ISO/TC 211 **N 5818**

ISO/TC 211 "Geographic information/Geomatics"
Secretariat: **SIS**
Committee manager: **Åhlin Mats Mr**

Text for FDIS 19157-1, Geographic information — Data quality — Part 1: General requirements, as sent to ISO/CS

Document type	Related content	Document date	Expected action
Project / Draft	Project: ISO/FDIS 19157-1	2022-11-01	**INFO**

Description

The FDIS ballot starts 2022-11-11 and ends 2023-01-06.

ISO/TC 211 N 5818

2022-09-19

ISO 19157-1:2022

ISO TC 211/WG 9

Geographic information — Data quality — Part 1: General requirements

FDIS stage

Contents

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 211, *Geographic information/Geomatics,* in collaboration with the European Committee for Standardization (CEN) Technical Committee CEN/TC 287, *Geographic Information,* in accordance with the Agreement on technical cooperation between ISO and CEN (Vienna Agreement).

This first edition cancels and replaces the first edition (ISO 19157:2013), which has been technically revised. It also incorporates the Amendment ISO 19157:2013/Amd 1:2018.

The main changes are as follows:

— terminology has been harmonized;

— the unique identification of normative components has been added;

— the definition of data quality model extension has been added;

— data quality measures have been moved into a new project on a standard data quality measures register;

– the conformance requirements in have been updated;

– the usage of package prefixes for type name has been omitted;

– the ‘usability’ data quality element has been removed from the model;

– new clause on extending standard quality model and the quality measures has been added;

– abstract test suit has been revised;

– requirements for XML schema implementation has been added;

– information on backwards compatibility with superseded edition of the standard has been included;

A list of all parts in the ISO 19157 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Geographic data are increasingly being shared, interchanged and used for purposes other than their producers' intended ones. Information about the quality of available geographic data is vital to the process of selecting a dataset in that the value of data are directly related to their quality. A user of geographic data can have multiple datasets from which to choose. Therefore, it is necessary to compare the quality of the datasets to determine which best fulfils the requirements of the user.

The purpose of describing the quality of geographic data is to facilitate the comparison and selection of the dataset best suited to application needs or requirements. Complete descriptions of the quality of a dataset will encourage the sharing, interchange and use of appropriate datasets. Information on the quality of geographic data allows a data producer to evaluate how well a dataset meets the criteria set forth in its product specification and assists data users in evaluating a product's ability to satisfy the requirements for their particular application. For the purpose of this evaluation, clearly-defined procedures are used in a consistent manner.

To facilitate comparisons, it is essential that the results of the quality are expressed in a comparable way and that there is a common understanding of the data quality measures that have been used. These data quality measures provide descriptors of the quality of geographic data through comparison with the universe of discourse. The use of incompatible measures makes data quality comparisons impossible to perform. This document standardizes the components and structures of data quality measures and defines commonly used data quality measures.

This document recognizes that a data producer and a data user can potentially view data quality from different perspectives. Conformance quality levels can be set using the data producer's product specification or a data user's data quality requirements. If the data user requires more data quality information than that provided by the data producer, the data user can follow the data producer's data quality evaluation process flow to get the additional information. In this case the data user requirements are treated as a product specification for the purpose of using the data producer process flow.

The objective of this document is to provide a framework for defining the quality of geographic data. This includes principles for evaluating quality, a conceptual model for handling quality information, a structure and content of data quality measures, and guidelines for reporting a quality evaluation. The framework is extensible, with rules for how to add additional data quality measures, and also has provision for complex dimensions of data quality.

Geographic information — Data quality — Part 1: General requirements

1. Scope

This document establishes the principles for describing the quality of geographic data. It:

— defines a well-considered system of components for describing data quality;

— defines the process for defining additional, domain specific components for describing data quality;

— specifies components and the content structure of data quality measures;

— describes general procedures for evaluating the quality of geographic data;

— establishes principles for reporting data quality.

This document is applicable to data producers providing quality information to describe and assess how well a dataset conforms to its product specification and to data users attempting to determine whether or not specific geographic data are of sufficient quality for their particular application.

This document does not attempt to define minimum acceptable levels of quality for geographic data. Such information is usually present as a requirement in a data product specification, defined in accordance with ISO 19131, for example.

2. Normative references

The following referenced documents, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO 19103:2015, *Geographic information — Conceptual schema language*

ISO 19109:2015, *Geographic information — Rules for application schema*

ISO 19115-1:2014, *Geographic information — Metadata — Part 1: Fundamentals*

3. Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
accuracy
closeness of agreement between a test result or measurement result and the true value

Note 1 to entry: In this document, the true value can be a reference value that is accepted as true.

[SOURCE: ISO 3534-2:2006, 3.3.1, modified – Notes 1, 2 and 3 to entry have been deleted. A new Note to entry has been added.]

3.2
conformance
conformity
fulfilment of a requirement

Note 1 to entry: When there is no ambiguity, the modifier “conformance” may be omitted. For example, “test report” is the same as “conformance test report”.

[SOURCE: ISO 19105:2022, 3.4]

3.3
conformance quality level
threshold value or set of threshold values for data quality results used to determine how well a dataset meets the criteria set forth in its data product specification or user requirements

3.4
correctness
correspondence with the universe of discourse

3.5
coverage
feature that acts as a function to return values from its range for any direct position within its domain

[SOURCE: ISO/DIS 19123-1]

3.6
data product specification
specification of a data product together with additional information that will enable it to be created, supplied to and used by another party

Note 1 to entry: A data product specification provides a description of the universe of discourse and a specification for mapping the universe of discourse to a data product. It may be used for production, sales, end-use or other purposes.

Note 2 to entry: specification is a document stating requirements [ISO 9000:2015, 3.8.7].

Note 3 to entry: data product is a dataset or a dataset series that may be supplied [ISO/FDIS 19131, modified – Note to entry has been removed].

[SOURCE: ISO/FDIS 19131, 3.9, modified – Note 2 and 3 to entry have been added]

3.7
data quality
degree to which a set of inherent characteristics of data fulfils requirement

[SOURCE: ISO 8000-2:2020, 3.8.1, modified – Note has been deleted]

3.8
data quality measure
variable to which a value is assigned as the result of measurement of a data quality characteristic

[SOURCE: ISO/IEC 25012:2008, 4.5, modified – Note has been deleted]

3.9
data quality unit
combination of a scope and data quality elements

3.10
dataset
identifiable collection of data

Note 1 to entry: A dataset can be a smaller grouping of data which, though limited by some constraint such as spatial extent or *feature type*, is located physically within a larger dataset. Theoretically, a dataset can be as small as a single *feature* or *feature attribute* contained within a larger dataset. A hardcopy map or chart can be considered a dataset.

[SOURCE: ISO 19115-1:2014, 4.3]

3.11
dataset series
collection of datasets sharing common characteristics

[SOURCE: ISO 19115-1:2014, 4.4]

3.12
feature
abstraction of real world phenomena

Note 1 to entry: A feature may occur as a type or an instance. *Feature type* or *feature instance* will be used when only one is meant.

[SOURCE: ISO 19101-1:2014, 4.1.11]

3.13
feature attribute
characteristic of a feature

Note 1 to entry: A feature attribute has a name, a data type and a value domain associated with it. A feature attribute for a feature instance also has an attribute value taken from the value domain.

[SOURCE: ISO 19101-1:2014, 4.1.12 modified – Examples 1 and 2, and Notes 2 and 3 have been deleted.]

3.14
feature instance
individual of a given feature type having specified feature attribute values

[SOURCE: ISO 19101-1:2014, 4.1.14]

3.15
feature operation
operation that every instance of a feature type may perform

[SOURCE: ISO 19110:2016, 3.7 - modified, Example and Note have been removed.]

3.16
feature type
class of features having common characteristics

[SOURCE: ISO 19156:2011, 4.7]

3.17
geographic data
data with implicit or explicit reference to a location relative to the Earth

[SOURCE: ISO 19109:2015, 4.13, modified – Note has been deleted.]

3.18
item
anything that can be described and considered separately

Note 1 to entry: An item can be any part of a *dataset*, such as a *feature*, feature relationship, *feature attribute*, or combination of these.

[SOURCE: ISO 2859-5:2005, 3.4, modified – Example has been removed. Note 1 to entry has been added.]

3.19
lineage
provenance, source(s) and production process(es) used in producing a resource

[SOURCE: ISO 19115-1:2014, 4.9]

3.20
metadata
information about a resource

[SOURCE: ISO 19115-1:2014, 4.10]

3.21
metaquality
information describing the quality of data quality

3.22
quality
degree to which a set of inherent characteristics of an object fulfils requirements

[SOURCE: ISO 9000:2015, 3.6.2, modified - Note 1 and 2 have been removed.]

3.23
quality evaluation
systematic examination of the extent to which an entity is capable of fulfilling specified requirements

[SOURCE: ISO/IEC/IEEE 24765:2017, 3.3267, modified – Note 1 has been removed]

3.24
register
set of files containing identifiers assigned to items with descriptions of the associated items

[SOURCE: ISO 19135-1:2015, 4.1.9]

3.25
requirement
need or expectation that is stated, generally implied or obligatory

[SOURCE: ISO 9000:2015, 3.6.4, modified – Notes 1, 2, 3, 4, 5 and 6 have been removed.]

3.26
quality evaluation report
quality report
free text document providing fully detailed information about data quality evaluations, results and measures used

3.27
uncertainty
measurement uncertainty
parameter, associated with the result of measurement, that characterizes the dispersion of values that could reasonably be attributed to a measurand

Note 1 to entry: Uncertainty of measurement comprises, in general, many components. Some of these components may be evaluated from the statistical distribution of the results of series of measurements and can be characterized by experimental standard deviations. The other components, which can also be characterized by standard deviations, are evaluated from assumed probability distributions based on experience or other information.

[SOURCE: ISO 19116: 2019, 3.28, modified – Note 1 to entry has been removed and replaced with Note 2 to entry from ISO 19101-2: 2018, 3.40]

3.28
universe of discourse
view of the real or hypothetical world that includes everything of interest

[SOURCE: ISO 19101-1:2014, 4.1.38]

4. Abbreviated terms and packages

4.1 Abbreviated terms

ADQR	aggregated data quality result
AQL	acceptance quality limit
GML	Geographic Markup Language
UML	Unified Modelling Language
XML	Extensible Markup Language

4.2 Abbreviated packages

Abbreviations are used to denote the package that contains a class. Those abbreviations precede class names, connected by a "_". The International Standard in which those classes are located is indicated in parentheses. A list of those abbreviations is listed in Table 1.

Table 1 — Package abbreviations

CI	Citation [ISO 19115-1:2014]
DS	Dataset [ISO 19115-1:2014]
GF	General Feature [ISO 19109:2015]
LI	Lineage [ISO 19115-1:2014]
MD	Metadata [ISO 19115-1:2014]

5. Conformance

5.1 General

In this document two conformance classes are defined (see Clause 5.2 and Clause 5.3). The related tests are provided in the abstract test suit in Annex A.

Requirements, recommendations and permissions are explicitly marked and assigned a requirement, recommendation or permission identifier.

5.2 Content of a data quality model

Table 2 describes the conformance class for the content of a data quality definition.

Table 2 — Content conformance class

Conformance class	**https://standards.isotc211.org/19157/-1/1/conf/content**
Standardization target type	Instance of a data quality definition, regardless of data encoding.
Dependency	https://standards.isotc211.org/19103/-/1/ (Conceptual schema language) https://standards.isotc211.org/19109/-/2/ (Rules for application schema) https://standards.isotc211.org/19115/-1/1/ (Metadata – Part 1: Fundamentals)
Requirements class	https://standards.isotc211.org/19157/-1/1/req/content (see Clause 6)
Tests	All tests in Clause A.1.

5.3 XML encoding of a data quality model

Table 3 describes the conformance class for the XML representation of a data quality model.

Table 3 — XML encoding conformance class

Conformance class	**https://standards.isotc211.org/19157/-1/1/conf/xml**
Standardization target type	XML document representing a data quality model
Dependency	https://standards.isotc211.org/19157/-1/1/conf/content
Requirements class	https://standards.isotc211.org/19157/-1/1/req/xml (see Clause 12)
Tests	All tests in Clause A.2.

6. General requirements for geographic information quality

6.1 General

Clauses 8 to 11 describe the components of data quality based on the conceptual Unified Modeling Language (UML) model defined in this document, which is part of the ISO/TC 211 harmonized model. Furthermore, additional descriptions, requirements and recommendations are included. A dictionary of elements in the UML model is provided in Annex C.

The requirements, recommendations and permissions defined in this document are summarized in Tables 3, 4, 5 and 6.

6.2 Data quality — general requirements, recommendations and permissions

Table 4 lists the requirements defined in this document.

Table 4 — List of requirements

Requirement class	**https://standards.isotc211.org/19157/-1/1/req/content**
Standardization target type	Instance of a data quality definition regardless of data encoding
Dependency	https://standards.isotc211.org/19103/-/1/ (Conceptual schema language)
Dependency	https://standards.isotc211.org/19109/-/2/ (Rules for application schema)
Dependency	https://standards.isotc211.org/19115/-1/1/ (Metadata – Part 1: Fundamentals)
Requirement 1	https://standards.isotc211.org/19157/-1/1/req/content/dataQuality
Requirement 2	https://standards.isotc211.org/19157/-1/1/req/content/additionalQualityElement
Requirement 3	https://standards.isotc211.org/19157/-1/1/req/content/qualityMeasure
Requirement 4	https://standards.isotc211.org/19157/-1/1/req/content/additionalQualityMeasure
Requirement 5	https://standards.isotc211.org/19157/-1/1/req/content/dataQualityMetadata
Requirement 6	https://standards.isotc211.org/19157/-1/1/req/content/qualityEvaluationReport
Requirement 7	https://standards.isotc211.org/19157/-1/1/req/content/aggregatedResult
Requirement 8	https://standards.isotc211.org/19157/-1/1/req/xml

In addition to the requirements class in Table 4, a recommendations class is defined in Table 5 and a permission class is defined in Table 6. These recommendations and permissions, when implemented, can contribute to improved content of a data quality definition. However, recommendations and permissions are optional and do not have an impact on the results of the conformance testing, thus the recommendation class in Table 5 and a permission class is defined in Table 6 are not related to any of the conformance classes defined in Clause 4.

Table 5 — List of recommendations

Recommendation class	**https://standards.isotc211.org/19157/-1/1/req/optionalContent**
Standardisation target type	Instance of a data quality definition regardless of data encoding
Dependency	https://standards.isotc211.org/19157/-1/1/req/content
Recommendation 1	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/dateTime
Recommendation 2	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/identifier
Recommendation 3	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/additionalQualityMeasure
Recommendation 4	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/description
Recommendation 5	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/qualityEvaluationProcess
Recommendation 6	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/evaluationMethodType
Recommendation 7	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/reportReference
Recommendation 8	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/aggregatedResult
Recommendation 9	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/sameQualityElement
Recommendation 10	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/differentQualityElement
Recommendation 11	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/qualityEvaluationReport
Recommendation 12	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/derivedResultMetadata
Recommendation 13	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/hierarchy

Table 6 lists the permissions defined in this document.

Table 6 — List of permissions

Permission class	**https://standards.isotc211.org/19157/-1/1/req/permittedContent**
Standardization target type	Instance of a data quality definition regardless of data encoding
Dependency	https://standards.isotc211.org/19157/-1/1/req/content
Permission 1	https://standards.isotc211.org/19157/-1/1/per/permittedContent/additionalQualityElement
Permission 2	https://standards.isotc211.org/19157/-1/1/per/permitted Content/conformanceResult
Permission 3	https://standards.isotc211.org/19157/-1/1/per/permittedContent/sourceReference

7. Overview of data quality

Working with data quality includes:

— understanding of the concepts of data quality related to geographic data. Annex B is a description of data quality concepts used to establish the components for describing the quality of geographic data;

— defining data quality conformance levels in data product specifications or based on user requirements. Establishment of data product specifications is described in ISO 19131.

— specifying quality aspects in application schemas;

— evaluating data quality and metaquality;

— reporting data quality and metaquality.

NOTE 1 The development of application schemas is described in ISO 19109.

A data quality evaluation can be applied to dataset series, a dataset or a subset of data within a dataset, sharing common characteristics so that its quality can be evaluated.

Data quality elements and their descriptors are used to describe how well a dataset meets the criteria set forth in its data product specification or user requirements and provide quantitative quality information.

Requirement 1:	https://standards.isotc211.org/19157/-1/1/req/content/dataQuality
	Data quality shall be described in conformance with the components of data quality and data quality measures as defined by the UML figures in Clause 8 and 9 in this document.

When data quality information describes data that have been created without a detailed data product specification or with a data product specification that lacks quantitative measures and descriptors, data quality may be evaluated in a non-quantitative subjective way with a descriptive result (see 8.5.4.4) used to report the evaluation result for each of the applicable data quality elements.

Some quality related information is provided by purpose, usage and lineage. This information is reported as metadata in conformance with ISO 19115-1.

NOTE 2 Purpose describes the rationale for creating a dataset and contains information about its intended use, which can potentially not be the same as the actual use of the dataset. Usage describes the application(s) for which a dataset has been used, either by the data producer or by other data users. Lineage describes the history of a dataset and recounts the life cycle of a dataset from collection and acquisition through compilation and derivation to its current form. This general, non-quantitative information is illustrative for users and can help assessing the quality of a dataset, especially in cases where it is used for a particular application that differs from the intended application (see also 10.2.3).

