

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 *Land Use* (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 *Land Use* (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 *Land Use* (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 6 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 6 – Data quality elements used in the spatial data theme *Land Use*

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
7.1.2	Completeness	Omission	data absent from the dataset, as described by the scope	dataset
7.1.3	Logical consistency	Domain consistency	adherence of values to the value domains	dataset
7.1.4	Positional accuracy	Absolute or external accuracy	closeness of reported coordinate values to values accepted as or being true	dataset
7.1.5	Thematic accuracy	Classification correctness	comparison of the classes assigned to features or their attributes to a universe of discourse	dataset series; dataset; spatial object type; spatial object
7.1.6	Thematic accuracy	Non-quantitative attribute correctness	correctness of non-quantitative attributes	dataset series; dataset; spatial object type; spatial object

Recommendation 10 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 11 Commission should be evaluated and documented using rate of excess as specified in the table below.

Name	Rate of excess
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	This data quality element is not applicable to SLU or GLU. For ELU and PLU, the land use features (ExistingLandUseObject and ZoningElement) are mutually exclusive, therefore possible existence of polygons' excess can be managed by topological principles. This data quality element applies only for SupplementaryRegulation in PLU
Evaluation scope	Data set
Reporting scope	Data set
Parameter	-
Data quality value type	Real
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Value 1,89 (measured or estimated value) means that 1,89 % of all features in SupplementaryRegulation element are excess items.
Measure identifier	3 (ISO 19157)

7.1.2 Completeness – Omission

Recommendation 12 Omission should be evaluated and documented using rate of missing items as specified in the table 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	All land use features in schemas (ELU, PLU, SLU, GLU) should have at least one land use type defined as HILUCS value. Rate of missing items shows the known or estimated rate of the missing HILUCS values in relation to all land use features in a schema.
Evaluation scope	Data set
Reporting scope	Data set
Parameter	-
Data quality value type	Real
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality

Example	<p>Rate of missing items is set to 1.89 for ELU meaning that 1.89 % of all polygons are missing the HILUCS value.</p> <p>Rate of missing items for GLU dataset is set to 5 meaning 5 % of pixels are classified as “nodata”.</p> <p>Rate of missing items values 0 % are not reported (data is complete).</p>
Measure identifier	5 (ISO 19157)

Name	Missing item
Alternative name	-
Data quality element	Completeness
Data quality sub-element	Omission
Data quality basic measure	Error indicator
Definition	Indicator that shows that a specific item is missing in the data.
Description	<p>Indicator shall have value “true” when a certain land use type (HILUCS class) is missing in the dataset (data not collected) but that specific land use type exists (in real world) in the area the dataset covers and the data producer knows / recognizes that existence.</p> <p>Indicator shall have value “false” when a certain land use type is missing in the dataset and that land use type does not exist in the real world either.</p>
Evaluation scope	Data set
Reporting scope	Data set
Parameter	-
Data quality value type	Boolean (true indicates that an item is missing, false indicates that an item is missing in data and also in the real world)
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	<p>HILUCS class 3 “Services” does not exist in the dataset. When the data producer knows that this class occurs in the area the dataset covers he sets the missing item value for this class “true”. That indication covers also the sub-classes (3.1, 3.1.1. etc).</p> <p>HILUCS class 1.2 “Forestry use” data exists in the dataset but the sub-classes 1.2.1, 1.2.2 and 1.2.3 do not. Data producer knows that classes 1.2.2 and 1.2.3 exist in the area but 1.2.1 does not. Missing item will have value “true” for HILUCS classes 1.2.2 and 1.2.3 and a value “false” to 1.2.1.</p> <p>HILUCS class 1.2 “Forestry use” data exists in the dataset but sub-classes 1.2.1, 1.2.2 and 1.2.3 do not. Data producer does not know whether there exist sub-classes 1.2.1 - 1.2.3 in the area or not. Values for missing items are not set.</p>
Measure identifier	7 (ISO 19157)

7.1.3 Logical consistency – Domain consistency

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

Recommendation 13 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

7.1.4 Positional accuracy – Absolute or external accuracy

Recommendation 14 Absolute or external accuracy should be evaluated and documented using mean value of positional uncertainties (2D) as specified in the table below.

