

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 Cover* (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 Cover* (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 Cover* (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 9 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 9 – Data quality elements used in the spatial data theme *Land Cover*

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	spatial object type
7.1.2	Completeness	Omission	data absent from the dataset, as described by the scope	spatial object type
7.1.3	Logical consistency	Conceptual consistency	adherence to rules of the conceptual schema	dataset series; dataset; spatial object type; spatial object
7.1.4	Logical consistency	Domain consistency	adherence of values to the value domains	spatial object
7.1.5	Logical consistency	Format consistency	degree to which data is stored in accordance with the physical structure of the dataset, as described by the scope	dataset; spatial object type;
7.1.6	Logical consistency	Topological consistency	correctness of the explicitly encoded topological characteristics of the dataset, as described by the scope	spatial object
7.1.7	Positional accuracy	Absolute or external accuracy	closeness of reported coordinate values to values accepted as or being true	spatial object

7.1.8	Positional accuracy	Relative or internal accuracy	closeness of the relative positions of features in the scope to their respective relative positions accepted as or being true	spatial object
7.1.9	Thematic accuracy	Classification correctness	comparison of the classes assigned to features or their attributes to a universe of discourse	spatial object
7.1.10	Thematic accuracy	Non-quantitative attribute correctness	correctness of non-quantitative attributes	spatial object
7.1.11	Thematic accuracy	Quantitative attribute accuracy	accuracy of quantitative attributes	spatial object

Recommendation 11 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 12 Commission should be evaluated and documented using **excess item** as specified in the tables below.

Name	excess item
Alternative name	-
Data quality element	completeness
Data quality sub-element	commission
Data quality basic measure	error indicator
Definition	indication that an item is incorrectly present in the data
Description	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set: LandCoverDataSet
Parameter	-
Data quality value type	Boolean (true indicates that the item is in excess)
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Presence of excess items in a dataset: - Two or more points collected on top of each other - Two or more surfaces collected on top of each other
Measure identifier	1

7.1.2 Completeness – Omission

Recommendation 13 Omission should be evaluated and documented using **missing item** as specified in the tables below.

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	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set: LandCoverDataSet

Parameter	-
Data quality value type	Boolean (true indicates that an item is missing)
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Missing item found: - Uncontrolled gap in the dataset
Measure identifier	5

7.1.3 Logical consistency – Conceptual consistency

Recommendation 14 Conceptual consistency should be evaluated and documented using **conceptual schema non-compliance** as specified in the tables below.

Name	conceptual schema non-compliance
Alternative name	-
Data quality element	logical consistency
Data quality sub-element	conceptual consistency
Data quality basic measure	error indicator
Definition	indication that an item is not compliant to the rules of the relevant conceptual schema
Description	-
Evaluation scope	spatial object: LandCoverUnit spatial object type: LandCoverUnit data set: LandCoverDataSet, LandCoverGridCoverage
Reporting scope	data set: LandCoverDataSet, LandCoverGridCoverage
Parameter	-
Data quality value type	Boolean (true indicates that an item is not compliant with the rules of the conceptual schema)
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Presence of items in the coverage violating conceptual consistency: - The coverage contains line / point features instead of polygons - Overlapping polygons are found in the coverage, while the concept does not allow overlaps - Polygons smaller, than the pre-defined Minimum Mapping Unit are present in the coverage
Measure identifier	8

7.1.4 Logical consistency – Domain consistency

Recommendation 15 Domain consistency should be evaluated and documented using **value domain non-conformance** as specified in the tables below.

Name	value domain non-conformance
Alternative name	-
Data quality element	logical consistency
Data quality sub-element	domain consistency
Data quality basic measure	error indicator
Definition	indication of if an item is not in conformance with its value domain
Description	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set : LandCoverDataSet, LandCoverGridCoverage
Parameter	-
Data quality value type	Boolean (true indicates that an item is not in conformance with its value domain)
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Presence of items in the coverage violating domain consistency:

	- The coverage contains points / polygons with attributes out of attribute domain (e.g. non existing class codes)
Measure identifier	14

7.1.5 Logical Consistency – Format consistency

Recommendation 16 Format consistency should be evaluated and documented using **physical structure conflicts** as specified in the tables below.