This document recognizes that data quality elements may have associated quality which is termed metaquality. Metaquality describes the quality of the data quality results in terms of defined characteristics.

NOTE 3 The concept of metaquality is described in 8.3.7 and the descriptors of metaquality are defined in 8.5.5.

Figure 1 provides an overview of data quality information.

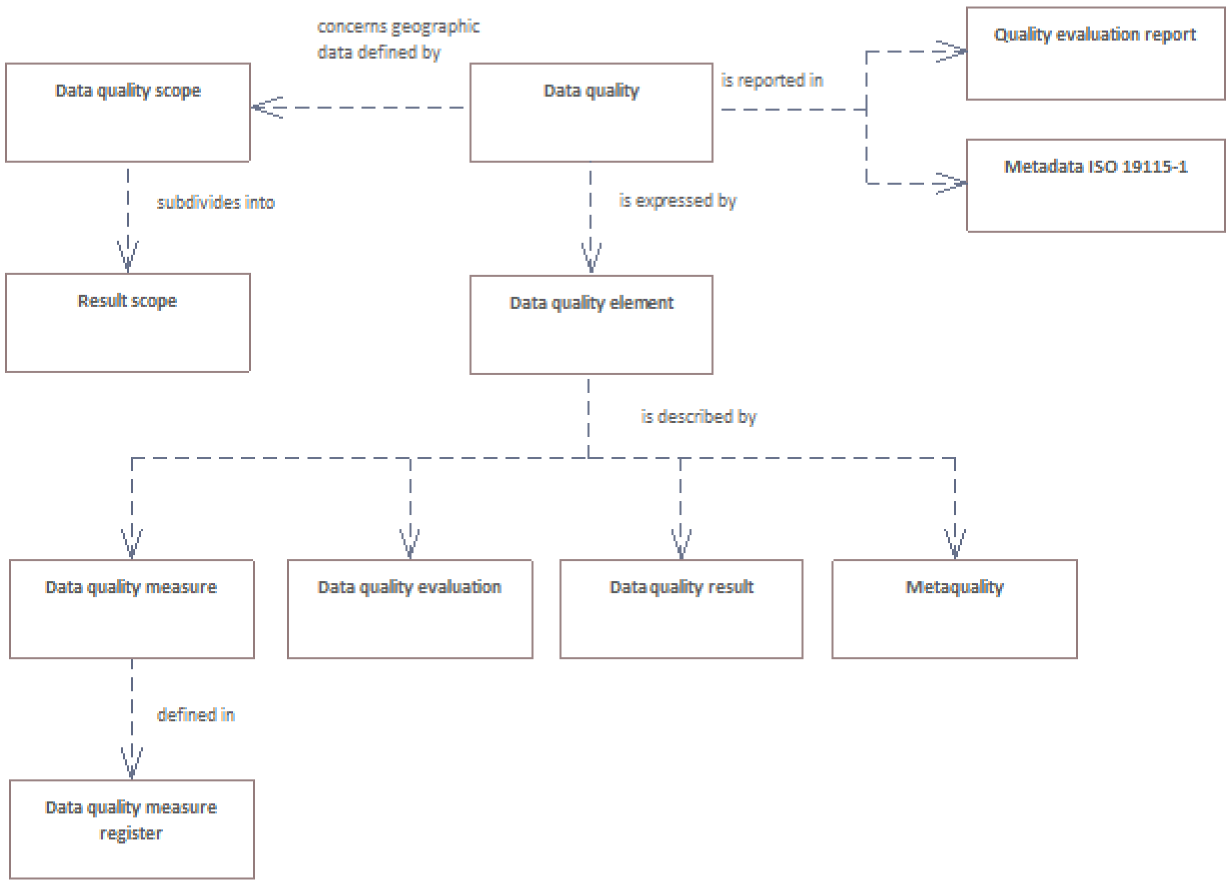


Figure 1 — Conceptual model of quality for geographic data

8. Components of data quality

8.1 Overview of the components

The components of data quality are described in Clause 8. Figure 2 presents an overview of the components and the connections between them. See the data dictionary defined in Annex C for more details about components and their attributes.

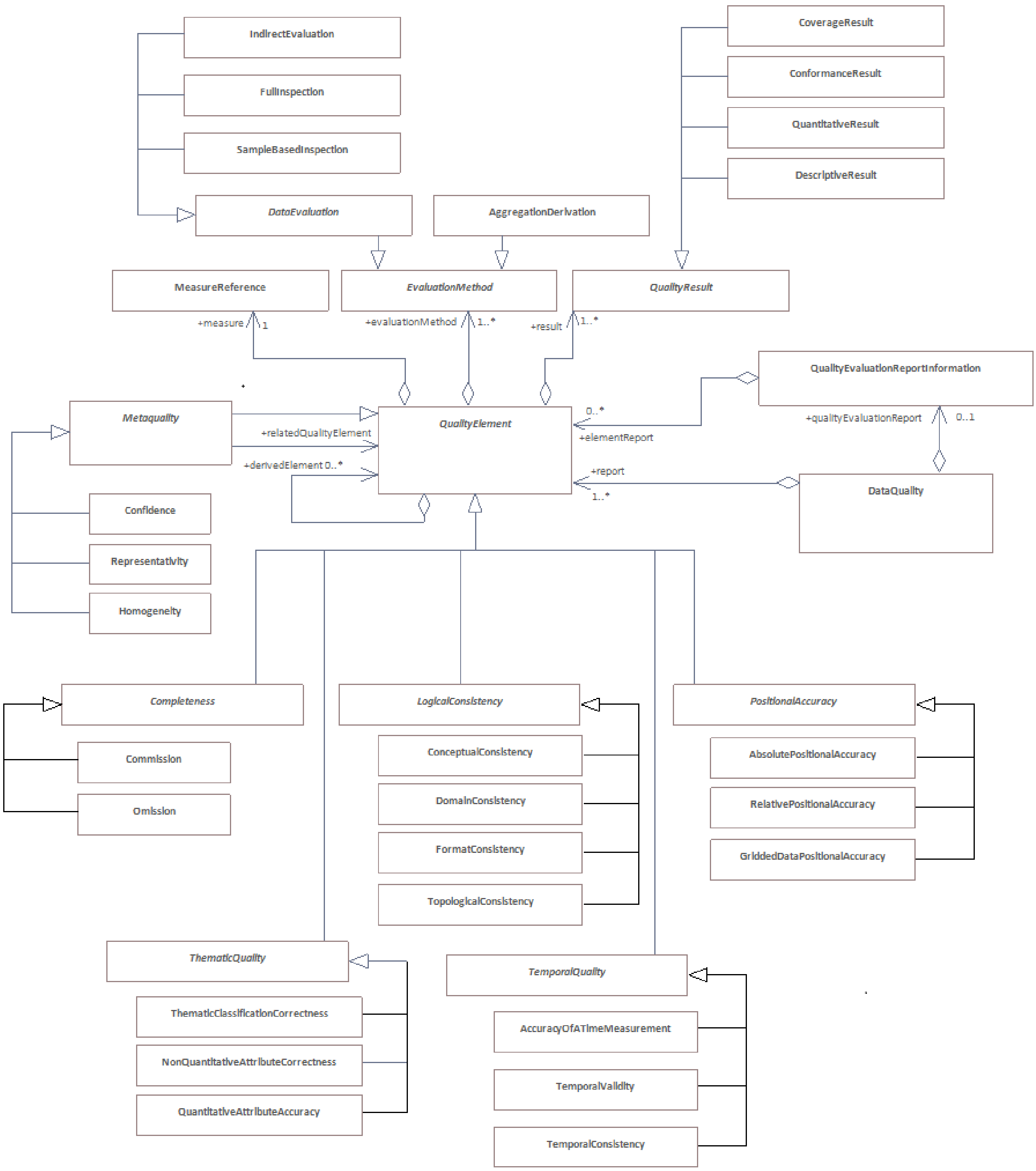


Figure 2 — Overview of the components of data quality

8.2 Data quality unit

When describing the quality of geographic data, different quality elements and different subsets of the data may be considered. In order to describe these, data quality units are used. A data quality unit is the combination of a scope and data quality elements, see Figure 3 and Table C.1.

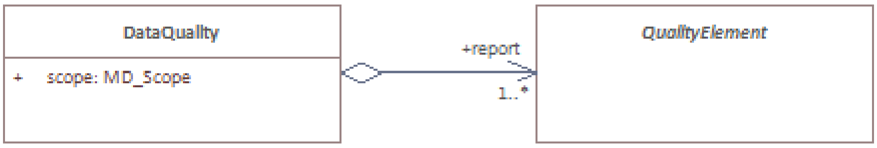


Figure 3 — Data quality unit

The scope of the data quality unit(s) specifies the extent, spatial and/or temporal, and/or common characteristic(s) that identify the data on which data quality is to be established and evaluated.

Since scopes are often different for individual data quality elements, one data quality report (metadata or quality evaluation report) may encompass several data quality units. These different scopes may be, for example, spatially separate, overlapping or even sharing the same extents.

NOTE In case a dataset is a result of integration of datasets of various scopes, data quality values or weights of quality elements, such new dataset needs to establish its new data quality unit, i.e. define its own data quality scope and specify quality elements of interest.

The following are examples of what defines a data quality scope (see also MD_Scope in ISO 19115-1):

a) a dataset series;

b) a dataset;

c) a subset of data defined by one or more of the following characteristics:

   1) types of items (sets of feature types, feature attributes, feature operations or relationships among features);

   2) specific items (sets of feature instances, attribute values or instances of feature relationships);

   3) geographic extent;

   4) temporal extent (the time frame of reference and accuracy of the time frame);

d) a coverage.

8.3 Data quality elements

8.3.1 General

A data quality element is a component describing a certain aspect of the quality of geographic data and, in this document, these have been organized into groups of quality elements, each group defined in 8.3.2 through to 8.3.7. The relationship among various aspects of quality and their corresponding quality elements are illustrated in Figure 4, where groups of elements are represented by abstract classes and elements are represented by their subclasses.

NOTE Depending on the tool used and/or organizational preferences, implementations of the data quality conceptual model presented in Figure 4 vary. Advising on best implementation practice is out of the scope of this document.

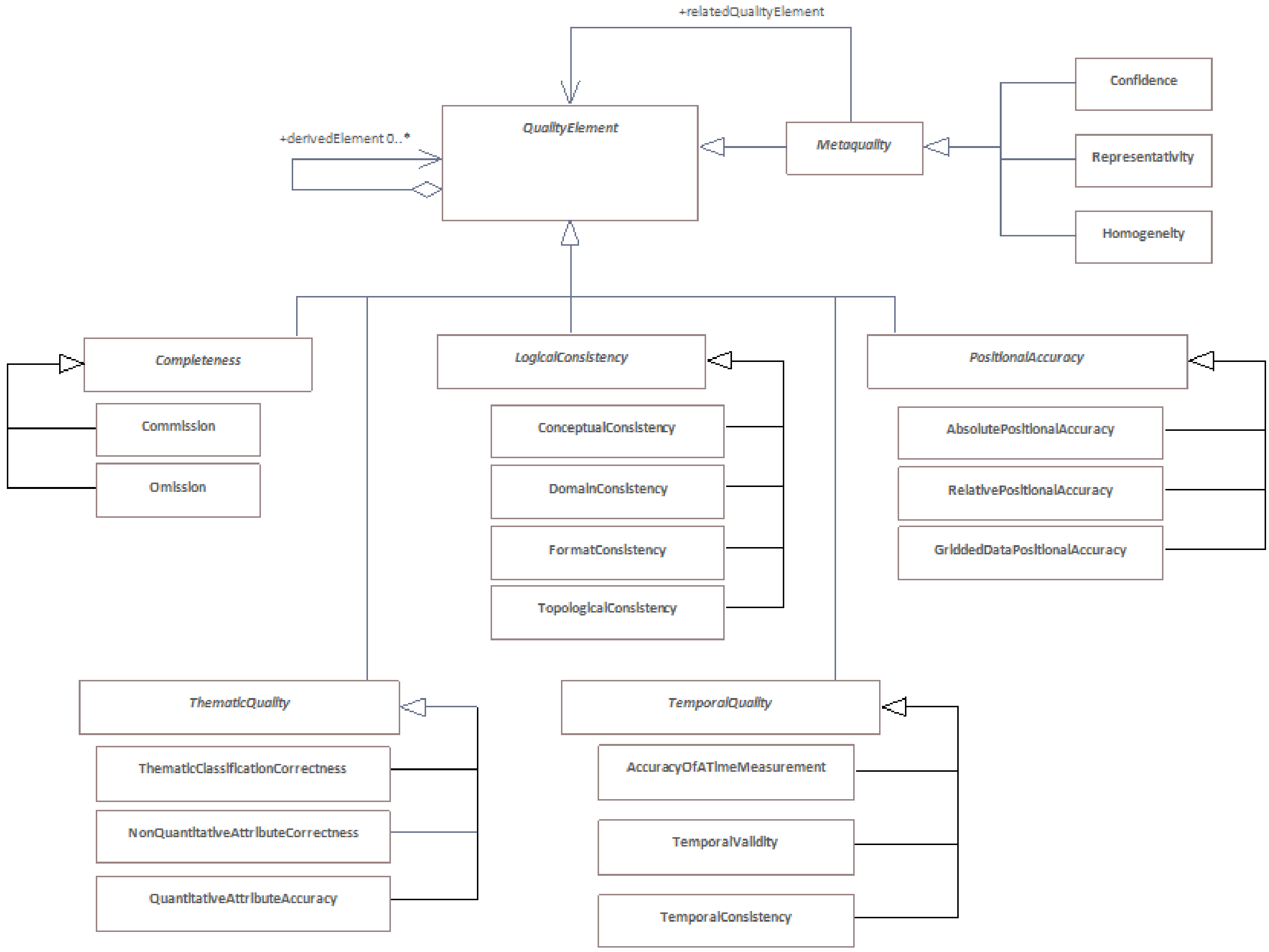


Figure 4 — Overview of the data quality elements

Permission 1.	https://standards.isotc211.org/19157/-1/1/per/permittedContent/additionalQualityElement
	In case quality elements defined in this standard are not sufficient for expressing the quality of the dataset, new elements can be defined and used according to the requirements specified in Clause 8.4

8.3.2 Completeness

Completeness is defined as a degree to which a dataset (or dataset series) representing a feature (or features) has values for all expected attributes and related entity instances at the time when the dataset has been created (ISO 8000-2; ISO/IEC 25012). As such, completeness refers to the presence and/or absence of features, their attributes and relationships. In this document, two elements (and their associated measures) define completeness:

— commission: excess data present in a dataset;

— omission: data absent from a dataset.

8.3.3 Logical consistency

Logical consistency is defined as the degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical). If these logical rules are documented elsewhere (for example, in a data product specification) then the source should be referenced (for example, in the data quality evaluation). This document specifies logical consistency with data quality elements (and their associated measures):

— conceptual consistency: adherence to rules of the conceptual schema;

— domain consistency: adherence of values to the value domains;

— format consistency: degree to which data are stored in accordance with the physical structure of the dataset;

— topological consistency: correctness of the explicitly encoded topological characteristics of a dataset.

8.3.4 Positional accuracy

Positional accuracy relates to measurement accuracy (ISO/IEC GUIDE 98-3; ISO/IEC GUIDE 99) and it represents a closeness of agreement between a measured position of features and a position accepted as true within a spatial reference system. Depending on the scope and type of the reference system, this document specifies the following three data quality elements for expressing positional accuracy:

— absolute or external accuracy: closeness of reported coordinate values to values accepted as true in a standard coordinate reference system;

— relative or internal accuracy: closeness of the relative positions of features in a related dataset to their respective relative positions accepted as true in a local coordinate reference system;

— gridded data positional accuracy: closeness of gridded data spatial position values to values accepted as true.

Positional accuracy is an estimate of the uncertainty of measurement results (ISO 19116) and should therefore be expressed with measures of uncertainty.

8.3.5 Temporal quality

Temporal quality is defined as the quality of the temporal attributes and temporal relationships of features. Temporal quality can be described by the following data quality elements (and their associated measures):

— accuracy of a time measurement: closeness of reported time measurements to values accepted as true;

— temporal consistency: correctness of the order of events.

— temporal validity: validity of data with respect to time.

NOTE 1 Timeliness, currentness (ISO/IEC 25012), up-to-dateness and actuality are used as synonyms to temporal validity.

NOTE 2 Temporal validity is different from domain consistency of temporal values. For example, a value '33 March 2020' representing a date indicates a domain inconsistency. An example of temporal invalidity is a dataset that contains objects that did not exist in the timestamp associated with their record.

8.3.6 Thematic quality

Thematic quality is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. This document specifies thematic quality with three data quality elements (and their associated measures):

— classification correctness: comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference data);

— non-quantitative attribute correctness: measure of whether a non-quantitative attribute is correct or incorrect;

— quantitative attribute accuracy: closeness of the value of a quantitative attribute to a value accepted as or known to be true.

8.3.7 Metaquality elements

Metaquality elements are a set of quantitative and qualitative statements about a quality evaluation and its result. The knowledge about the quality and the suitability of the evaluation method, the measure applied, and the given result may be of the same importance as the result itself.

