data quality element	logical consistency
Data quality sub-element	domain consistency
Data quality basic measure	error count
Definition	count of all items in the dataset that are not in conformance with their value domain
Description	
Evaluation scope	spatial object / spatial object type
Reporting scope	data set
Parameter	-
Data quality value type	integer

7.1.5 Positional accuracy – Absolute or external accuracy

Recommendation 18 Absolute or external accuracy should be evaluated and documented using *Root mean square error* as specified in the tables below.

Name	Root mean square error
Alternative name	RMSE
Data quality element	Positional accuracy
Data quality sub-element	Absolute or external accuracy
Data quality basic measure	not applicable
Definition	Standard deviation, where the true value is not estimated from the observations but known a priori An indication of the accuracy of the data set as an aggregate.
Description	The true value of an observable Z is known as z_t . From this, the estimator $\sigma_z = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{mi} - z_t)^2}$ yields to the linear root mean square error $RMSE = \alpha_z$. This measure is an indication of the accuracy of the features within the data set or feature type, and should reflect the feature within the data set or feature type that is least accurate.
Evaluation scope	spatial object type, data set, data set series
Reporting scope	data set
Parameter	
Data quality value type	Measure
Data quality value structure	
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	
Measure identifier	39

7.2 Minimum data quality requirements

No minimum data quality requirements are defined for the spatial data theme Protected Sites.

7.3 Recommendation on data quality

No minimum data quality recommendations are defined.