

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 *Agricultural and Aquaculture Facilities* (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 *Agricultural and Aquaculture Facilities* (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 *Agricultural and Aquaculture Facilities* (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 3 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 3 – Data quality elements used in the spatial data theme *Agricultural and Aquaculture Facilities*

Section	Data quality element	Data quality sub-element	Definition	Evaluation Scope
7.1.1	Completeness	Omission	data absent from the dataset, as described by the scope	dataset
7.1.2	Logical consistency	Conceptual consistency	adherence to rules of the conceptual schema	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	spatial object type
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	Temporal quality	Temporal validity	validity of data specified by the scope with respect to time	dataset series; dataset; spatial object type; spatial object

Recommendation 7 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 – Omission

Recommendation 8 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	Information contained on provided datasets could only show a constrained set of elements because of different causes as Geo-reference issues, Thematic Scope Constraints or Data Quality and Accuracy.
Evaluation scope	data set
Reporting scope	data set
Parameter	Calculation Method: Real, Estimated.
Data quality value type	<i>Value type for reporting a data quality result. A data quality value type shall be provided for a data quality result. Examples include Boolean, Real, Integer, Ratio (numerator of type integer : denominator of type integer), Percentage, Measure(s) (value(s) + unit(s)), Ratio</i>
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Legislation requirements establish limits (e.g Combustion Capacity, Number of Animals,...) to the Entities to be registered on Thematic Data Sets. Limits and conditions on the Accuracy requested by the legislation (e.g. REGULATION (EC) No 1166/2008)
Measure identifier	7

7.1.2 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.1.2 and A.1.4-A.1.7) of a data set.

Recommendation 9 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

7.1.3 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 (test A.1.3) of a data set.

Recommendation 10 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 11 Absolute or external accuracy should be evaluated and documented using **Mean value of positional uncertainties** as specified in the tables below.

Name	Mean value of positional uncertainties
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>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> <p>1D: $e_i = x_{mi} - x_{ti}$ 2D: $e_i = \sqrt{(x_{mi} - x_{ti})^2 + (y_{mi} - y_{ti})^2}$ 3D: $e_i = \sqrt{(x_{mi} - x_{ti})^2 + (y_{mi} - y_{ti})^2 + (z_{mi} - z_{ti})^2}$</p> <p>The mean positional uncertainties of the horizontal absolute or external positions are 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). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result. This data quality measure is different from the standard deviation.</p>
Evaluation scope	Spatial object : Activity Complex/Holding
Reporting scope	data set
Parameter	-
Data quality value type	Measure
Data quality value structure	
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	<p>REGULATION (EC) No 1166/2008) on farm structure surveys and the survey on agricultural production methods “... ANNEX III List of farm structure survey characteristics CHARACTERISTICS UNITS/CATEGORIES I. General characteristics – Location of the holding – – Latitude (within an arc of 5 minutes or less) Degrees: Minutes – – Longitude (within an arc of 5 minutes or less) Degrees: Minutes ”</p> <p>E-PRTR User Manual: http://www.eionet.europa.eu/schemas/eprtr/EPTRTRUserManual.pdf</p> <p>“ ... 3.1.4 Facility Report elements</p> <p>GeographicalCoordinate:</p> <p>The coordinates of the location of the facility should be expressed in longitude and latitude coordinates giving a precision of the order of at least □□500 meters and referring to the geographical centre of the site of the</p>

	facility. Required. ...”
Measure identifier	28

7.1.5 Thematic accuracy – Classification correctness

Recommendation 12 Classification correctness should be evaluated and documented using **Misclassification rate** 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	number of incorrectly classified features in relation to the number of features that are supposed to be there
Description	-
Evaluation scope	data set
Reporting scope	data set
Parameter	-
Data quality value type	Real, Percentage, Ratio
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	Similarity in terms and scopes can derive in incorrect categorizations (E.g A Combustion plant could be consider as "Installation" or Facility- Activity Complex depending on the capacity and size and its emplacement , isolated or as part of a wider Facility).
Measure identifier	61

7.1.6 Temporal quality – Temporal validity

Recommendation 13 Temporal validity should be evaluated and documented using **Value Domain Conformance Rate** as specified in the tables below.

Name	Value Domain Conformance Rate
Alternative name	-
Data quality element	Temporal quality
Data quality sub-element	Temporal validity
Data quality basic measure	Error rate
Definition	number of items in the dataset that are in conformance with their value domain in relation to the total number of items in the dataset
Description	-
Evaluation scope	spatial object type: Activity Complex
Reporting scope	data set
Parameter	-
Data quality value type	-
Data quality value structure	Single value
Source reference	-
Example	Changes in the legal and real world entities through the time line derive in different classifications and registries in different thematic Datasets. This can derive in duplications and updates of entities in datasets. Temporal validity and refresh of data is an important issue.

Measure identifier	17
--------------------	----

7.2 Minimum data quality requirements

No minimum data quality recommendations are defined.

7.3 Recommendation on data quality

No minimum data quality recommendations are defined.