Name	Mean value of positional uncertainties (2D)										
Alternative name	-										
Data quality element	Positional accuracy										
Data quality sub-element	Absolute or external accuracy										
Data quality basic measure	Not applicable										
Definition	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.										
Description	<p>Element is valid only for land use schemas ELU and PLU. Measurement unit is meters.</p> <p>For ELU a estimation of positional error in sampled points can be evaluated:</p> <p>For a number of points (N), the measured positions are given as x_{mi}, y_{mi} and z_{mi} coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, x_{ti}, y_{ti} and z_{ti}, are considered to represent the true positions. The errors are calculated as</p> $e_i = \sqrt{(x_{mi} - x_{ti})^2 + (y_{mi} - y_{ti})^2}$ <p>The mean positional uncertainties of the horizontal absolute or external positions is then calculated as</p> $\bar{e} = \frac{1}{N} \sum_{i=1}^N e_i$ <p>A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines, etc.). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result.</p> <p>NOTE: This data quality measure is different than the standard deviation.</p> <p>For PLU (polygons) the positional accuracy is portrayed as an estimated value based to the scale of the data as follows:</p> <table border="1"> <thead> <tr> <th>Scale, 1:n</th> <th>Estimated geometric accuracy; normal data quality</th> </tr> </thead> <tbody> <tr> <td>1 000 000</td> <td>500 m</td> </tr> <tr> <td>500 000</td> <td>250 m</td> </tr> <tr> <td>250 000</td> <td>125 m</td> </tr> <tr> <td>100 000</td> <td>50 m</td> </tr> </tbody> </table>	Scale, 1:n	Estimated geometric accuracy; normal data quality	1 000 000	500 m	500 000	250 m	250 000	125 m	100 000	50 m
Scale, 1:n	Estimated geometric accuracy; normal data quality										
1 000 000	500 m										
500 000	250 m										
250 000	125 m										
100 000	50 m										

	50 000	25 m
	20 000	10 m
	10 000	5 m
	5 000	2,5 m
	2 000	1 m
	1 000	0,5 m
Evaluation scope	Data set	
Reporting scope	Data set	
Parameter	-	
Data quality value type	Measure or real	
Data quality value structure	-	
Source reference	ISO/DIS 19157 Geographic information – Data quality	
Example	<p>Value 30.5 for ELU dataset means that the mean positional uncertainty of a land use type (position of the outer limit) is ± 30.5 meters compared to the true position.</p> <p>Value 50 for PLU dataset means that dataset's estimated positional accuracy corresponds to the 1:100 000 scale.</p>	
Measure identifier	28 (ISO 19157)	

7.1.5 Thematic accuracy – Classification correctness

Recommendation 15 Classification correctness should be evaluated and documented using misclassification rate as specified in the table below.

Name	Misclassification rate
Alternative name	-
Data quality element	Thematic accuracy
Data quality sub-element	Classification correctness
Data quality basic measure	Error rate
Definition	Number of incorrectly classified features relative to the number of features that should be there.
Description	<p><i>Land Use</i> type classes in dataset in relation to the land use types in the real world.</p> <p>This element is used only for existing land use: ELU, SLU, GLU.</p>
Evaluation scope	Data set
Reporting scope	Data set
Parameter	-
Data quality value type	Real
Data quality value structure	
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Value 1.89 indicates that 1.89 % of all features in a dataset (polygons, points or pixels) have different land use type than the corresponding feature in a real world.
Measure identifier	61 (ISO 19157)

7.1.6 Thematic accuracy – Non-quantitative attribute correctness

Recommendation 16 Non-quantitative attribute correctness should be evaluated and documented using rate of incorrect attribute values as specified in the table below.

Name	Rate of incorrect attribute values
Alternative name	-
Data quality element	Thematic accuracy
Data quality sub-element	Non-quantitative attribute correctness
Data quality basic measure	Error rate

Definition	Number of attribute values where incorrect values are assigned in relation to the total number of attribute values.
Description	This element is used only for planned land use: PLU (Zoning Element and SupplementaryRegulation) features and is used to non-quantitative attributes like Inspire-id, RegulationNature, ProcessStepGeneral and SupplementaryRegulationValue.
Evaluation scope	Data set
Reporting scope	Data set
Parameter	-
Data quality value type	Real ; percentage
Data quality value structure	
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Value 1.89 means that 1.89 % of all non-quantitative attributes have incorrect value.
Measure identifier	67 (ISO 19157)

7.2 Minimum data quality requirements

No minimum data quality requirements are defined for the spatial data theme *Land Use*.

7.3 Recommendation on data quality

Recommendation 17 For the data quality elements listed in Table 7, all data sets related to the spatial data theme *Land Use* should meet the specified target results.

Table 7 – Recommended minimum data quality results for spatial data theme *Land Use*

Section	Data quality element and sub-element	Measure name(s)	Target result(s)	Condition
7.1.1	Completeness - Commission	Error rate	0 %	
7.1.2	Completeness - Omission	Rate of missing items	0 %	
7.1.3	Logical consistency - Domain consistency	Correct items rate	Min 95 %	
7.1.4	Positional accuracy - Absolute or external accuracy	Estimation of positional error	Corresponding normal data quality (table above)	
7.1.5	Thematic accuracy - classification correctness	Error rate	Max 5 %	