See D.5.3 for an example of metaquality evaluation.

Metaquality should be described using the following elements:

— Confidence: trustworthiness of a data quality result.

NOTE 1 Quantitative figures for confidence can be obtained by statistical parameters such as standard deviation or a confidence interval on a given confidence level.

EXAMPLE Confidence originates primarily from the method used and of its reliability, as well, to a lesser extent, from the concerned population.

— Representativity: degree to which the sample used has produced a result which is representative of the data within the data quality scope.

NOTE 2 A statistical method based on sampling could be considered as reliable as a global method when all the geographic zones and concerned time periods are covered and the population is sufficiently large. It is not only the size of the sample which is crucial but also how well it represents the actual state of the data. See also 10.2.2 and Annex E.

— Homogeneity: expected or tested uniformity of the results obtained for a data quality evaluation.

NOTE 3 Homogeneity consists of comparing the evaluation results of several segments of a global dataset. This comparison can be expressed using root mean square errors for example (as in E.5.3). In the case of a general process, homogeneity cannot be evaluated because the result is global.

NOTE 4 These tests are often conducted when data has been captured by different operators, depending on the acquisition zone or the acquisition date.

8.4 Extending the data quality information model

This document recognizes that for many domain-specific purposes, it is necessary or convenient to extend the standard data quality information model as defined in 8.3. The extension includes adding data quality elements (or one of its specializations as defined in 8.3) and data quality measures.

EXAMPLE Data producers can wish to provide means for their users to record feedback on a dataset. For example, an information model for such feedback (i.e. new quality elements) can be defined with elements from [36].

Requirement 2:	https://standards.isotc211.org/19157/-1/1/req/content/additionalQualityElement
	New quality element shall be defined as a specialization of *QualityElement*, (see 8.3) in compliance with the rules for metadata extensions (ISO 19115-1:2014, Annex C) and rules for domain profiles of standard schemas (ISO 19109:2015, 8.3). NOTE If applicable, elements defined in other International Standards (e.g. ISO/IEC 25000 series) can be used for extending the set of quality components as defined in Clause 8.

8.5 Descriptors of data quality elements

8.5.1 General

A data quality element is described by a reference to a quality measure, an evaluation method, quality result, and a metaquality element. These are shown in Figure 5, and are described in 8.5.2, 8.5.3, 8.5.4, and 8.5.5.

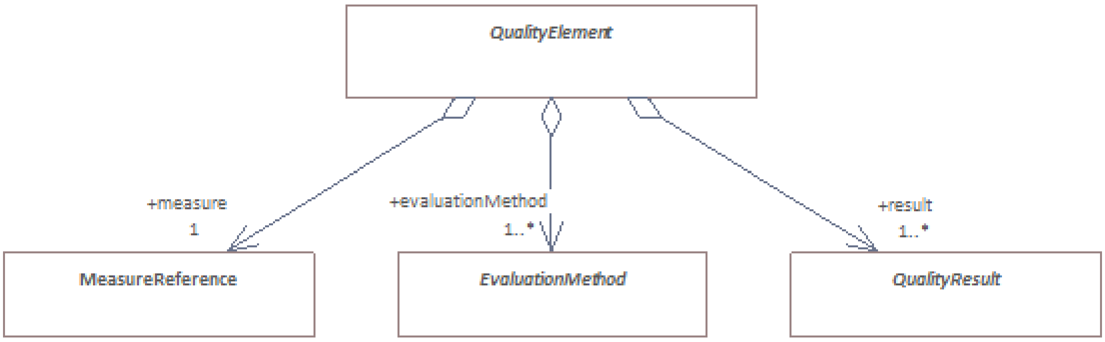


Figure 5 — Data quality element descriptors

8.5.2 Measure Reference

Data quality elements refer to measures by means of a measure reference (see Figure 6), providing an identifier of a measure fully described elsewhere.

NOTE The whole description can be found within a measure register or catalogue, which can form part of a data product specification or a quality evaluation report.

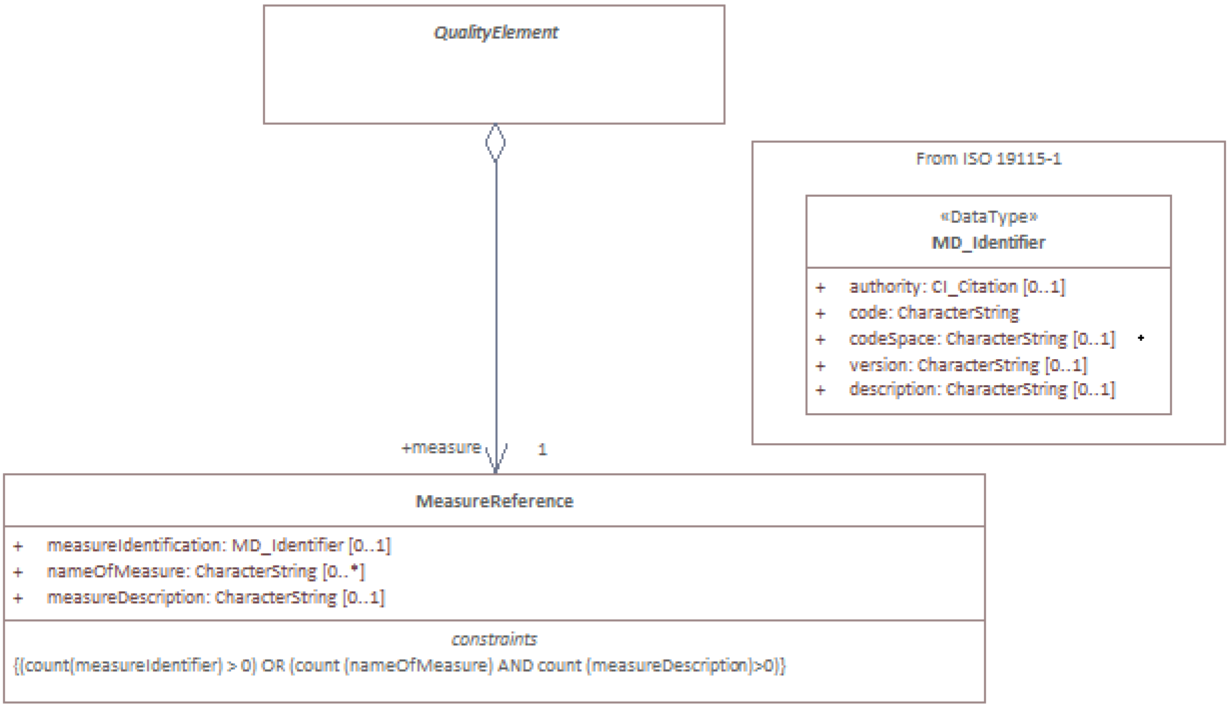


Figure 6 — Data quality measure reference

Data quality measures are further described in Clause 9 of this document.

EXAMPLE The percentage of the values of an attribute which are correct.

This document recognizes that the quality of a dataset is measured using a variety of methods. A single data quality measure can be insufficient for fully evaluating the quality of the data specified by a data quality scope and providing a measure of quality for all possible utilizations of a dataset. A combination of data quality measures can give useful information. Multiple data quality measures may be reported for the data specified by a data quality scope. The data quality report should then include one instance of *QualityElement* for each measure applied.

8.5.3 Evaluation method

Data quality evaluation method describes those procedures and methods which are applied to the geographic data to arrive at a data quality result, see Figure 7. Different evaluations are often used for the various data quality elements.

Data quality evaluation method is used for describing, or for referencing documentation describing, the methodology used to apply a data quality measure to the data specified by a data quality scope.

NOTE Data quality evaluation is further described in Clause 10.

EXAMPLE Examples of documentation are data product specifications (e.g. ISO 19131), published articles or accepted industry standards.

Recommendation 1:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/dateTime
	One date or range of dates (compliant with ISO 19108) should be included for each evaluation. If the evaluation was carried out on non-consecutive dates, each single date should be included.

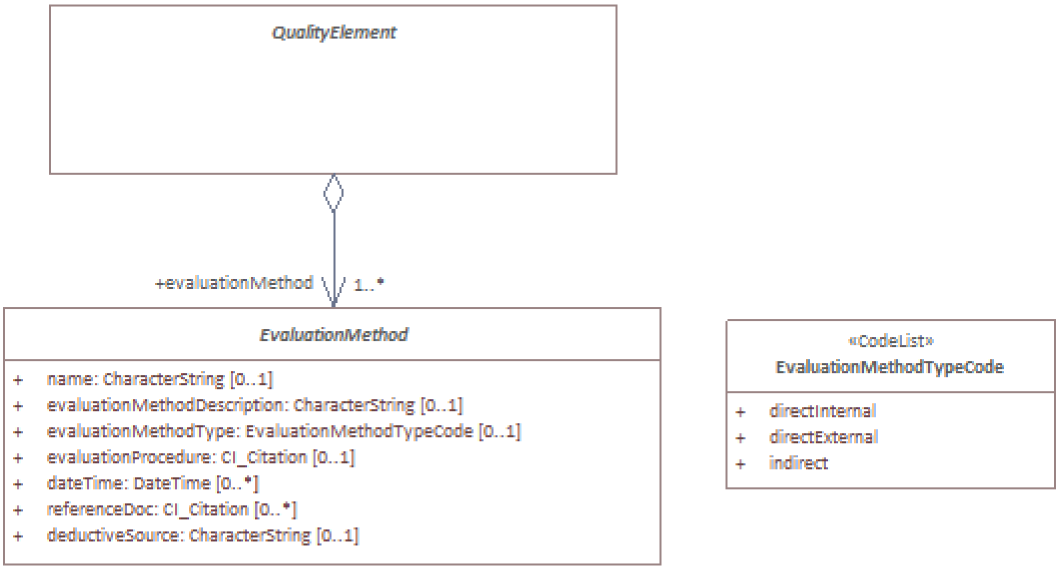


Figure 7 — Data quality evaluation method

8.5.4 Quality Result

8.5.4.1 General

Data quality result is provided for each data quality element. This could be a quantitative result, a conformance result, a descriptive result or a coverage result, see also Figure 8.

NOTE Different types of results can be provided for the same data quality element.

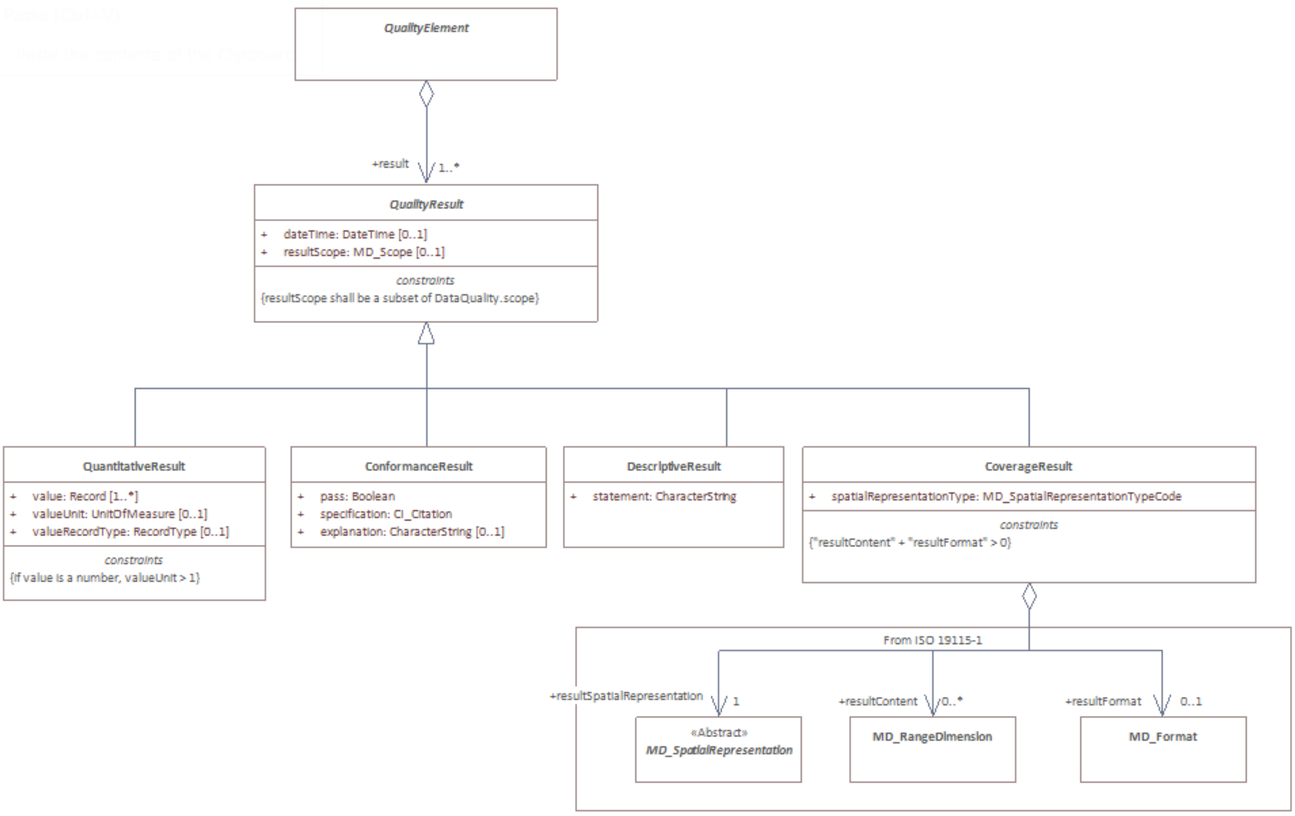


Figure 8 — Data quality result

Quality frequently differs between various parts of the dataset for which quality is evaluated. Therefore, several evaluations may be applied for the same data quality element to more completely and, in more detail, describe quantitative data quality. To avoid repeating the measure and evaluation procedure descriptions in several instances of data quality element (*QualityElement*), several results with individual result scopes can be used.

EXAMPLE A dataset contains features of identical type but whose positions have been established with various methods yielding different positional accuracies. The same quality evaluation method and the same measure are, however, applied for the whole dataset, and provide different results depending on the data acquisition method. In this case, it may be desirable to have several results with individual result scopes (the area covered by each data acquisition method) and one data quality scope (the dataset).

8.5.4.2 Quantitative result

Quantitative result may be a single value or multiple values, depending on the values of attributes *valueType* and *valueStructure* defined in the description of the measure applied.

The attribute *valueRecordType* is used to describe how the *valueType* and *valueStructure* defined in the measure are implemented to provide the value of the quantitative result.

NOTE The attribute valueRecordType is of type RecordType, which is a generic data type defined in ISO 19103. Its value changes depending on which implementation solution is used for providing the quantitative result. An example of Extensible Markup Language (XML) implementation for recordType is provided in ISO/TS 19139-1.

EXAMPLE 1 Using an XML implementation: simple example: value = 5, valueRecordType = gco:Integer, valueUnit = “metre.”

EXAMPLE 2 Within the description of the measure, the *valueType* is an integer, the *valueStructure* is an $n \times n$ (square) matrix. The *value* attribute of the quantitative result provides the result matrix itself, within a numeric encoding using a particular XML type called MatrixType. The attribute *valueRecordType* provides the description of the type MatrixType in XML. If another encoding is used, the attribute *valueRecordType* will change to provide the description of the type Matrix in the other encoding, and the implementation of the attribute value will change accordingly, but the value itself will not change.

EXAMPLE 3 Measure “Rate of excess items” is used to evaluate the number of excess items in the dataset in relation to the number of items that are expected to be present. The quantitative result value is of value type Real. The value unit is used in this case to show that the value is a percentage, the value has been multiplied by 100. In this example the value unit is “%”.

8.5.4.3 Conformance result

A conformance result is the outcome of comparing the value or set of values obtained from applying a measure to the data specified by a data quality scope with a specified conformance quality level (e.g. as recorded in an ISO 19131 compliant data product specification).

When a conformance quality level is defined, the obtained result is compared with this to evaluate if the quality of the data meets the specified level of quality.

A conformance result may be provided for each measure. The conformance quality level is typically specified in suitable reference documentation such as the data product specification (e.g. an ISO 19131 compliant data product specification) or a user defined requirements specification. If conformance is evaluated, a reference to the relevant reference documentation shall be made and the conformance quality level used shall be specified.

Conformance result is not to be used for specifying the conformance quality level in an ISO 19131 compliant data product specification.

Recommendation 2:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/identifier
	An *identifier* should be provided for the *specification* attribute of *ConformanceResult*. NOTE: *specification* is of type *CI_Citation*, in which, as defined by ISO 19115-1, the *identifier* is optional.

Permission 2.	https://standards.isotc211.org/19157/-1/1/per/permittedContent/conformanceResult
	If evaluation has been performed against conformance levels originating from different sources, more than one data quality *ConformanceResult* may be provided for the same measure.

8.5.4.4 Descriptive result

In some cases (e.g. with thematic and geoscientific observations), it is not possible to produce a quantitative result for a data quality element. A subjective evaluation of an element can then be expressed with a textual statement as a data quality descriptive result.

EXAMPLE The relative positional accuracy is higher between a geological feature and a nearby feature from a base map (roads, rivers, lakes etc.) than the absolute positional accuracy on the geological feature itself.

This descriptive result can also be used to provide a short synthetic description of the result of the data quality evaluation, to accompany the complete quantitative result or replace it, if no quantitative value can be provided.