Name	physical structure conflicts
Alternative name	-
Data quality element	logical consistency
Data quality sub-element	format consistency
Data quality basic measure	error indicator
Definition	indication that items are stored in conflict with the physical structure of the dataset
Description	-
Evaluation scope	data set : LandCoverDataSet, LandCoverGridCoverage
Reporting scope	data set : LandCoverDataSet, LandCoverGridCoverage
Parameter	-
Data quality value type	Boolean (true indicates physical structure conflict)
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Presence of items in the coverage violating format consistency: - File format, file or attribute names or attribute types do not correspond to the specifications
Measure identifier	119

7.1.6 Logical Consistency – Topological consistency

Recommendation 17 Topological consistency should be evaluated and documented using **number of invalid self-intersect errors** as specified in the tables below.

Name	number of invalid self-intersect errors
Alternative name	loops
Data quality element	logical consistency
Data quality sub-element	topological consistency
Data quality basic measure	error count
Definition	count of all items in the data that illegally intersect with themselves
Description	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set : LandCoverDataSet
Parameter	-
Data quality value type	Integer
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Number of loops („figure eight” forming land cover polygons) present
Measure identifier	26

7.1.7 Data Quality – Positional accuracy – Absolute or external accuracy

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

Name	root mean square error of planimetry
Alternative name	RMSEP

Data quality element	positional accuracy
Data quality sub-element	absolute or external accuracy
Data quality basic measure	-
Definition	radius of a circle around the given point, in which the true value lies with probability P
Description	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set : LandCoverDataSet
Parameter	-
Data quality value type	Real
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Absolute or external positional accuracy of land cover data is usually determined by the absolute positional accuracy of the Earth Observation imagery which serves as basis for derivation of the LC data.
Measure identifier	47

7.1.8 Data Quality – Positional accuracy – Relative or internal accuracy

Recommendation 19 Relative or internal accuracy should be evaluated and documented using **root mean square error of planimetry** as specified in the tables below.

Name	root mean square error of planimetry
Alternative name	RMSEP
Data quality element	positional accuracy
Data quality sub-element	relative or internal accuracy
Data quality basic measure	-
Definition	radius of a circle around the given point, in which the true value lies with probability P
Description	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set : LandCoverDataSet
Parameter	-
Data quality value type	Real
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Relative or internal accuracy of land cover data is usually determined as the accuracy of delineation of land cover unit boundaries relative to the underlying Earth Observation imagery which serves as basis for derivation of the LC data.
Measure identifier	47

7.1.9 Data Quality – Thematic accuracy – Classification correctness

Recommendation 20 Classification correctness should be evaluated and documented using **misclassification rate or misclassification matrix** as specified in the tables below.

Name	misclassification rate
Alternative name	-
Data quality element	thematic accuracy
Data quality sub-element	classification correctness
Data quality basic measure	error rate
Definition	incorrectly classified area relative to the true area of the target land cover class
Description	-
Evaluation scope	spatial object: LandCoverUnit

Reporting scope	data set : LandCoverDataSet
Parameter	-
Data quality value type	Real
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Misclassification rate is usually estimated by sampling and reported relative to a land cover class area. Quantitative accuracy parameters are to be published together with the estimation of the uncertainty on a pre-defined significance level: - class commission error rate is 12,1% ± 2,8% - class omission error rate is 16,4% ± 3,5% Significance level: 68,3%
Measure identifier	61

