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 *Buildings* (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 *Buildings* (sections 0 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 *Buildings* (see sections 0 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 *Buildings*

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	Positional accuracy	Absolute or external accuracy	closeness of reported coordinate values to values accepted as or being true	spatial object type
7.1.4	Usability	--	degree of adherence of a dataset to a specific set of requirements	dataset

Recommendation 1 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 2 Commission should be evaluated and documented using <error rate, from ISO/DIS 19157> as specified in the tables below.

NOTE: this recommendation applies to any data set related to theme *Buildings*, whatever profile is used. In case an extended profile is used, priority should be given to report omission related to feature types Building and BuildingPart.

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	Items that should have been present are defined in the data capture rules of the data producer that have to be documented, e.g. in the template for additional information (provided in annex D)
Evaluation scope	Data set
Reporting scope	Data set
Parameter	
Data quality value type	Real (e.g. percentage, ratio)
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	2% (The dataset has 2% building parts more than the ones necessary to model the universe of discourse)
Measure identifier	3 (ISO/DIS 19157:2012)

7.1.2 Completeness – Omission

Recommendation 3 Omission should be evaluated and documented using <error rate, from ISO/DIS 19157> as specified in the tables below.

NOTE: this recommendation applies to any data set related to theme *Buildings*, whatever profile is used. In case an extended profile is used, priority should be given to report omission related to feature types Building and BuildingPart.

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	Items that should have been present are defined in the data capture rules of the data producer that have to be documented, e.g. in the template for additional information (provided in annex E)
Evaluation scope	Data set
Reporting scope	Data set
Parameter	–
Data quality value type	Real (e.g. percentage, ratio)
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	5% (The dataset has 5% less buildings than the ones necessary to model the universe of discourse)
Measure identifier	7 (ISO/DIS 19157:2012)

7.1.3 Positional accuracy – Absolute or external accuracy

Recommendation 4 Absolute or external accuracy should be evaluated and documented using <Name of the measure(s), from ISO/DIS 19157> as specified in the tables below.

NOTE: this recommendation applies to any data set related to theme *Buildings*, whatever profile is used.

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	Not applicable
Definition	Radius of a circle around the given point, in which the true value lies with probability P
Description	The true values of the observed coordinates *X* and *Y* are known as x_t and y_t. From this the estimator $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(x_{mi}-x_t)^2+(y_{mi}-y_t)^2\right]}$ yields to the linear root mean square error of planimetry RMSEP = σ
Evaluation scope	data set;
Reporting scope	data set
Parameter	-
Data quality value type	Measure
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	-
Measure identifier	47 (ISO/DIS 19157:2012)

Recommendation 5 Absolute or external accuracy of the vertical component of feature types, should be evaluated and documented using *Root mean square error* as specified in the table below.

NOTE: this recommendation applies to only to data sets related to theme *Buildings* and having vertical component, i.e. to data sets with 3D or 2,5D data.

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*
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 = σ_z.
Evaluation scope	data set
Reporting scope	data set
Parameter	-
Data quality value type	Measure
Data quality value structure	Single value
Source reference	ISO/DIS 19157 Geographic information – Data quality
Example	-
Measure identifier	39 (ISO/DIS 19157:2012)

7.1.4 Usability

Recommendation 6 Usability should be evaluated and documented using < **Computation of population use case pass, Assessment of vulnerability to earthquake use case conformanc**e> as specified in the tables below.

NOTE: this recommendation applies to any data set related to theme *Buildings*, whatever profile is used.

Name	**< Computation of population use case pass, >**
Alternative name	
Data quality element	usability element
Data quality sub-element	
Data quality basic measure	correctness indicator
Definition	indication that all the requirements for computation of population are fulfilled
Description	R1: completeness for the buildings whose current use is residential, industrial or commerceAndServices must be better than 90% R2: at least, 95 % of the buildings of interest must be represented by a GM_Surface R3: the completeness and thematic accuracy on attribute numberOfFloorAboveGround must be better than 90% or the completeness and thematic accuracy on attribute heightAboveGround must be better than 90% R4: the completeness and thematic accuracy on attribute currentUse must be better than 90% R5: positional geometric accuracy on buildings must be better than 5 m.
Evaluation scope	data set
Reporting scope	data set
Parameter	
Data quality value type	Boolean (true if the requirements R1 to R5 are fulfilled)
Data quality value structure	Single value
Source reference	
Example	
Measure identifier	

Name	**<** Assessment of vulnerability to earthquake use case conformance**>**
Alternative name	
Data quality element	usability element
Data quality sub-element	
Data quality basic measure	correctness indicator
Definition	indication that all the requirements for assessing vulnerability to earthquake are fulfilled
Description	R1: completeness for the buildings having currentUse in real world should be better than 90% R2: at least, 95 % of the buildings must be represented by a GM_Surface R3: the completeness and thematic accuracy on attribute numberOfFloorAboveGround must be better than 90% or the completeness and thematic accuracy on attribute heightAboveGround must be better than 90% R4: the date of construction must be available with resolution equivalent or better than 5 years for the buildings constructed after 1970, 10 years for buildings constructed between 1950 and 1970 , 20 years between 1930 and 1950, for at least 90% of buildings R5: the completeness and thematic accuracy on attribute material of structure must be better than 90%
Evaluation scope	data set
Reporting scope	data set
Parameter	
Data quality value type	A (if requirements R1 to R5 are fulfilled) B (if requirements R1 to R4 are fulfilled)
Data quality value structure	single
Source reference	
Example	
Measure identifier	

7.2 Minimum data quality requirements

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

7.3 Recommendation on data quality

Data related to theme *Buildings* may be provided according several levels of detail. The level of detail has to be adapted to the use case(s).