8.5.4.5 Coverage result

Data quality result can be provided as an information in a coverage. A UML model for a coverage result is provided in Figure 8.

8.5.5 Descriptors of a metaquality element

A metaquality element is described by the same descriptors as for the quality element: measure, evaluation method and result (see Figure 9).

NOTE The related quality element is the element on which the metaquality element applies.

See E.5.3 for an example of metaquality evaluation.

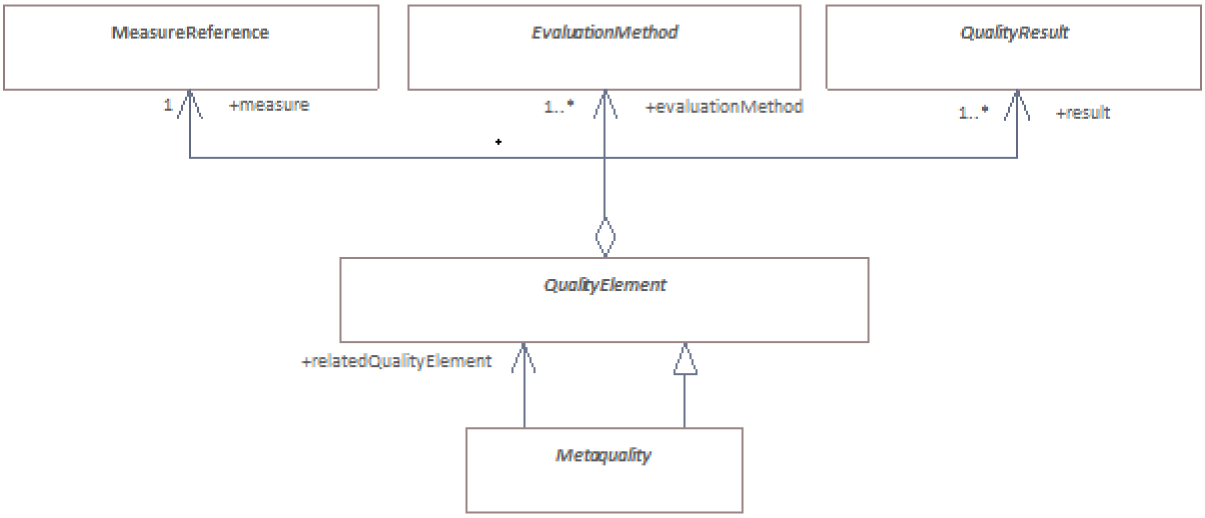


Figure 9 — Metaquality descriptors

9. Data quality measures

9.1 General

To facilitate dataset comparisons, evaluations and data quality reports (metadata or a quality evaluation report) have to be expressed in a comparable way, and it is necessary to have a common understanding of the data quality measures that have been used.

Requirement 3:	https://standards.isotc211.org/19157/-1/1/req/content/qualityMeasure
	When possible, measures defined in a quality measures register for geographic information shall be used in data quality evaluation and to represent the result of the quality evaluation in quality evaluation report and metadata. NOTE 1 If no measure as defined in a quality measures register for geographic information is applicable, additional, user defined measures can be used (see 9.3). NOTE 2 Elements defined in other standards (e.g. ISO/IEC 25000 series) can be used for extending the quality model, if applicable.

9.2 Standardized data quality measures

9.2.1 General

The structure of standard data quality measures is illustrated in Figure 10.

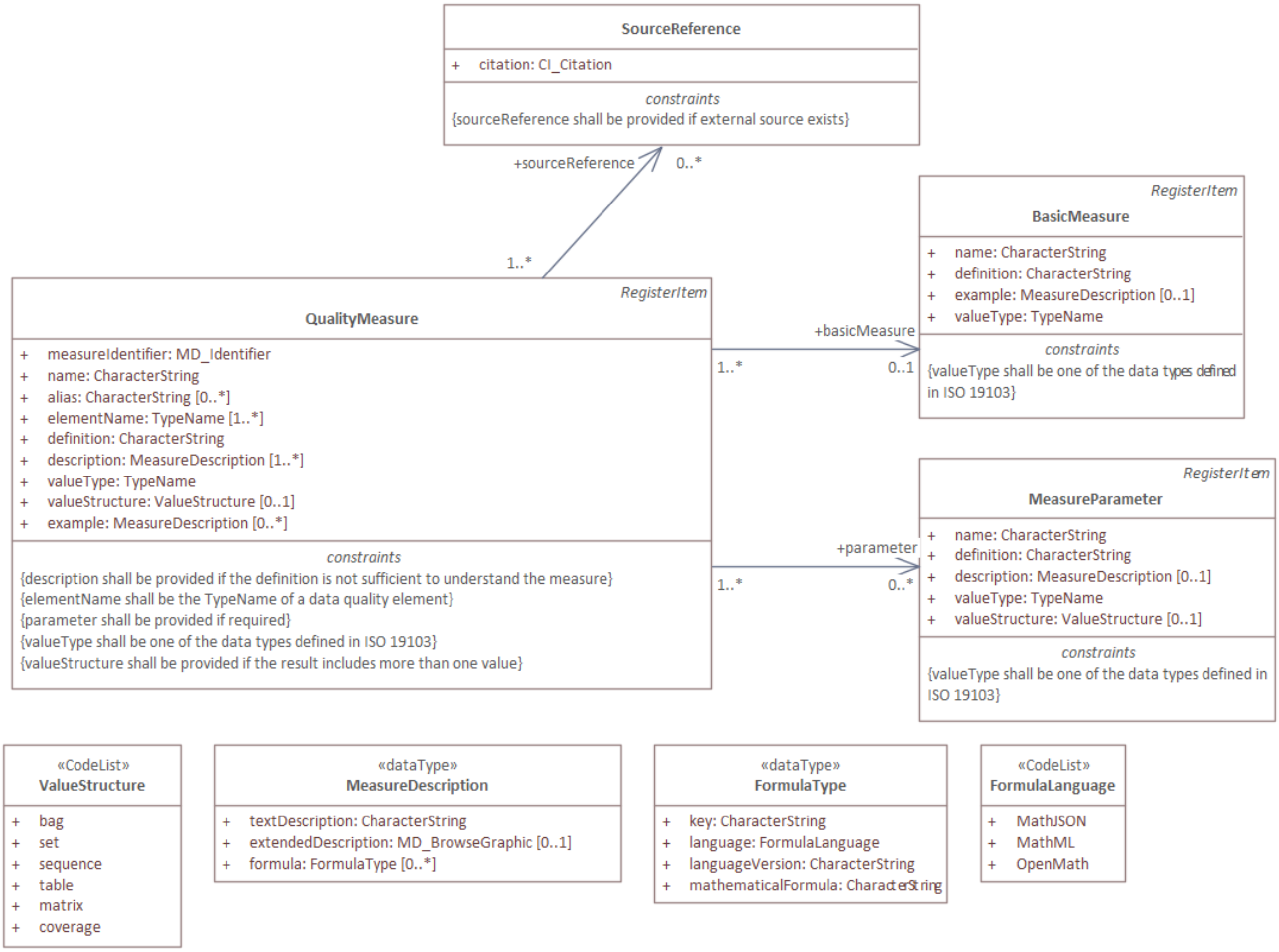


Figure 10 — Data quality measures

Each data quality measure is described by the following components:

— measure identifier (9.2.2);

— name (9.2.3);

— alias (9.2.4);

— element name (9.2.5);

— basic measure (9.2.6);

— definition (9.2.7);

— description (9.2.8);

— parameter (9.2.9);

— value type (9.2.10);

— value structure (9.2.11);

— source reference (9.2.12);

— example (9.2.13).

Multiple measures can be defined for each data quality element. The choice of which one to use will depend on the type of the data and its intended purpose. A list of standardized data quality measures is given in a data quality measure register defined in conformance with ISO 19135-1:2015.

9.2.2 Measure identifier

Identifier is a value uniquely identifying a measure within a namespace.

NOTE This identifier enables references to the data quality measure within the data quality elements (see 8.5.2).

9.2.3 Name

Name is the name of the measure.

NOTE If the measure already has a commonly used name, this name is be used. If no name exists, a name that reflects the nature of the measure is chosen.

9.2.4 Alias

Alias is another recognized name for the same data quality measure. It may be a different commonly used name, or an abbreviation, or a short name. More than one alias may be provided.

9.2.5 Element name

Element name is the name of the data quality element (see 8.3 and 8.4) to which a measure applies. More than one element name may be provided.

9.2.6 Basic measure

Basic measures are defined by their *name*, *definition*, *example* and *valueType* (see Figure 10).

The basic measures should also be used for creating new measures if applicable. For example, to report unclosed surface patches or other application-dependent measures.

Recommendation 3:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/additionalQualityMeasure
	Whenever possible, a *BasicMeasure* should be used as the basis for defining a new quality measure. NOTE A variety of measures are based on counting of erroneous items. There are also several measures dealing with the uncertainty of numerical values. In order to avoid repetition, the most common methods of constructing count-related measures, as well as general statistical measures, for one- and two-dimensional random variables are defined in terms of basic measures.

9.2.7 Definition

Definition is a representation of a concept by a descriptive statement which serves to differentiate it from related concepts (ISO 19104).

NOTE If the measure is derived from a basic measure, the definition is based on the basic measure definition and specialized for this measure.

9.2.8 Description

Measure description is the description of the measure including methods of calculation, with all formulae and/or illustrations needed to establish the result of applying the measure.

If the measure uses the concept of errors, it should be stated how an item is classified as incorrect. This is the case when the quality can only be reported as correct or incorrect.

Recommendation 4:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/description
	If a measure uses the concept of errors for reporting correct or incorrect items, *description* should contain an explanation of when item is classified as incorrect.

9.2.9 Parameter

A measure parameter is an auxiliary variable used by the measure.

9.2.10 Value type

Value type is the data type used for reporting the result of the measure. The data types defined in ISO 19103 shall be used.

9.2.11 Value structure

A result may consist of multiple values. In such cases, the result shall be structured using the value structure as given in C.3.3.

9.2.12 Source reference

Source reference is the citation of the documentation of the measure.

9.2.13 Example

Example is an example of applying the measure or the result obtained for the measure. More than one example may be provided.

9.3 User defined data quality measures

Due to the nature of quality and geographic data, the list of standardized data quality measures cannot be complete. There can potentially be cases where the user of this document has to devise other data quality measures. When possible, these measures shall be defined using the data quality basic measures and the measure shall be defined using the structure given in Clause 9. Additional examples of data quality measures can be found in ISO 19160-3:2020 Annex C [20].

Requirement 4:	https://standards.isotc211.org/19157/-1/1/req/content/additionalQualityMeasure
	New quality measure shall be defined as a specialisation of *QualityMeasure*, or one of its subtypes (see Clause 9) according to the rules for metadata extensions (ISO 19115-1:2014, Annex C) and rules for domain profiles of standard schemas (ISO 19109:2015, 8.3). If the quality measure is derived from a basic measure, the basic measure shall be referenced. If an item is adopted from an external source the reference to the external source shall be given. If a measure uses an auxiliary variable, a reference *MeasureParameter* shall be provided for the measure.

10. Data quality evaluation

10.1 The process for evaluating data quality

10.1.1 Introduction

Quality evaluation processes are used in different phases of a product life cycle, having different objectives in each phase. The phases of the life cycle considered here are specification, production, delivery, use and update.

The process for evaluating data quality is a sequence of steps to produce a data quality result.

10.1.2 The process flow

The quality evaluation process is a sequence of steps followed to produce a quality evaluation result. Figure 11 illustrates a possible workflow for evaluating data quality; see also Annex D for a description of the concepts for evaluating and reporting data quality.

Recommendation 5:	https://standards.isotc211.org/19157/-1/1/rec/content/qualityEvaluationProcess
	When evaluated geographic dataset is heterogeneous and quality is specified for its individual parts, separate evaluation should be carried out for individual parts of the dataset.

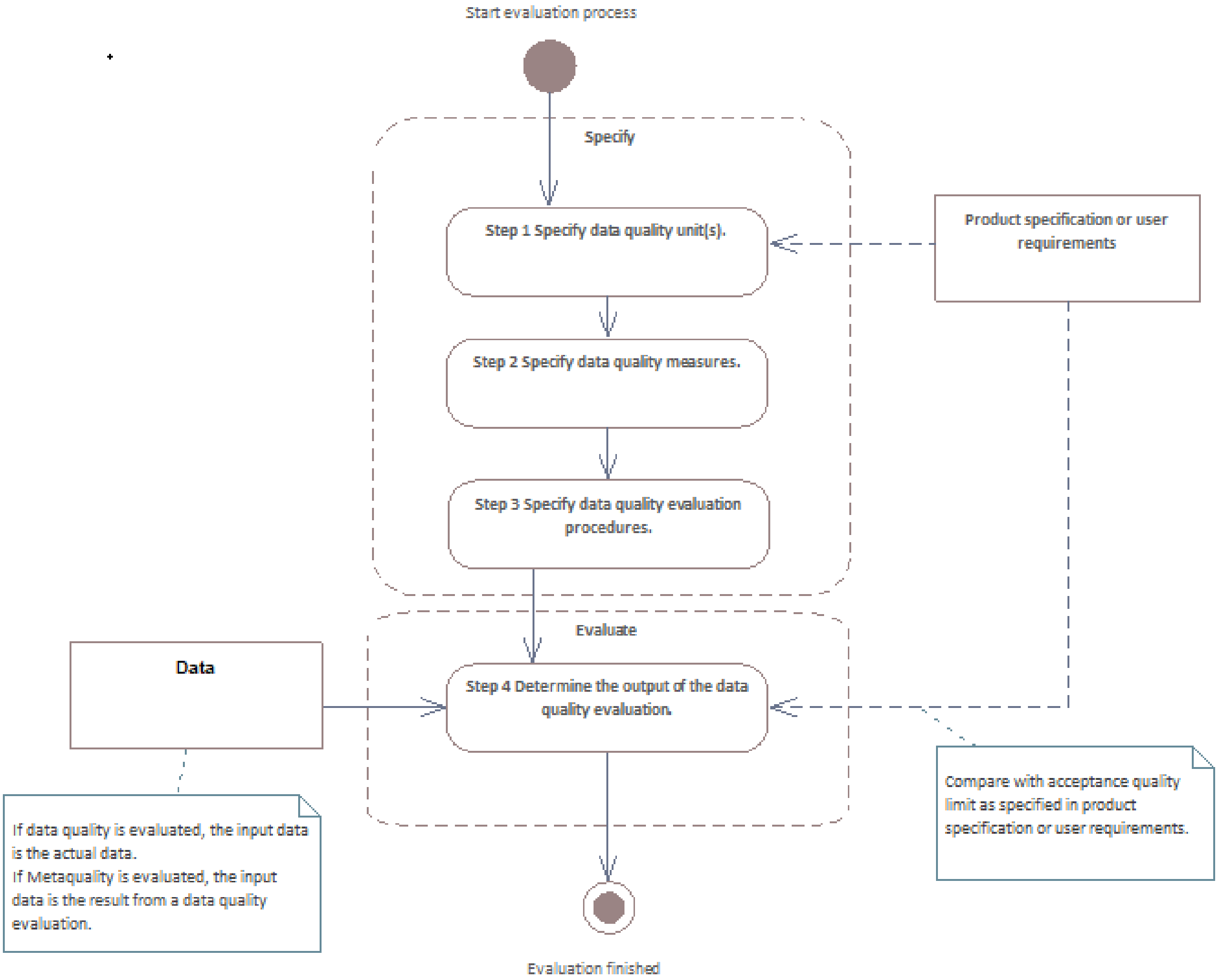


Figure 11 — Evaluating data quality

10.1.3 Process steps

Table 7 specifies the process steps involved in quality evaluation as illustrated in Figure 11.

Table 7 — Process steps

Process step	Action	Description
1	Specify data quality unit(s)	A data quality unit is composed of a scope and quality element(s), see 8.2. All data quality elements relevant to the data for which quality is to be described should be used. NOTE The data quality elements to be tested are described in 8.3, and Annex D provides guidelines for the use of quality elements. As described in 8.4 additional elements can be defined and used in quality evaluation.
2	Specify data quality measures	A measure should be specified for each data quality element. If no measure can be identified, a descriptive result may be provided
3	Specify data quality evaluation procedures	A data quality evaluation procedure consists of applying one or more evaluation methods. Including: – Standardized assessment method including conformance levels – Sampling strategy
4	Determine the output of the data quality evaluation	A result is the output of applying the evaluation. A conformance to data product specification or user requirements is part of quality evaluation output. The output is presented as metadata (e.g. in MD_Usage defined in ISO 19115-1) and a quality evaluation report.

Evaluation of metaquality may be performed after obtaining the output of the quality evaluation. The workflow described above is also a possible workflow for evaluating metaquality, with the following process steps: specify the metaquality element and the quality evaluation for which metaquality is evaluated, then specify a measure and an evaluation method and determine the output of the metaquality evaluation.

10.2 Data quality evaluation methods

10.2.1 Classification of data quality evaluation methods

A data quality evaluation procedure comprises one or more data quality evaluation methods. Data quality evaluation methods can be divided into two main classes: direct and indirect. Direct evaluation methods determine data quality through the comparison of the data with internal and/or external reference information. Indirect evaluation methods infer or estimate data quality using information on the data such as lineage.