Name	misclassification matrix																																																											
Alternative name	confusion matrix																																																											
Data quality element	thematic accuracy																																																											
Data quality sub-element	classification correctness																																																											
Data quality basic measure	-																																																											
Definition	matrix that indicates the area class (i) classified as class (j). In the practice these areas are estimated by the number of random samples falling on class(i) and class (j)																																																											
Description	The misclassification matrix (MCM) is a quadratic matrix with n columns and n rows. n denotes the number of classes under consideration. MCM (i,j) = [# items of class (i) classified as class (j)] The diagonal elements of the misclassification matrix contain the correctly classified items, and the off diagonal elements contain the number of misclassification errors.																																																											
Evaluation scope	spatial object: LandCoverUnit																																																											
Reporting scope	data set : LandCoverDataSet																																																											
Parameter	Name: n Definition: number of classes under consideration Value Type: Integer																																																											
Data quality value type	Integer																																																											
Data quality value structure	Matrix (n × n)																																																											
Source reference	ISO/DIS 19157 Geographic information – Data quality; Congalton, R. (1991) A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. Remote Sensing of Environment, 37:35-46.																																																											
Example	<div>The assignment of an item to a certain class can either be correct or incorrect. As appropriate reference land cover data are rarely available, land cover classification is usually compared at sample locations with reference interpretation based on EO imagery or a field check.</div> <table><tr><td colspan="2"></td><td colspan="4">Reference data</td><td></td></tr><tr><td rowspan="5">Classified data</td><td>Forest (F)</td><td>F</td><td>I</td><td>U</td><td>W</td><td>Sum</td><td>User's accuracy</td></tr><tr><td>Industrial (I)</td><td>68</td><td>7</td><td>3</td><td>0</td><td>78</td><td>87.2%</td></tr><tr><td>Urban (U)</td><td>12</td><td>112</td><td>15</td><td>10</td><td>149</td><td>75.2%</td></tr><tr><td>Water (W)</td><td>3</td><td>9</td><td>89</td><td>0</td><td>101</td><td>88.1%</td></tr><tr><td>Sum</td><td>0</td><td>2</td><td>5</td><td>56</td><td>63</td><td>88.9%</td></tr><tr><td colspan="2">Producer's accuracy</td><td>83</td><td>130</td><td>112</td><td>66</td><td>391</td><td></td></tr><tr><td colspan="2"></td><td>81.9%</td><td>86.2%</td><td>79.5%</td><td>84.8%</td><td></td><td></td></tr></table> <div>Overall accuracy: 84%</div>			Reference data					Classified data	Forest (F)	F	I	U	W	Sum	User's accuracy	Industrial (I)	68	7	3	0	78	87.2%	Urban (U)	12	112	15	10	149	75.2%	Water (W)	3	9	89	0	101	88.1%	Sum	0	2	5	56	63	88.9%	Producer's accuracy		83	130	112	66	391				81.9%	86.2%	79.5%	84.8%		
		Reference data																																																										
Classified data	Forest (F)	F	I	U	W	Sum	User's accuracy																																																					
	Industrial (I)	68	7	3	0	78	87.2%																																																					
	Urban (U)	12	112	15	10	149	75.2%																																																					
	Water (W)	3	9	89	0	101	88.1%																																																					
	Sum	0	2	5	56	63	88.9%																																																					
Producer's accuracy		83	130	112	66	391																																																						
		81.9%	86.2%	79.5%	84.8%																																																							
Measure identifier	62																																																											

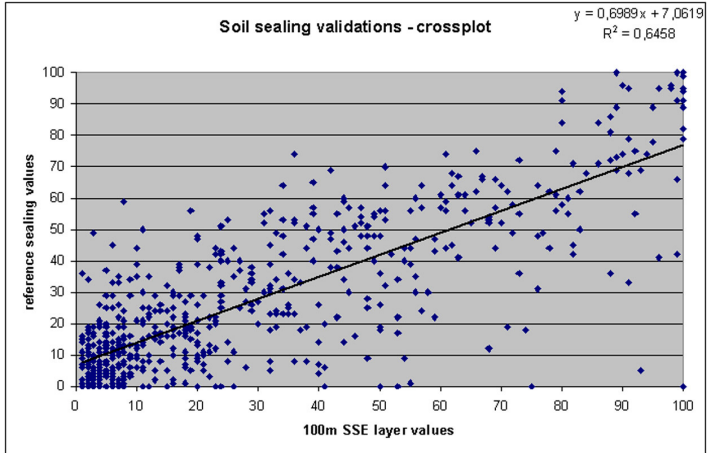
7.1.10 Data Quality – Thematic accuracy – Non-quantitative attribute accuracy

Recommendation 21 Non-quantitative attribute accuracy should be evaluated and documented using **rate of incorrect attribute values** as specified in the tables below.

