

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 *Atmospheric Conditions and Meteorological Geographical Features* (section 7.4).

It may also define requirements or recommendations about the targeted data quality results applicable for data sets related to the spatial data theme *Atmospheric Conditions and Meteorological Geographical Features* (sections 7.5 and 7.6).

In particular, the data quality elements, sub-elements and measures specified in section 7.4 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 *Atmospheric Conditions and Meteorological Geographical Features* (see sections 7.5 and 7.6).

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

7.1 WMO operational quality procedures

Almost all WMO data quality issues are process based and ultimately refer to documents specifying WMO regulations and other descriptive documents summarized in section 7.2. WMO regulations apply globally, and not just to WMO Member States who are also EC Member States.

Meteorological measurements compliant with WMO regulations go through operational procedures:

- a) To ensure the best possible quality of the data which are used in the real-time operations;
- b) In non-real time, to protect and improve the quality and integrity of data destined for storage and retrieval;
- c) To provide the basis for feedback of information on errors and questionable data to the source of the data.

Minimum standards for quality control of data apply to all WMO operational centres (cf. *Manual on the Global Data-processing and Forecasting System, WMO-No. 485*). They include quality control at various stages of processing. They apply to both real-time and non-real-time processing and lead to various records of quality-control actions. WMO also establishes standard operating and quality control procedures for atmospheric composition measurements (cf. WMO Global Atmosphere Watch (GAW) report series).

Checking includes:

- Detection of missing data at centres
- Adherence to prescribed coding formats
- Internal consistency
- Time consistency
- Space consistency
- Physical and Climatological limits

Records to be maintained include:

- Information to identify source of data such as station, aircraft, ship
- Type of deficiency (non-receipt, incomplete or incorrect reports, etc.)
- Identification of deficient element (whole report, specific parameter, etc.)
- Frequency of occurrence of data deficiencies (according to station type and element)

In non real time, checking includes in addition:

- Review of recorded data in comparison with observations
- Inter-comparison of parameters and calculations
- Check of supplementary data
- Check of extreme values

These quality reports are often held locally and not distributed with the data, which may be distributed worldwide. There is no requirement to collect and distribute this information – which would be new collections of data quality. Very few datasets relevant to this theme will hold or link to this quality data.

Similarly, numerical model output goes through thorough and systematic evaluation and quality assessment. Standard procedures have been developed for the production and exchange of verification results.

The question of the quality of meteorological data is closely related to its representivity. Depending on the way in which it is generated, the representivity of meteorological data can vary to a very large extent:

- in space:
 - local representativeness, ranging from a few m² to a few km² at most, over very homogeneous terrain
 - wider area representativeness, up to ~100 km² or more
- in time:
 - so-called instantaneous data (i.e. a few seconds)
 - average (or other statistical combinations) over periods of hours, days, months, etc.

Local data come from in-situ measurements and are available only for the locations of observing sites; area representative data come mainly from

- numerical models, available for all locations
- and remote-sensing (satellite based or not).

The WMO quality assurance process is very comparable to the ISO 19158 standard on quality control accreditation for data supply. For each Member as a supplier of observational and forecast data, WMO acts as the client encompassing all other Members. WMO defines the observing standards and the quality control processes at each stage of the data collection and dissemination process by which the data is distributed around the world.

As meteorological observations are transitory there is seldom any opportunity to perform repeat observations. Many of the Data Quality Classes of ISO 19157 are either not relevant, or rather the quality measures required are process based using non-quantitative standards and have only descriptive results referenced to WMO regulations and guides. Quantitative quality measures in WMO are usually post-facto tasks of monitoring and verification. The monitoring process studies the availability, timeliness and quality of data with respect to easily detectable errors. Verification involves matching observations with forecast values from numerical models in a cross validation exercise. Both are separate data gathering and product generation processes in slow time, which generate large amounts of new data.

7.2 WMO regulations on data quality

WMO regulations on data quality are included as sub-topics in a number of different WMO standards documents (there is no single WMO data quality reference document):

- The Guide on the Global Data-processing System (WMO-No. 305) Chapter 6 is the authoritative reference on all matters related to quality control procedures.

- Observational data are collected to quality standards declared in the WMO Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8 (Seventh edition 2008)).
- In the Manual on the GOS (WMO-No 544 Volume 1 Global Aspects Part V Quality Control), WMO recommends that rigorous quality control should be exercised at all stages, including periodic calibration, validation and maintenance of the equipment in order to maintain the quality of the observations.
- The Manual on the Global Data-processing and Forecasting System (Volume I Global Aspects WMO-No. 485) in Part II Section 2 defines the responsibilities and minimum standards of Quality Control at GDPFS Centres in real- and non-real-time.
- Guidelines on quality management procedures and practices for Public Weather Services (WMO/TD No. 1256), 2005 extends the quality aspects to the delivery of data and weather information outside the meteorological community.
- The series of Global Atmosphere Watch (GAW) Research and monitoring reports (e.g. Quality Assurance Project Plan (QAPjP) for Continuous Ground Based Ozone Measurements (WMO TD No. 634) define the quality control procedures for atmospheric composition measurements.

There is also European legislation covering air quality;

- Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe defines data quality objectives in Annex 1 (<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32008L0050:EN:NOT>)

7.3 Quality recommendation

For WMO and non-WMO derived data, the general principles are likely to hold:

- datasets will not be able to repeat measurements
- quality measures are descriptive, and qualitative, not quantitative.
- in case quality reports are available they are not collected and distributed on a regular basis.

Therefore, we recommend the following:

Recommendation 10 For WMO based data the quality metadata should refer to the qualitative process based WMO regulations for data quality

Recommendation 11 For non-WMO based data the quality metadata should refer to a statement of quality processes used by the producer.

7.4 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 *Atmospheric Conditions and Meteorological Geographical Features*

Section	Data quality element	Data quality sub-element	Definition	Evaluation Scope
7.4.1	Logical consistency	Conceptual consistency	adherence to rules of the conceptual schema	dataset series; dataset; spatial object type; spatial object

7.4.2	Logical consistency	Domain consistency	adherence of values to the value domains	dataset series; dataset; spatial object type; spatial object
-------	---------------------	--------------------	--	--

Recommendation 12 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.

7.4.1 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 13 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

7.4.2 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 14 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

7.5 Minimum data quality requirements

No minimum data quality requirements are defined for the spatial data theme Atmospheric Conditions and Meteorological Geographical Features.

7.6 Recommendation on data quality

No minimum data quality recommendations are defined.