Recommendation 6:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/evaluationMethodType
	Direct evaluation methods (*directInternal* or *directExternal*) should be used in preference to indirect evaluations.

The direct evaluation methods are further sub-classified by the source of the information needed to perform the evaluation, if internal or external. Figure 12 shows the classes used describing the evaluation methods.

NOTE Lineage is described in ISO 19115-1.

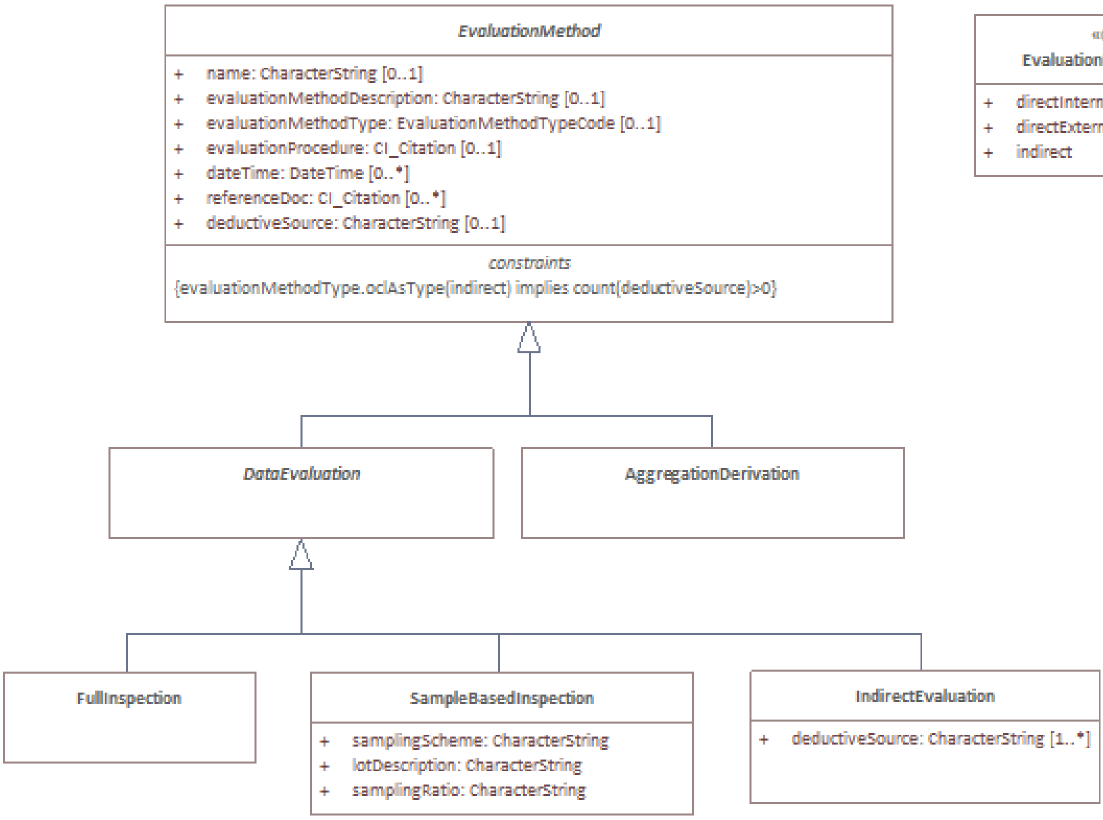


Figure 12 — Data quality evaluation methods

10.2.2 Direct evaluation

A direct evaluation method is a method of evaluating the quality of a dataset based on inspection of the items within the dataset.

The direct evaluation methods can be classified as internal or external. Internal direct data quality evaluation uses only data that resides in the dataset being evaluated. External direct quality evaluation requires reference data external to the dataset being tested.

NOTE 1 Reference data are data accepted as representing the universe of discourse.

For both external and internal direct evaluation methods, one of the following inspection methods may be used:

— full inspection: data quality evaluation tests every item specified by the data quality scope;

— sample-based inspection: data quality evaluation is performed on subsets of the geographic data defined by the data quality scope

If the evaluation is based on a sample-based inspection the sample strategy should be explained in the quality evaluation report.

The full inspection is preferable. In case full inspection is not feasible, the sample-based inspection can be used. Annex E discusses sampling methods in detail.

10.2.3 Indirect evaluation

An indirect evaluation method is a method of evaluating the quality of a dataset based on external knowledge or experience of the data product and can be subjective.

This external knowledge may include information about dataset's usage, lineage and purpose (see ISO 19115-1), or other data quality reports on the dataset or data used to produce the dataset. Data quality may be estimated, for example, from knowledge about the source, tools and methods used for the capturing of the data and evaluated against procedures and specifications worked out for this product. Indirectly evaluated data quality may also be based on experience alone or determined from the user feedback (for example, as reported in accordance with [36]).

In some cases it can potentially be misleading or not even possible to report indirectly evaluated data quality as quantitative results. In those cases the data quality may be described in textual form using a descriptive result; see 8.5.4.4.

10.3 Aggregation and derivation

Additional results may be produced by aggregating or deriving existing results without carrying out a new data quality evaluation.

Aggregation combines quality results from data quality evaluations based on different data quality elements or different data quality scopes.

Additional results may also be derived from existing results, for example, when a conformance result is obtained by comparing a quantitative result to a conformance level. This is useful e.g. if the result is expressed differently than the conformance level.

NOTE 1 Aggregation can be used to aggregate results of different data quality elements to describe the conformance to a data product specification.

NOTE 2 Aggregation is further described in Annex G. How to report aggregation is described in 11.2.1 and Annex D.

NOTE 3 How to report derivation is described in 11.2.2 and Annex D.

EXAMPLE If the result is expressed with a significance level of 95 % and the conformance level is expressed with a significance level of 99 %, the result could be recalculated to be of the same significance level as the conformance level.

11. Data quality reporting

11.1 General

Data quality shall be reported as metadata and may be complemented with a quality evaluation report.

Requirement 5:	https://standards.isotc211.org/19157/-1/1/req/content/dataQualityMetadata
	Data quality shall be reported as metadata in compliance with Clause 8, Clause 11 and Annex C of this standard, and in compliance with ISO 19115-1.

In order to provide more details than reported as metadata, and including details of the data quality evaluation, such as the specific process steps executed in the evaluation, a quality evaluation report may additionally be created. The metadata description includes a reference to the quality evaluation report when it exists (see Figure 13).

Requirement 6:	https://standards.isotc211.org/19157/-1/1/req/content/qualityEvaluationReport
	An optional quality evaluation report shall not replace the metadata.

Recommendation 7:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/reportReference
	In case a quality evaluation report exists for the reported data quality metadata, a reference to this report should be provided.

NOTE 1 See also B.4.3.2 for more information about how to report data quality and the complementary role between metadata and quality evaluation report.

NOTE 2 See E.4 for examples of how to report data quality.

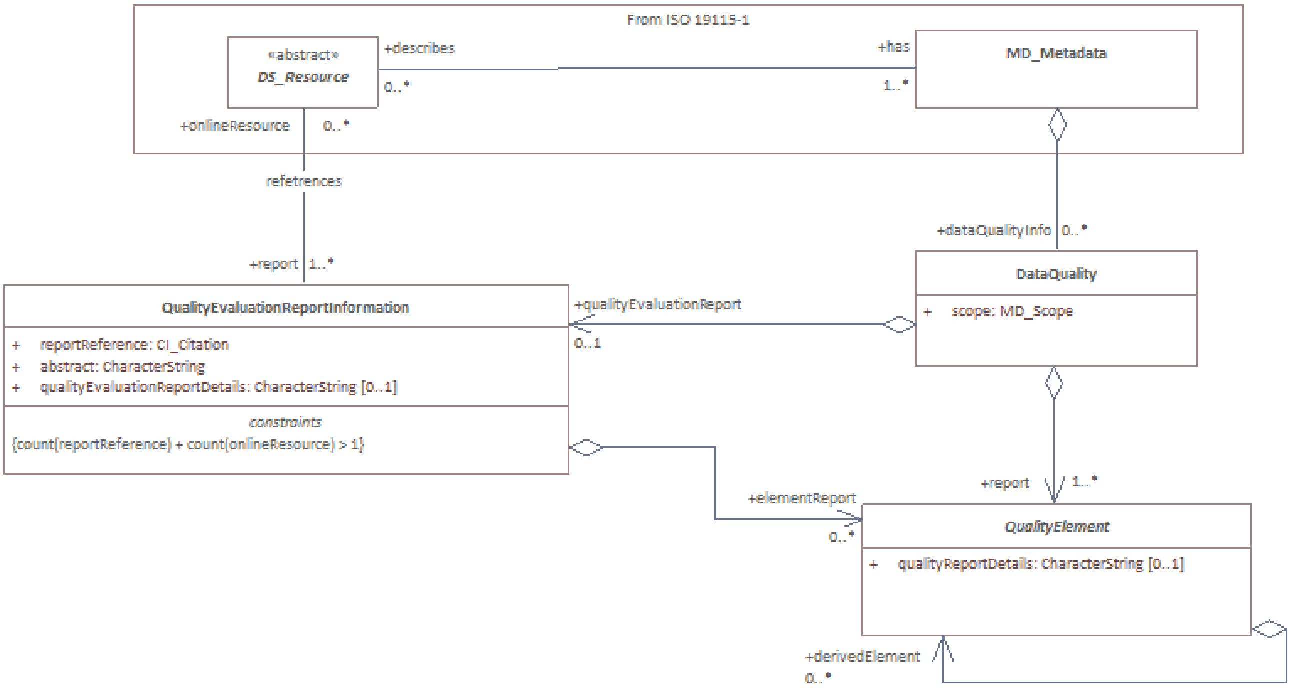


Figure 13 — Reporting data quality

11.2 Particular cases

11.2.1 Reporting aggregation (aggregated results)

A quality evaluation report may include information about aggregated results.

Recommendation 8:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/aggregatedResult
	Where the result has been aggregated, a quality evaluation report should be provided to complete the information provided in the metadata. This report includes fully detailed information on the original result (with measure(s) and evaluation procedure(s)), aggregated result and aggregation method.

Recommendation 9:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/sameQualityElement	
	When several quality results for the same data quality element are aggregated into a single result of this element, the result should be reported in metadata as a result for this data quality element. See D.4.1.2 and D.4.1.3 for examples.	

Recommendation 10:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/differentQualityElement
	When several quality results for different data quality elements are aggregated into a single result, this should be reported in metadata as a result of evaluation of a dataset's conformance with the data product specification or user requirements. See D.4.1.4 for an example.

Requirement 7:	https://standards.isotc211.org/19157/-1/1/req/content/aggregatedResult
	When aggregated result is reported in metadata a reference to the original quality evaluation results on individual element(s) shall be provided.

11.2.2 Reporting derivation (derived results)

When derived results are only reported in metadata, a quality evaluation report should also be generated to provide the original data quality results from which the derived result has been determined. The metadata should then provide the reference to the quality evaluation report and the original data quality result.

Recommendation 11:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/qualityEvaluationReport
	A quality evaluation report detailing the source of derivation (e.g. quality results, derivation method) should be generated for a derived result.

Recommendation 12:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/derivedResultMetadata
	Metadata on derived result should provide a reference to the quality evaluation report and the original data quality result.

Conformance result is often derived from a quantitative result. If only the conformance result is provided in metadata, then the quantitative results should be provided in a quality evaluation report.

11.2.3 Reference to the original data quality result

Permission 3.	https://standards.isotc211.org/19157/-1/1/per/permittedContent/sourceReference
	When derived or aggregated result(s) are reported in metadata, the reference to the original data quality result may be provided using two attributes: — The attribute *derivedElement* references a quality element [and its result(s)] described in the metadata; — The attribute *qualityEvaluationReportDetails* references the part of the quality evaluation report where the original result(s) are described.

11.2.4 Hierarchy principle

This document recognizes the principle of the hierarchical level:

Data quality specified at upper level (e.g. series) is applicable at lower level (e.g. dataset), see Table 8.

Table 8 — Hierarchical levels

Upper level ↑	Series	
	Dataset	
	Subset	
↓	Feature type	Attribute type
Lower level	Feature instance	Attribute instance

NOTE Quality for an instance of a feature, a feature attribute or associations between features can be reported as an attribute for that instance as defined in ISO 19109.

Recommendation 13:	https://standards.isotc211.org/19157/-1/1/rec/optionalContent/hierarchy
	If the data quality differs between upper and lower level, then supplemental information should be provided at lower level.

12. Requirements for XML encoding

The exchange of data quality specification in XML format requires encoding rules. This clause specifies such rules and an XML schema. The requirements class for the XML encoding is documented in Table 3.

Requirement 8:	https://standards.isotc211.org/19157/-1/1/req/xml
	A data quality XML document shall be conformant with the XML schema https://schemas.isotc211.org/19157/-1/dqc/1.0.0/dqc.xsd.

The XML schema follows the rules stated in ISO/TS 19139-1 for translation of UML models to XML schema.

The XML schema also uses the patterns for decoupling XML namespaces outlined in ISO/TS 19115-3:2016, Clause 8.

The XML schema definitions pertain to the following namespace: https://schemas.isotc211.org/19157/-1/dqc/1.0. This namespace is abbreviated dqc.

This XML schema implements all the UML classes defined in this document and imports all relevant classes from other International Standards.

More details regarding the XML encoding can be found in Annex H.

Annex A
(normative)

Abstract test suite

A.1 Content of a data product specification

Conformance test	**https://standards.isotc211.org/19157/-1/1/conf/content/allContent**
Reference	All normative statements in requirements class https://standards.isotc211.org/19157/-1/1/req/content (see Table 4 in Clause 6)
Test purpose:	Verify that the content of a data quality definition conforms to the UML model and additional requirements related to the elements in the UML model.
Test method:	Identify the individual object and their attributes and associations to other objects. Verify that each such element is in conformance with multiplicity and data type as expressed in the UML model. For each individual object check for additional requirements beyond the semantics in the model. Verify that each such element meets specified criteria.
Test type:	Basic

A.2 XML encoding

Conformance test	**https://standards.isotc211.org/19157/-1/1/conf/xml/xmlEncoding**
Reference	All normative statements in requirements class: https://standards.isotc211.org/19157/-1/1/req/xml (see Clause 12 and Annex H)
Test purpose:	Verify the syntax of a data quality definition from the XML document.
Test method:	Validate the XML document using the XML schema https://schemas.isotc211.org/19157/-1/dqc/1.0.0/dqc.xsd.
Test type:	Basic

Annex B
(informative)

Data quality concepts and their use

B.1 Framework of data quality concepts

A dataset may be produced for a specific application or for a set of presupposed applications. The quality of a dataset can only be assessed by knowledge about its data quality elements and, for some cases, indirectly by its non-quantitative quality information usage, lineage and purpose (see ISO 19115-1). The data quality elements evaluate the difference between the dataset and the universe of discourse (i.e. the perfect dataset that corresponds to the data product specification). The non-quantitative quality information provides general information from which quality-related knowledge may be derived.

Data quality concepts provide an important framework for data producers, as well as, for data users. A data producer is given the means for validating how well a dataset reflects its universe of discourse as defined in the data product specification. Data users can assess the quality of a dataset to ascertain if it is able to satisfy the requirements of the data user's application (see Figure B.1).

It should be noted that quality results reported are valid against the data product specification or the user requirements used. Care should be taken when comparing quality results where the universe of discourse (i.e. the specification or user requirements) is different. If these are changed, then quality evaluation should be repeated against the changed specification or requirements.

EXAMPLE Typical examples of this are related to model transformation in Spatial Data Infrastructures or generalization. For example, if the geometry of a feature type is changed, then positional accuracy results are changed as well.