Name	rate of incorrect attribute values
Alternative name	-
Data quality element	thematic accuracy
Data quality sub-element	non-quantitative attribute accuracy
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	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set : LandCoverDataSet
Parameter	-
Data quality value type	Real
Data quality value structure	-
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	
Measure identifier	67

7.1.11 Data Quality – Thematic accuracy – Quantitative attribute accuracy

Recommendation 22 Quantitative attribute accuracy should be evaluated and documented using a **scatter-plot** as specified in the tables below.

Name	scatter plot
Alternative name	-
Data quality element	thematic accuracy
Data quality sub-element	quantitative attribute accuracy
Data quality basic measure	-
Definition	graphical comparison of reference land cover density values with density values provided by the product
Description	-
Evaluation scope	spatial object: LandCoverUnit
Reporting scope	data set : LandCoverDataSet
Parameter	-
Data quality value type	graphical
Data quality value structure	-
Source reference	http://sia.eionet.europa.eu/EAGLE/EAGLE_5rdMeeting_g2_MONINA_FFM/EL_Validation_of_HR_layers_finaldraft.pdf
Example	 <p>The scatter plot displays a positive correlation between reference sealing values and 100m SSE layer values. The y-axis is labeled 'reference sealing values' and ranges from 0 to 100. The x-axis is labeled '100m SSE layer values' and also ranges from 0 to 100. A regression line is drawn through the data points, with the equation $y = 0,6989x + 7,0619$ and $R^2 = 0,6458$ displayed in the top right corner. The plot is titled 'Soil sealing validations - crossplot'.</p>
Measure identifier	-

7.2 Minimum data quality requirements

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

7.3 Recommendation on data quality

Recommendation 23 For the data quality elements listed in the table below, all data sets related to the spatial data theme *Land Cover* should meet the specified target results.

Table 10 – Recommended minimum data quality results for spatial data theme Land Cover

Section	Data quality element and sub-element	Measure name(s)	Target result(s)	Condition
7.1.1	Completeness - Commission	excess item	false	Mandatory, if the recognition of commission errors requires only automatic procedures (e.g. filtering out overlapping polygons or duplicate points for an area frame sampling).
7.1.2	Completeness - Omission	missing item	false	Mandatory, if the recognition of omission errors requires only automatic procedures (e.g. finding uncontrolled gaps).
7.1.3	Logical consistency - Conceptual consistency	conceptual schema non-compliance	False (zero violations in dataset)	Mandatory
7.1.4	Logical consistency – Domain consistency	value domain non-conformance	False (zero violations in dataset)	Mandatory -
7.1.5	Logical consistency – Format consistency	physical structure conflicts	False (zero violations in dataset)	Mandatory
7.1.6	Logical consistency - Topological consistency	number of invalid self-intersect errors	0 (zero violations in dataset)	Mandatory, if the dataset is a polygon coverage
7.1.7	Positional accuracy - Absolute or external accuracy	RMSE	Not more than the pixel size of the Earth Observation imagery, which serves as basis for derivation of the LC data.	

7.1.9	Thematic accuracy – Classification correctness	misclassification rate / misclassification matrix	Max 15% misclassification rate is a widely used criteria, but can not used as a general target, because the misclassification rate strongly depends on the level of details (number of classes, geometric resolution). The minimum recommendation is to measure classification correctness and tell the results in a form of misclassification rate and/or misclassification matrix.	
-------	---	--	---	--