The level of detail of a building data set is characterised both by the choice of the represented features and by the geometric representation of these features (mainly the positional accuracy). For instance, a data set with only buildings will not be used at the same range of scales as a data set with buildings and some of its constitutive elements (building part, building unit, walls, roofs, rooms …).

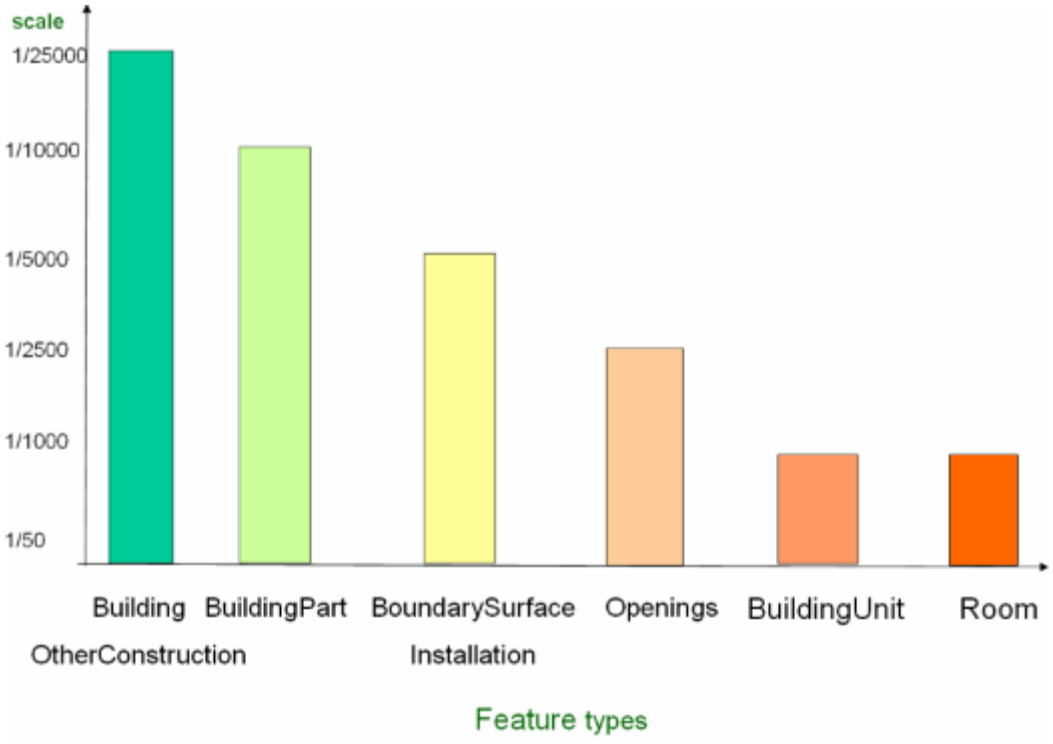

Figure 59: The scale range suitable for the feature types of theme Buildings

The level of detail of a building data set has to be consistent, i.e. the positional accuracy of the geometric representation has to be adapted to the scale range of the feature types the data set contains.

The following Table 7 gives the recommended minimum scale and accuracy for the various possible levels of detail of building data.

Table 7: Recommended minimum scale and accuracy for building data sets.

	LoD 0 (2D) **LoD1 (3D)**	**LoD 0 (2D)** **LoD1 (3D)**	**LoD 2**	
			LoD2 (2D)	LoD2 (3D)
Feature types	Building	Building BuildingPart	Building BuildingPart Installation	Building BuildingPart Installation BoundarySurface
Scale	1/25 000	1/10 000	1/5 000	1/5 000
Accuracy	5 m	2 m	1 m	1 m

	LoD 3	**LoD4**	
	LoD 3 (3D)	Lod4 (2D)	LoD4 (3D)
Feature types	Building BuildingPart Installation BoundarySurface Openings	Building BuildingPart Installation BuildingUnit	Building BuildingPart Installation BoundarySurface Opening BuildingUnit – Room InternalBuildingInstallation
Scale	1/2 500	1/1 000	1/1 000
Accuracy	0,5 m	0,2 m	0,2 m

Recommendation 1 A data set related to INSPIRE theme *Buildings* should have the minimum positional accuracy indicated in table n°2 above.

NOTE: the recommendation must be understood as follows:

- a data set containing only feature type Building should have an absolute positional accuracy equal or better than 5 m
- a data set containing only feature types Building and BuildingPart should have an absolute positional accuracy equal or better than 2 m
- a data set corresponding to LoD2 should have an accuracy better than 1 m. A data set is considered of LoD2 if it contains at least one feature type typical of LoD2 and no feature type of more detailed level (LoD3 or LoD4).