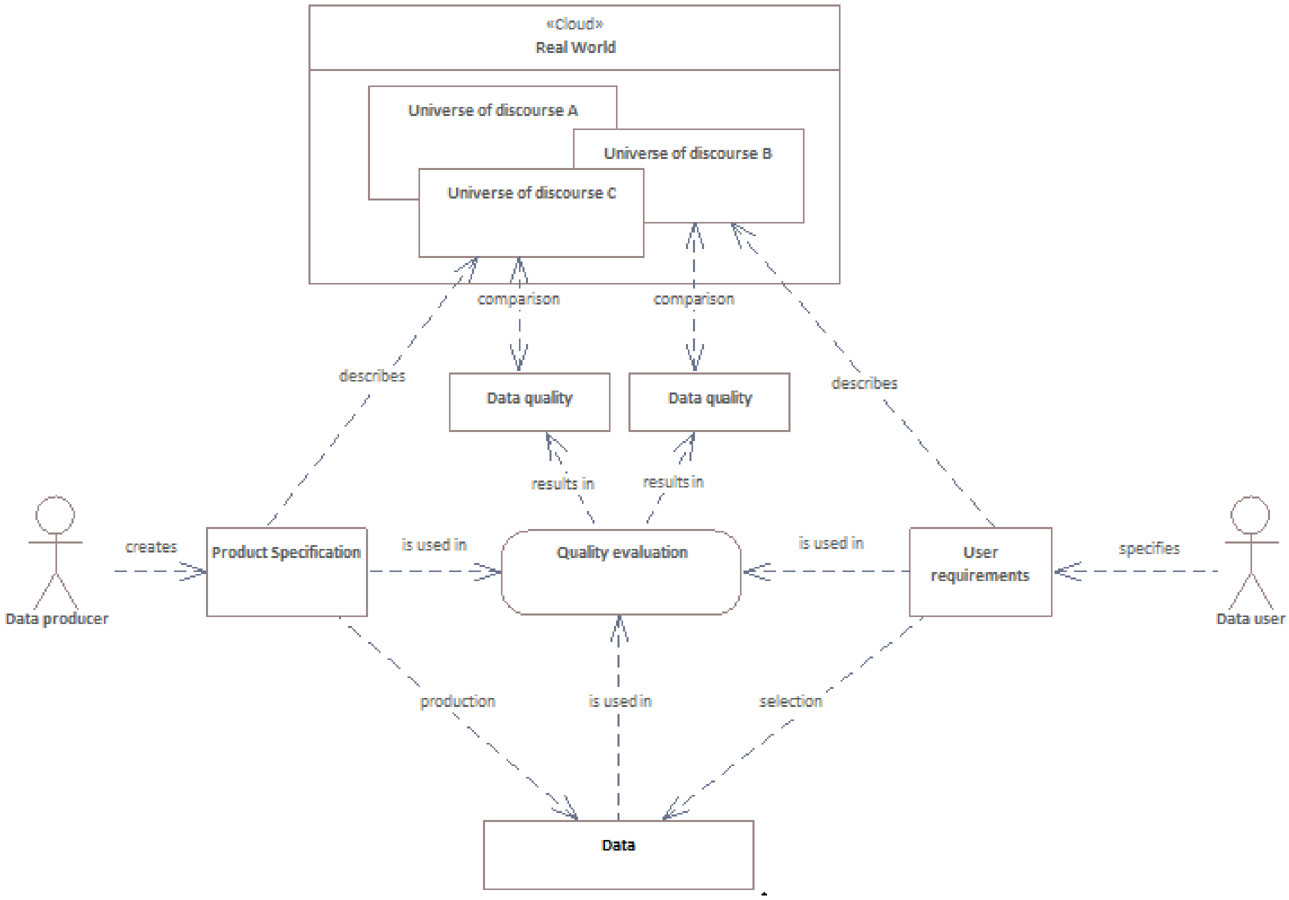


Figure B.1 — Framework of data quality concepts

B.2 The structure of datasets and components for quality description

A dataset may belong to a dataset series meaning that all of the series' datasets are based on the same data product specification. The quality of all member datasets belonging to a dataset series may be the same.

A dataset can be viewed as containing a large but finite number of subsets of data. Subsets of data which share a commonality such as belonging to the same feature type, feature attribute or feature relationship or sharing a collection criteria or geographic or temporal extent do often have similar quality. A subset of data can be as small as a feature instance, attribute value or occurrence of a feature relationship and, theoretically, data quality concepts allow each feature instance, attribute value and occurrence of a feature relationship of a dataset to have its own quality. The quality of subsets of data within a dataset cannot be assumed to be the same as the quality of other parts of the dataset to which they belong. Data quality concepts allow for reporting the quality of a dataset and, additionally, the differing quality of subsets of this dataset by identifying these groupings as the data specified by data quality scopes. The quality information reported for multiple data quality scopes smaller than the whole dataset for which quality is reported, provide a more complete and detailed picture of quality than the overall quality for the total dataset.

For a data producer, a data product specification describes a universe of discourse and contains the rules for constructing a dataset. For a data user, user requirements describe a universe of discourse, which may or may not match a dataset's universe of discourse. The quality of a dataset is how well it represents a universe of discourse. The quality of the same dataset can therefore differ depending on which universe of discourse it is evaluated against.

The quality of a dataset is described by data quality elements and their descriptors. Some quality related information may also be provided by the non-quantitative elements usage, lineage and purpose. Metaquality provides quality information about quality evaluation.

Data quality elements allow for the evaluation of how well a dataset meets the criteria set forth in its data product specification or user requirements. Data quality elements can be evaluated in various ways and at different stages of a dataset's life cycle. Not all data quality elements are applicable to all types of datasets. Some data quality elements are applicable to larger datasets, while others are more suitable for subsets of data within a larger dataset. Some data quality elements are applicable for single instances of data as well as for larger numbers while some only are applicable for multiple instances.

This document identifies data quality elements primarily as a means of identifying and reporting separate categories of quality information. It additionally recognizes that data quality elements frequently are interrelated. For example, a coordinate error may generate at least two kinds of errors, a positional error and a topological error; see Annex G. The meaning of the data quality elements in terms of the product and manner in which the data quality elements are handled are the responsibility of the quality evaluator.

All quality elements may be used to evaluate the conformance of the dataset against data product specification or a set of user defined requirements. The evaluation may be based on a specific application or user requirements that cannot be described using the quality elements described above. In this case, components defined in other standards from the series (e.g. LI_Lineage from ISO 19115-1) can be used, or new elements can be defined according to the requirements for new quality elements defined in 8.4. Likewise, existing standard elements defined elsewhere may be referenced.

Data producers can demonstrate how a dataset is suitable for various identified usages. This can be done as free text in a dataset's identification (in the *abstract* of the MD_Identification class from ISO 19115-1:2014) or, with MD_Usage class from ISO 19115-1, when a more detailed and structured explanation is needed.

Requirements for a dataset's fitness for use in the specific dataset's usage as defined by the producer, can be provided in a data product specification by defining conformance quality levels for values of relevant quality elements.

EXAMPLE The property registers and the corresponding digital property map is suitable for property formation in the local reference system inside urbanised areas.

local positional accuracy real property boundary:	0,05 m
external positional accuracy real property boundary:	0,2 m
omission joint property units:	< 1 %

B.3 When to use quality evaluation procedures

Quality evaluation procedures may be used in different phases of a product's life cycle. The stages of a product's life cycle during which quality evaluation may be applied are as follows:

— Development of a data product specification or user requirements: When developing a data product specification or defining user requirements, quality evaluation procedures may be used to facilitate the establishment of conformance quality levels that should be met by the final product. A data product specification or user requirements may include conformance quality levels for the data and quality evaluation procedures to be applied during production and updating.

— Quality control during dataset creation: At the production stage, the producer may apply quality evaluation procedures, either explicitly established or not contained in the data product specification, as part of the process of quality control. The description of the applied quality evaluation procedures, when used for production quality control, may be reported as lineage metadata including, but not necessarily limited to, the quality evaluation procedures applied, conformance quality levels established and the results.

— Inspection for conformance to a data product specification: On completion of the production, a quality evaluation process may be used to produce and report data quality results. These results may be used to determine whether a dataset conforms to its data product specification or not. If the dataset passes inspection (composed of a set of quality evaluation procedures), the dataset is considered to be ready for use. The results of the inspection operation should be reported in accordance with Clause 11. See also the example in Annex D describing evaluation and reporting of data quality. The outcome of the inspection will be either acceptance or rejection of the dataset. If the dataset is rejected, then, after the data has been corrected, a new inspection will be required before the product can be deemed to be in conformance with the data product specification.

— Evaluation of dataset conformance to user requirements: Quality evaluation procedures may be used to establish if a dataset meets the conformance quality levels specified in user requirements. Indirect as well as direct methods may be used in analyses of dataset conformance to user requirements.

— Quality control during dataset update: Quality evaluation procedures are applied to dataset update operations, both to the items being used for update and to benchmark the quality of the dataset after an update has occurred.

B.4 Reporting quality information

B.4.1 Why report data quality

The need to report data quality exists for a number of reasons including the following:

— to aid discovery and encourage use of the dataset;

— to demonstrate the compliance to a data product specification or to user requirements;

— as part of supplier management initiatives;

— to permit downstream judgements about the quality of information derived from the dataset;

— to permit rational (optimal) decision-making when it is known that all data contains imperfections.

B.4.2 When to report quality information

Datasets are continually being created, updated and merged with the result that the quality or a component of the quality of a dataset may change. The quality of a dataset can be affected by three conditions:

— when any quantity of data is deleted from, modified or added to a dataset,

— when a dataset’s data product specification is modified or new user specified data quality requirements are identified,

— when the real world has changed.

The first condition, a modification to a dataset, may occur frequently. Many datasets are not static. There is an increase in the interchange of information, the use of datasets for multiple purposes and an accompanying update and refinement of datasets to meet multiple purposes. If the reported quality of a dataset is likely to change with modifications of the dataset, the quality of this dataset should be reassessed and updated as required when changes occur.

Complete knowledge of all applicable data quality elements should be available when a dataset is created. Only the data producer's usage (assuming the data producer actually uses the dataset) of a dataset can initially be reported. There is a reliance on data users to report uses of a dataset that differ from its intended purpose so that continual updates to this particular data quality overview element can be made to reflect occurring, unforeseen uses.

The second condition, a modification to a dataset's data product specification, is most likely to occur before initial dataset construction and prior to the release of quality information. It is conceivable, however, that as a dataset is used, and its data product specification is updated so that future modifications to the dataset will better meet the actual needs. As the data product specification changes, the quality of the current dataset also changes. The quality information for a dataset should always reflect the current dataset given its current data product specification.

The third condition, a change of the real world, occurs continuously. Changes may be caused by natural phenomena such as movements in the earth's crust or erosion, but it is most often a result of human activity. Changes are often very rapid and dramatic. For this reason, the date of data collection is equally important as the date of quality evaluation when judging the quality of a dataset. In some cases, when known, even the rate of change is of interest. The update frequency of the dataset may also be of interest in some cases. However, this document recognizes that it might not be possible to create a new data quality report every time the real world changes.

B.4.3 How to report quality information with metadata and quality evaluation report

B.4.3.1 General

Quality information may be reported as metadata and as a quality evaluation report. These two mechanisms complement each other by allowing the reporting of data quality evaluation with different levels of detail:

— The metadata aims at providing short, synthetic and generally-structured information to enable metadata interoperability and web services usage;

— The quality evaluation report may be used to provide fully detailed information about the data quality evaluation. The quality evaluation report is to be referenced from the metadata as explained in Clause 11.

For example, in the case of aggregation of different quality results, the quality evaluation report will provide full information on the original results (with evaluation procedures and measures applied), the aggregated result and the aggregation method whereas the metadata may describe only the aggregated result with a reference to the original results described in the quality evaluation report.

B.4.3.2 Reporting quality information as metadata

The class MD_Metadata, defined in ISO 19115-1, aggregates from zero to many data quality units (instances of the class DataQuality, as specified in this document); see Figure B.2.

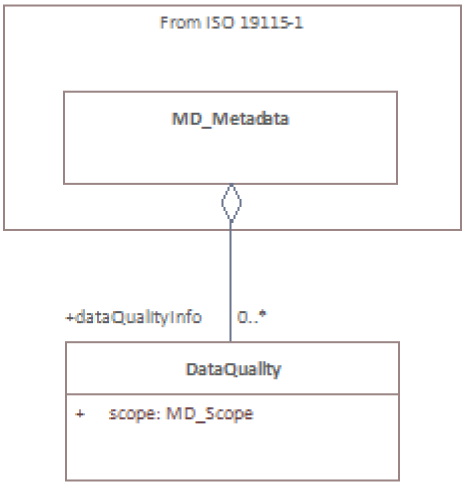


Figure B.2 — Data quality information

B.4.3.3 Reporting quality information within a quality evaluation report

The standardization of terminology (e.g. the data quality elements) and structure of the underlying data quality information will be of benefit to users familiar with the standard and facilitate better understanding and comparison. Further, a statement of compliance to the standard within the report may be of value to users.

A quality evaluation report should contain a scope to easily identify the extent to which the report covers the dataset under evaluation.

Each report should contain sufficient information to meaningfully describe the relevant aspects of data quality and their results. This may take the form of references to supporting documentation such as a data product specification or measure register.

The full structure of this quality evaluation report has intentionally not been standardized so that each particular organization is able to adapt it for its own needs, practices and evaluation procedures. The report may be in free text. However, the amount of quality information may be important. It is therefore important to present the report in a succinct, easily understood and easily retrievable way. For example, a report may follow the structure described in this document. An example of a quality evaluation report is provided in Annex D.

Annex C
(normative)

Data dictionary for data quality

C.1 Data dictionary overview

C.1.1 Introduction

This data dictionary describes the characteristics of the data quality model defined in Clauses 8, 9, 10 and 11. The dictionary is specified in tables with columns and rows in a hierarchy to establish relationships and an organization for the information.

The shaded table rows represent classes. The unshaded table rows represent class attributes and associations. The classes and class attributes within the data dictionary tables are defined by six table columns described in C.1.2 to C.1.7.

C.1.2 Name/role name

A label assigned to class or class attribute. Class names start with an upper-case letter. Spaces do not appear in a class name. Instead, multiple words are concatenated, with each new sub word starting with a capital letter (example: XnnnYmmm). Class names are unique within the entire data dictionary of this document. Class attribute names are unique within a class, not the entire data dictionary of this document. Class attribute names are made unique, within an application, by the combination of the class name and class attribute names. Role names are used to identify abstract model associations and are preceded by "Role name": to distinguish them from other class attributes. Names and role names may be in a language other than that used in this document.

C.1.3 Definition

This is the class or class attribute description.

C.1.4 Obligation/Condition

C.1.4.1 General

This is a descriptor indicating whether a class or class attribute shall always be documented in the dataset or sometimes be documented [i.e. contains value(s)]. This descriptor may have the following values: M (mandatory), C (conditional), or O (optional).

C.1.4.2 Mandatory (M)

The class or class attribute shall be documented.

C.1.4.3 Conditional (C)

Specifies an electronically manageable condition under which at least one class, class attribute or association is mandatory. "Conditional" is used for one of the three following possibilities:

— Expressing a choice between two or more options. At least one option is mandatory and shall be documented.

— Documenting a class, class attribute or association if another class has been documented.

— Documenting a class attribute or association if a specific value for another class attribute has been documented. To facilitate reading by humans, the specific value is used in plain text. However, the code shall be used to verify the condition in an electronical user interface.

If the answer to the condition is positive, then the class, class attribute or association shall be mandatory.

C.1.4.4 Optional (O)

The class, class attribute or association may or may not need to be documented. Optional class or optional class attribute have been defined to provide a guide to those looking to fully document their data. (Use of this common set of defined elements will help promote interoperability among geographic data users and producers world-wide.) If an optional class is not used, the class attributes contained within that class (including mandatory attributes) will also not be used. Optional classes may have mandatory class attributes; those class attributes only become mandatory if the optional class is used.

C.1.5 Maximum occurrence

Specifies the maximum number of instances the class, class attribute or association may have. Single occurrences are shown by “1”; repeating occurrences are represented by “N”. Fixed number occurrences other than one are allowed, and will be represented by the corresponding number (i.e. “2”, “3”…etc).

C.1.6 Data type

Specifies a set of distinct values for representing the class attributes; for example, integer, real, string, DateTime, and Boolean. The data type column is also used to define classes, stereotypes, and class associations.

NOTE Data types are defined in ISO 19103:2015, 6.5.2.

C.1.7 Domain

For a class (shaded rows), the domain indicates the line numbers covered by class attributes and associations for that class.

For a class attribute or association, the domain specifies the values allowed. The use of free text indicating unrestricted textual information used for the content of the field, or “unspecified domain” which may be any alphanumeric set of characters.

NOTE Domains are defined in ISO 19115-1:2013, B.1.7.

C.2 Data quality package data dictionary

C.2.1 Data quality

C.2.1.1 General

The global UML model for the whole data quality package is shown in Figure 2.

Table C.1 describes the properties of the UML model shown in Figure 3 and Figure 13.

Table C.1 — Data quality

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
1	DataQuality	Degree to which a set of inherent characteristics of data fulfils requirement.	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (MD_Metadata)	Lines 2–4
2	Scope	The specific data to which the data quality information applies	M	1	Class	MD_Scope << DataType >> (ISO 19115-1)
3	*Role name:* report		M	N	Association	QualityElement (Abstract class) (C.2.1.2)
4	*Role name:* QualityEvaluationReport		O	1	Association	QualityEvaluationReportInformation (C.2.1.6)

C.2.1.2 Data quality element

Table C.2 describes the UML model shown in Figure 4, Figure 5, Figure 10 and Figure 13.

Table C.2 — Data quality element

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
5	*QualityElement*	Aspect of quantitative quality information	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (DataQuality) (Abstract class)	Lines 6–10
6	QualityEvaluationReportDetails	Clause in the QualityEvaluationReport where this data quality element or any related data quality element (original results in case of derivation or aggregation) is described	O	1	Character string	Free text
7	*Role name:* measure	Reference to measure used	M	1	Association	MeasureReference (C.2.1.3)
8	*Role name:* evaluationMethod	Evaluation information	M	1	Association	EvaluationMethod (C.2.1.4)
9	*Role name:* result	Value (or set of values) obtained from applying a data quality measure or the outcome of evaluating the obtained value (or set of values) against	M	N	Association	QualityResult (Abstract class) (C.2.1.5)

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
		a specified acceptable conformance quality level				
10	*Role name:* derivedElement	In case of aggregation or derivation, indicates the original element	O	N	Association	QualityElement (Abstract class) (C.2.1.2)
11	*Completeness*	Presence and absence of features, their attributes and their relationships	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (QualityElement) (Abstract class)	Lines 6–10
12	Commission	Excess data present in the dataset, as described by the scope	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Completeness)	Lines 6–10
13	Omission	Data absent from the dataset, as described by the scope	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Completeness)	Lines 6–10
14	*LogicalConsistency*	Degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (QualityElement) (Abstract class)	Lines 6–10
15	ConceptualConsistency	Adherence to rules of the conceptual schema	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Logical Consistency)	Lines 6–10
16	DomainConsistency	Adherence of values to the value domains	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Logical Consistency)	Lines 6–10
17	FormatConsistency	Degree to which data are stored in accordance with the physical structure of the dataset, as described by the scope	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Logical Consistency)	Lines 6–10
18	TopologicalConsistency	Correctness of the explicitly encoded topological characteristics of the	Use obligation from	Use maximum occurrence from	Specified Class (Logical	Lines 6–10

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
		dataset as described by the scope	referencing object	referencing object	Consistency)	
19	*PositionalAccuracy*	Accuracy of the position of features	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (QualityElement) (Abstract class)	Lines 6–10
20	AbsolutePositionalAccuracy	Closeness of reported coordinate values to values accepted as true in a standard coordinate reference system	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Positional Accuracy)	Lines 6–10
21	RelativePositionalAccuracy	Closeness of the relative positions of features in the scope to their respective relative positions accepted as true in a local coordinate reference system	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Positional Accuracy)	Lines 6–10
22	GriddedDataPositionalAccuracy	Closeness of gridded data position values to values accepted as true	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Positional Accuracy)	Lines 6–10
23	*ThematicAccuracy*	Accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (QualityElement) (Abstract class)	Lines 6–10
24	ThematicClassificationCorrectness	Comparison of the classes assigned to features or their attributes to a universe of discourse	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Thematic Accuracy)	Lines 6–10
25	NonQuantitativeAttributeCorrectness	Correctness of non-quantitative attributes is correct or incorrect	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Thematic Accuracy)	Lines 6–10
26	QuantitativeAttributeAccuracy	closeness of the value of a quantitative attribute to a value accepted as or known to be true	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Thematic Accuracy)	Lines 6–10

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
27	*TemporalQuality*	Accuracy of the temporal attributes and temporal relationships of features	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (QualityElement) (Abstract class)	Lines 6–10
28	AccuracyOfATimeMeasurement	closeness of reported time measurements to values accepted as true	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Temporal Quality)	Lines 6–10
29	TemporalValidity	Validity of data with respect to time	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Temporal Quality)	Lines 6–10
30	*Metaquality*	Information about the reliability of data quality results	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (QualityElement) (Abstract class)	Lines 31 and 6–10
31	*Role name:* relatedQuality Element	Related quality element	M	1	Association	QualityElement (Abstract class) (C.2.1.2)
32	Confidence	Trustworthiness of a data quality result	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Metaquality)	Lines 31 and 6–10
33	Representativity	Degree to which the sample used has produced a result which is representative of the data within the data quality scope	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Metaquality)	Lines 31 and 6–10
34	Homogeneity	Expected or tested uniformity of the results obtained for a data quality evaluation	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Metaquality)	Lines 31 and 6–10

C.2.1.3 Measure reference

Table C.3 describes the UML model shown in Figure 6.

Table C.3 — Measure reference

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
35	MeasureReference	Reference to the measure used	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (Element)	Lines 36-38
36	measureIdentification	Identifier of the measure, value uniquely identifying the measure within a namespace	O	1	Class	MD_Identifier << DataType >> (see ISO 19115-1:2014, Table B.17.2)
37	nameOfMeasure	Name of the test applied to the data	C/ if measureIdentification not documented	N	Character string	Free text
38	measureDescription	Description of the measure	C/ if measureIdentification not documented	1	Character string	Free text

C.2.1.4 Data quality evaluation

Table C.4 describes the UML model shown in Figure 7, Figure 11 and Figure 12.

Table C.4 — Data quality evaluation

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
39	*EvaluationMethod*	Description of the evaluation method and procedure applied	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (QualityElement)	Lines 40-46
40	name	Name of the evaluation method	O	1	CharacterString	Free text
41	evaluationMethodDescription	Description of the evaluation method	O	1	CharacterString	Free text
42	evaluationMethodType	Type of method used to evaluate quality of the data	O	1	Class	EvaluationMethodTypeCode << CodeList >> (C.3.2)
43	evaluationProcedure	Reference to the procedure information	O	1	Class	CI_Citation << DataType >> (see ISO 19115-1:2014, Table B.16)
44	dateTime	Date or range of dates on which a data quality measure was applied	O	N	Class	DateTime (see ISO 19103:2015)

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
45	referenceDoc	Information on documents which are referenced in developing and applying a data quality evaluation method	O	N	Class	CI_Citation << DataType >> (see ISO 19115-1:2014, Table B.16)
46	deductiveSource	Information on which data are used as sources in deductive evaluation method	M	1	CharacterString	Free text
47	*DataEvaluation*	Data evaluation method	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Evaluation Method) (Abstract class)	Lines 40-46
48	FullInspection	data quality evaluation tests every item specified by the data quality scope	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DataEvaluation)	Lines 40-46
49	SampleBasedInspection	data quality evaluation is performed on subsets of the geographic data defined by the data quality scope.	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DataEvaluation)	Lines 40-46 and 50-52
50	samplingScheme	Information of the type of sampling scheme and description of the sampling procedure	M	1	CharacterString	Free text
51	lotDescription	Information of how lots are defined	M	1	CharacterString	Free text
52	samplingRatio	Information on how many samples on average are extracted for inspection from each lot of population	M	1	CharacterString	Free text
53	AggregationDerivation	Aggregation or derivation method	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Evaluation Method)	Lines 40-46

C.2.1.5 Data quality result

Table C. 5 describes the UML model shown in Figure 8.

Table C.5 — Data quality result

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
54	*QualityResult*	Generalization of more specific result classes	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (Element) (Abstract class)	Lines 55-56
55	resultScope	Scope of the result	O	1	Class	MD_Scope (ISO 19115-1:2014)
56	dateTime	date when the result was generated	O	1	Class	DateTime (see ISO 19103:2015)
57	Quantitative Result	The values or information about the value(s) (or set of values) obtained from applying a data quality measure	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Result)	Lines 58-60 and 55-56
58	value	Quantitative value or values, content determined by the evaluation procedure used, accordingly with the value type and valueStructure defined for the measure	M	N	Class	Record (see ISO 19103:2015)
59	valueUnit	Value unit for reporting a data quality result	C / If value is described with a number	1	Class	UnitOfMeasure (see ISO 19103:2015)
60	valueRecord Type	Value type for reporting a data quality result, depends of the implementation	O	1	Class	RecordType << Metaclass >> (see ISO 19103:2015)
61	Conformance Result	Information about the outcome of evaluating the obtained value (or set of values) against a specified acceptable conformance quality level	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (QualityResult)	Lines 62-64 and 55-56
62	specification	Citation of data product specification or user requirement against which data are being evaluated	M	1	Class	CI_Citation << DataType >> (see ISO 19115-1:2014, Table B.16)
63	explanation	Explanation of the meaning of	O	1	CharacterString	Free text

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
		conformance for this result				
64	pass	Indication of the conformance result where 0 = fail and 1 = pass	M	1	Boolean	1 = yes 0 = no
65	DescriptiveResult	Data quality descriptive result	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (Result)	Lines 66 and 55-56
66	statement	Textual expression of the descriptive result	M	1	CharacterString	Free text
67	CoverageResult	Result of a data quality measure organizing the measured values as a coverage	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified class (Result)	Lines 68-71 and 55-56
68	spatialRepresentationType	Method used to spatially represent the coverage result	M	1	Class	MD_SpatialRepresentationTypeCode <<Code List>> (ISO 19115-1:2014 AMD1)
69	*Role name:* resultSpatialRepresentation	Provides the digital representation of data quality measures composing the coverage result	M	1	Association	MD_SpatialRepresentation <<Abstract>> (ISO 19115-1:2014 AMD1)
70	*Role name:* resultContent	Provides the description of the content of the result coverage when the quality coverage is included with the resource being described. i.e. semantic definition of the data quality measures	C / if resultFormat not provided?	N	Association	MD_RangeDimension (ISO 19115-1:2014)
71	*Role name:* resultFormat	Provides information about the format of the result coverage data	C / if resultContent not provided?	1	Association	MD_Format (ISO 19115-1:2014)

C.2.1.6 Quality evaluation report information

Table C.6 describes the UML model shown in Figure 13.

Table C.6 — Quality evaluation report information

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
72	QualityEvaluationReportInformation	Reference to an external quality evaluation report	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 73-77
73	reportReference	Reference to the associated quality evaluation report	C/ if onlineResource not provided	1	Class	CI_Citation << DataType >> (see ISO 19115-1:2014, Table B.16)
74	abstract	Abstract for the associated quality evaluation report	M	1	CharacterString	Free text
75	QualityEvaluationReportDetails	Reference to the original results in the quality evaluation report	O	N	CharacterString	Free text
76	Role name: elementReport	Data quality elements covered by the quality evaluation report	O	N	Association	QualityElement
77	Role name: onlineResource		C/ if reportReference not provided	N	Association	DS_Resource <<Abstract>> (see ISO 19115-1:2014)

C.2.2 Data quality measure

C.2.2.1 General

The UML model for measures information is shown in Figure 10.

C.2.2.2 Data quality measures

Table C.7 describes the UML model shown in Figure 10.

Table C.7 — Data quality measures

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
78	QualityMeasure	Data quality measure	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 79-90
79	measureIdentifier	Value uniquely identifying the measure within a namespace	M	1	Class	MD_Identifier << DataType >> (see ISO 19115-1:2014, Table B.17.2)

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
80	name	Name of the data quality measure applied to the data	M	1	CharacterString	Free text
81	alias	Another recognized name, an abbreviation or a short name for the same data quality measure	O	N	CharacterString	Free text
82	elementName	Name of the data quality element for which quality is reported	M	N	Class	TypeName << interface >> (see ISO 19103:2015)
83	definition	Definition of the fundamental concept for the data quality measure	M	1	CharacterString	Free text
84	description	Description of the data quality measure, including all formulae and/or illustrations needed to establish the result of applying the measure	M	N	Class	MeasureDescription << Datatype >> (C.2.2.5)
85	valueType	Value type for reporting a data quality result (shall be one of the data types defined in ISO 19103:2015)	M	1	Class	TypeName << interface >> (see ISO 19103:2015)
86	valueStructure	Structure for reporting a complex data quality result	O	1	Class	ValueStructure << CodeList >> (C.3.3)
87	example	Illustration of the use of a data quality measure	O	N	Class	MeasureDescription (C.2.2.5)
88	*Role name:* basicMeasure	Name of the data quality basic measure from which the data quality measure is derived	C/if derived from basic measure	1	Association	BasicMeasure (C.2.2.3)
89	*Role name:* sourceReference	Reference to the source of an item that has been adopted from an external source	C/if an external source exists	N	Association	SourceReference (C.2.2.6)
90	*Role name:* parameter	Auxiliary variable used by the data quality measure, including its name, definition and optionally its description	C/if required	N	Association	MeasureParameter (C.2.2.4)

C.2.2.3 Data quality basic measure

Table C.8 describes the UML model shown in Figure 10.

Table C.8 — Data quality basic measure

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
91	BasicMeasure	Data quality basic measure	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 92-95
92	name	Name of the data quality basic measure applied to the data	M	1	CharacterString	Free text
93	definition	Definition of the data quality basic measure	M	1	CharacterString	Free text
94	example	Illustration of the use of a data quality measure	O	1	Class	MeasureDescription << Data type >> (C.2.2.5)
95	valueType	Value type for the result of the basic measure (shall be one of the data types defined in ISO 19103:2015)	M	1	Class	TypeName << interface >> (see ISO 19103:2015)

C.2.2.4 Data quality parameter

Table C.9 describes the UML model shown in Figure 10.

Table C.9 — Data quality parameter

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
96	MeasureParameter	Data quality parameter	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 97-101
97	name	Name of the data quality parameter	M	1	CharacterString	Free text
98	definition	Definition of the data quality parameter	M	1	CharacterString	Free text
99	description	Description of the data quality parameter	O	1	Class	MeasureDescription << Data type >> (C.2.2.5)
100	valueType	Value type of the data quality parameter (shall be one of the data types defined in ISO 19103:2015)	M	1	Class	TypeName << interface >> (see ISO 19103:2015)
101	valueStructure	Structure of the data quality parameter	O	1	Class	ValueStructure << CodeList >> (C.3.3)

C.2.2.5 Data quality measure description

Table C.10 describes the UML model shown in Figure 10.

Table C.10 — Data quality measure descriptor

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
102	MeasureDescription	Data quality measure description	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 103-105
103	textDescription	Text description	M	1	CharacterString	Free text
104	extendedDescription	Illustration	O	1	Class	MD_BrowseGraphic (see ISO 19115-1:2014, Table B.17.3)
105	formula	Description of formulas for the quality measure	0	N	Class	FormulaType << Datatype >> (C.2.2.7)

C.2.2.6 Data quality measure source reference

Table C.11 describes the UML model shown in Figure 10.

Table C.11 — Data quality measure source reference

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
106	SourceReference	Reference to the source of the data quality measure	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Line 107
107	citation	Reference to the source	M	1	Class	CI_Citation << DataType >> (see ISO 19115-1:2014, Table B.16)

C.2.2.7 Data quality formula type

Table C.12 describes the UML model shown in Figure 10

Table C.12 — Data quality formula type

	Name / role name	Definition	Obligation / condition	Maximum occurrence	Data type	Domain
108	FormulaType	Description of the formula	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 109-112
109	key	Explanation of the formula.	M	1	CharacterString	
110	language	The language used to express the formula	M	1	Class	FormulaLanguage << CodeList >> (c.3.3)
111	languageVersion	The version of the language	M	1	CharacterString	
112	mathematicalFormula	The formula in chosen language	M	1	CharacterString	

C.3 Code lists

C.3.1 Introduction

The stereotype classes << CodeList >> can be found below. These stereotype classes do not contain "obligation/condition", "maximum occurrence", "data type" and "domain" columns. As a << CodeList >> is extendable, none of these stereotype classes contain a value such as "other".

C.3.2 Evaluation method type

Table C.13 describes the UML model shown in Figure 12

Table C.13 — Evaluation method type

	Name	Definition
1	EvaluationMethodTypeCode	type of method for evaluating an identified data quality measure
2	directInternal	method of evaluating the quality of a dataset based on inspection of items within the dataset, where all data required is internal to the dataset being evaluated
3	directExternal	method of evaluating the quality of a dataset based on inspection of items within the dataset, where reference data external to the dataset being evaluated is required
4	indirect	method of evaluating the quality of a dataset based on external knowledge

C.3.3 Value structure

Table C.14 describes the UML model shown in Figure 12

Table C.14 — Value structure

	Name	Definition
1	ValueStructure	
2	bag	finite, unordered collection of related items (objects or values) that may contain duplicates (ISO 19103:2015)
3	set	unordered collection of related items (objects or values) with no repetition (ISO 19103:2015)
4	sequence	finite, ordered collection of related items (objects or values) that may be repeated (ISO 19103:2015)
5	table	an arrangement of data in which each item may be identified by means of arguments or keys (ISO/IEC 2382-4:1999)
6	matrix	rectangular array of numbers (ISO/TS 19129:2009)
7	coverage	feature that acts as a function to return values from its range for any direct position within its domain (ISO/DIS 19123-1)

C.3.4 Formula language

Table C.15 describes the UML model shown in Figure 10.

Table C.15 — Formula language

	Name	Definition
1	FormulaLanguage	the coding language of the formula
2	MathJSON	a JSON-based format to represent math formulas (w3c.gituhub)
3	MathML	XML-based language for describing mathematical notation (mozilla.org)
4	OpenMath	extensible standard for representing the semantics of mathematical objects (openmath.org)

Annex D
(informative)

Evaluating and reporting data quality

D.1 Introduction

This annex provides one main example describing evaluation and reporting of data quality.

A few additional examples are provided in E.5, pointing to the metadata reporting of particular cases like descriptive result, metaquality and sampling evaluation.

D.2 Dataset description

D.2.1 Data product specification

D.2.1.1 General

The data product specification defined below describes the universe of discourse. The specification defines those features, attributes and relationships that are considered important and should be present in the dataset.

NOTE This is not a complete example of a data product specification (see ISO 19131).

The product will comprise transport network (paths and roads), buildings (houses and industrial buildings) and trees.

D.2.1.2 Feature Types

Each feature type, with zero or more attributes, is listed in Table D.1. Each attribute name is followed by a value type (e.g. code list, data type, character string or integer) and by an optional value domain.

Table D.1 — Feature types

	Feature type	Attribute name	Value type	Value domain
Buildings	Industrial building			
	House	family name	CharacterString	
		number of occupants	Integer	
Transport network	Path			
	Road	condition	CodeList: Pavement	surfaced, unsurfaced
	Tree	height	CodeList: TreeHeight	A: from 1 m to 3 m; B: from 3 m to 5 m; C: from 5 m to 10 m; D: more than 10 m.

D.2.1.3 Rules

The feature types in Table D. 1 shall adhere to the following rules:

— trees with a height of less than 1 m shall not be recorded;

— the attribute “condition” of a road may have no value (“undetermined value”);

— the attributes “name” and “number of occupants” of a house may have no value (“undetermined value”).

D.2.1.4 Conformance quality levels

Overall data quality requirement: a dataset shall pass only if it conforms to quality requirements as listed below.

a) Only feature types and attributes defined in this data product specification can be present in the dataset.

Transport network:

b) Max two items can be missing for each feature type.

c) Max two items can be in excess for each feature type.

d) Max two feature instances can be misclassified as another of the Transport Network feature type and zero as other feature types.

Buildings:

e) Max two items can be missing for each feature type.

f) Max two items can be in excess for each feature type.

g) Max two feature instances can be misclassified as another of the Building feature types and zero as other feature types.

Trees:

h) Max 10 % missing trees.

i) Max 10 % trees in excess.

j) Max 20 % of the trees can have wrong height.

k) No feature instances can be misclassified as other feature types.

D.2.2 Representation of the real world, the universe of discourse and the dataset

The relationship between the three figures is as follows:

a) Figure D.1 represents the “real world”, which generally contains more features than will be contained in the dataset;

b) Figure D.2 represents the “universe of discourse” given by the data product specification; it is that part of the real world that is to be included in the dataset, if the dataset is completely and accurately produced;

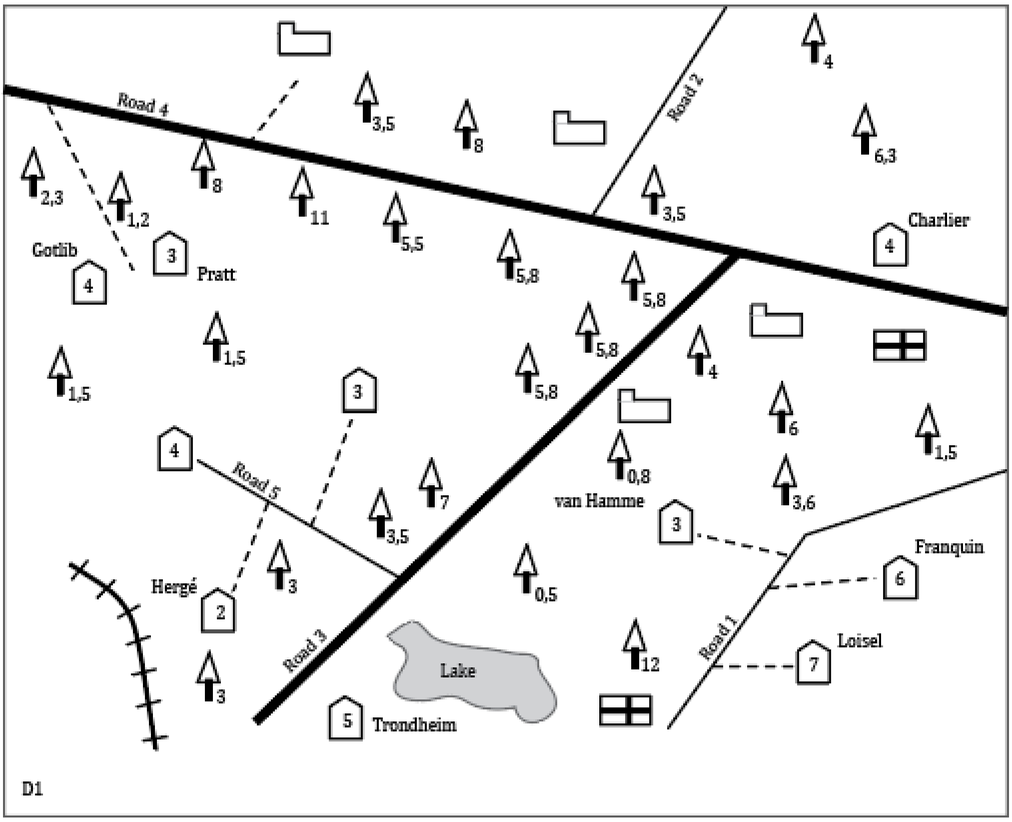
c) Figure D.3 represents the dataset as produced.

In all of the figures:

— the digit or letter representing domain of digits under the symbol of a tree is the height of the tree in metres,

— the digit in the symbol of a house is the number of occupants of the house,

— the name of the occupants of a house is noted beside the symbol of the house.



Key

hospital building

industrial building

tree with height of x metres

path

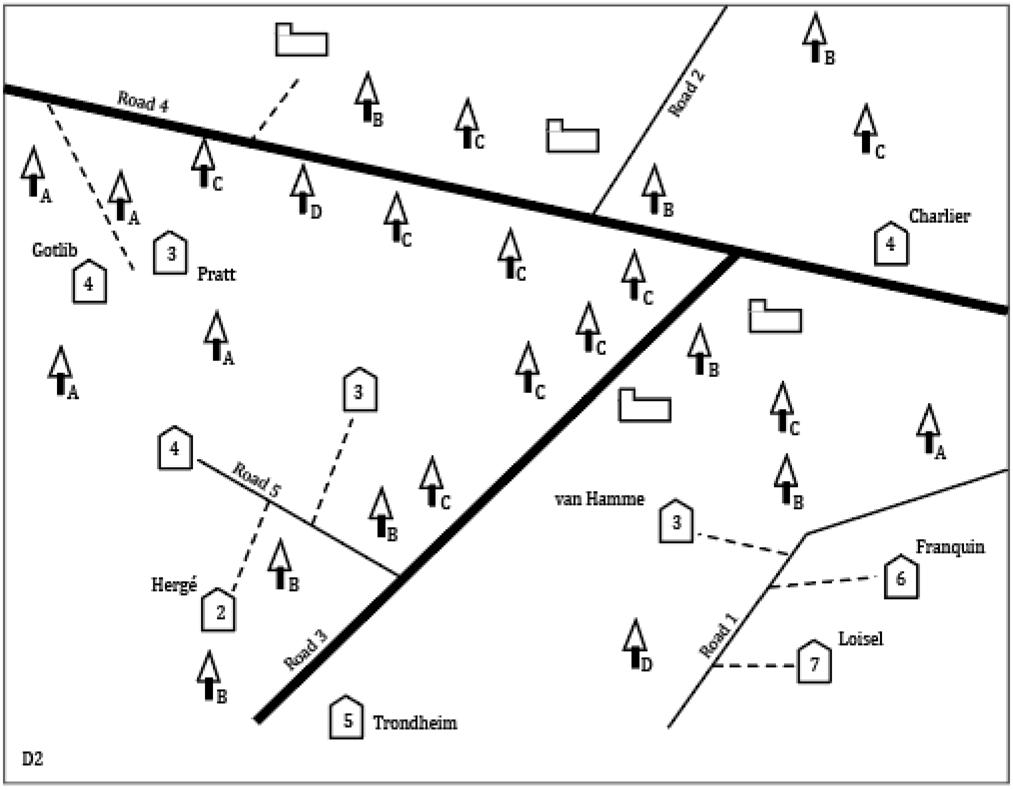
railway

house with x occupants

road: surfaced

road: unsurfaced

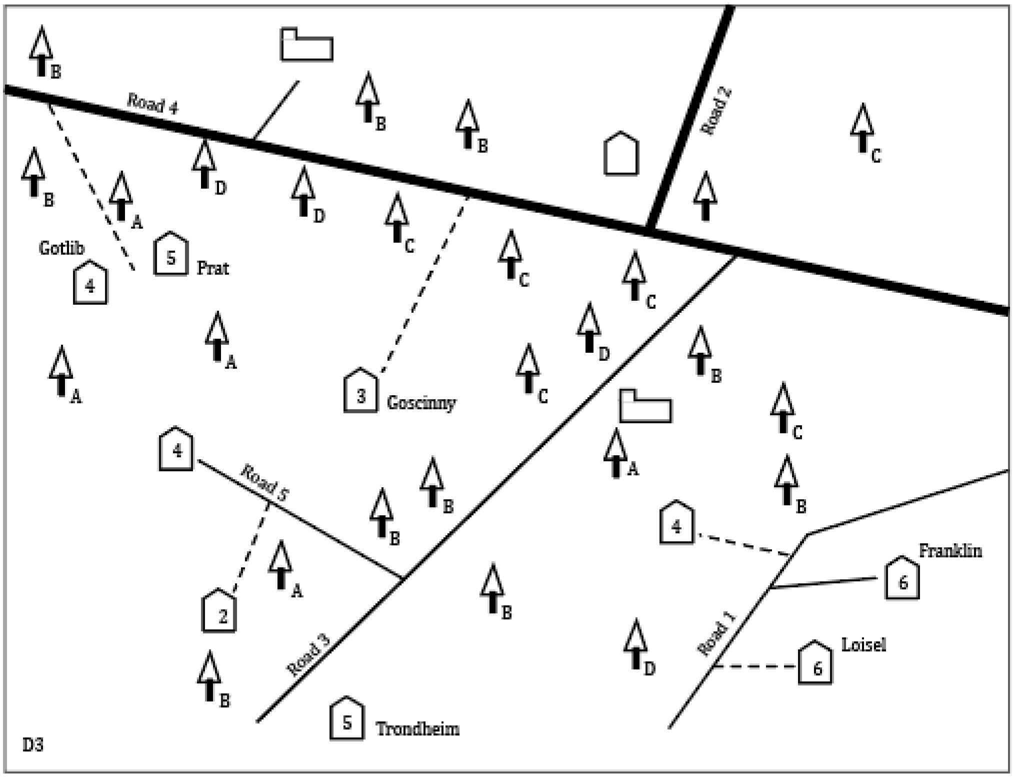
Figure D.1 — Graphical representation of the "real world"



Key

	industrial building		house with x occupants
	tree with height of 5 m to 10 m		road: surfaced
	path		road: unsurfaced

Figure D.2 — Graphical representation of the universe of discourse



Key

industrial building		house with x occupants	
tree with height of 5 m to 10 m		road: surfaced	
path		road: unsurfaced	

Figure D.3 — Graphical representation of the dataset

D.3 Quality evaluation process

D.3.1 Specify data quality unit(s)

A data quality unit is composed by a scope and quality element(s). In this example the completeness and thematic quality are evaluated to conform to the data product specification.

— The first quality unit is composed by conceptual consistency, completeness (commission and omission) and thematic classification correctness evaluated on the whole dataset.

— Two other quality units are composed by aggregated conceptual consistency, completeness (commission and omission) and thematic classification correctness evaluated on the transport networks and buildings.

— One quality unit is composed by quantitative attribute accuracy evaluated on feature type (tree).

Guidelines for choosing appropriate data quality elements are provided in Annex F.

D.3.2 Specify data quality measures

For describing logical consistency the following measure is used:

— Measure 9, “conceptual schema compliance”.

For describing completeness the following measures are used:

- Measure 1, “excess item”;
- Measure 2, “number of excess items”;
- Measure 3, “rate of excess items”;
- Measure 5, “missing item”;
- Measure 6, “number of missing items”;
- Measure 7, “rate of missing items”.

For describing thematic quality the following measure is used:

- Measure 62, “misclassification matrix”.

For describing overall conformance to the specification the following measure is used:

- Measure 101, “data product specification passed”.

D.3.3 Specify data quality evaluation procedures

For this example we use a direct external procedure.

Full inspection is used for this example.

NOTE An example of a sampling procedure is described in D.5.5.

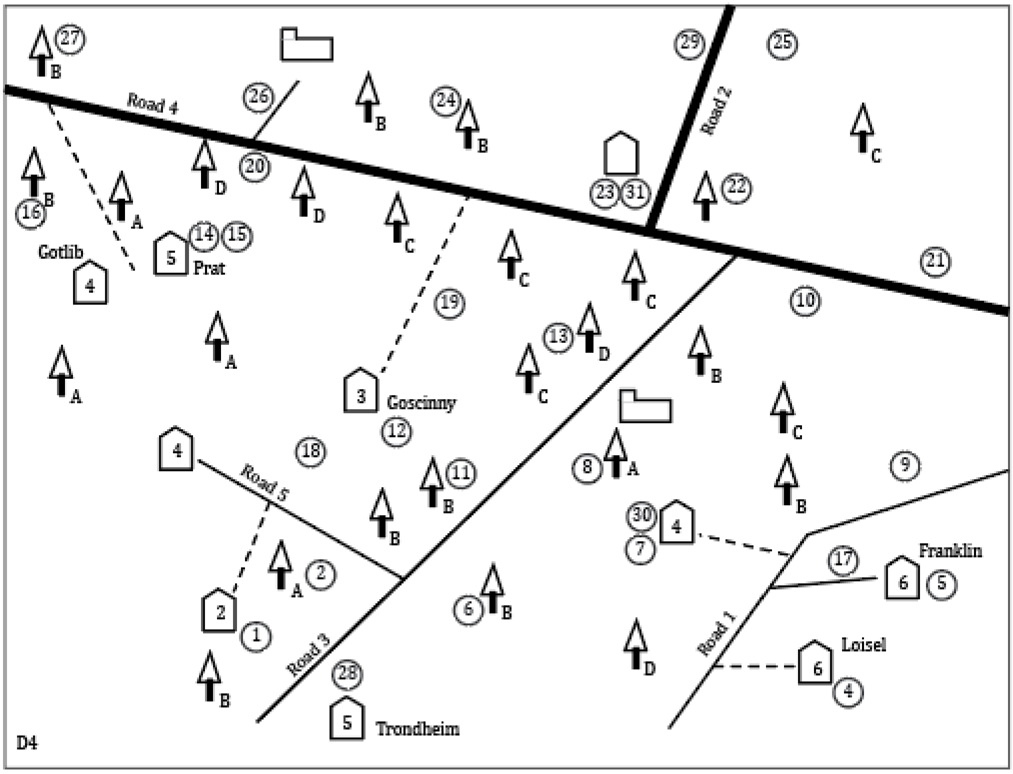
D.3.4 Specify conformance quality levels

The dataset is expected to be conformant to the data quality levels specified in D.2.1.4.

D.3.5 Determine the output of the data quality evaluation (Result)

D.3.5.1 Identification of errors

By comparing the dataset, represented by Figure D.3, with the universe of discourse, represented by Figure D.2, a list of errors in the example dataset can be produced, represented by Figure D.4.



Key

industrial building	error number
tree with height of 3,5 m	house with x occupants
path	road: surfaced
	road: unsurfaced

Figure D.4 — Graphical representation of dataset error locations

The following is a list of detected errors with error numbers given for reference.

— Errors of omission and commission in recording of trees. Three trees (No. 6, No. 8, No. 27) are in excess and two trees are missing (No. 9, No. 25).

— Errors of omission and commission in recording paths. One path is missing (No. 18) and one is in excess (No. 19).

— A house replaces an industrial building (No. 23).

— Two paths are miscoded as roads (No. 17, No. 26).

— A house is missing (No. 21).

— Attribute error on roads. Two roads have the wrong “condition” (No. 29, No. 28).

— Two trees with a height less than 1 m are represented in the dataset (No. 6, No. 8) as excess items.

— Tree height attribute class code missing. A tree is missing a class code while it is B in the universe of discourse (No. 22).

— Tree height attribute misclassified. Six trees have the wrong height class assigned (No. 2, No. 11, No. 13, No. 16, No. 20, No. 24).

— House name attribute “family name” errors. The houses named “van Hamme” (No. 7) and “Hergé” (No. 1) in the universe of discourse have no name in the dataset. The house named “Goscinny” in the dataset (No. 12) has no name in the universe of discourse.

— House name attribute “family name” errors. The houses named “Franquin” (No. 5) and “Pratt” (No. 15) in the universe of discourse are named “Franklin” and “Prat” respectively in the dataset.

— House occupant count attribute errors. The occupant count attribute is missing for one house (No. 31) and wrong for three houses (No. 4, No. 14, No. 30).

— Omission error in industrial buildings. One industrial building is missing (No. 10).

NOTE The classification of errors as omission/commission, completeness or thematic quality is subjective. For example, the misclassification of a house as an industrial building could alternately be considered as an error of omission of the one and commission of the other.

D.3.5.2 Logical consistency

Only feature types and attributes defined in the data product specification are present in the dataset. See the conformance result for conceptual consistency in Table D.2.

Table D.2— Conformance result for logical consistency

Scope	Quality element	Conformance requirements	Number of evaluations	Counts yes/no	Pass
Dataset	Conceptual consistency	1) Only feature types and attributes defined in the application schema can be present in the dataset.	1 (no errors detected)	1/0	Yes

D.3.5.3 Completeness

D.3.5.3.1 General

Completeness in this example is classified by feature class. The types of measures tested for are commission and omission. The results are shown in Tables D.3 to D.5.

D.3.5.3.2 Quantitative result

Table D.3 depicts a way to classify completeness using quantitative values.

Table D.3— Completeness by feature class

Feature class	Number of instances in the universe of discourse	Commission count	Commission percentage[a]	Omission count	Omission percentage[b]
Path	7	1	14	1	43
Road[c]	5	2	40	0	0
Tree	25	1	12	2	8
Industrial building	4	0	0	1	50
House	10	0	10	1	10

a Commission percentage = number of included items/number of items in the universe of discourse × 100.

b Omission percentage = number of omitted items/number of items in the universe of discourse × 100.

c If there is a small distinction between an unpaved road and a large path the error 2 commission of roads could be a misclassification, see D.3.5.4.2.

D.3.5.3.3 Derived conformance result

Table D.4 presents the conformance results derived from the quantitative results.

Table D.4— Completeness conformance

Evaluation ID	Quality element	Measure and measure ID	Feature type	Conformance requirements	AQL[a]	Error Count	Pop	Pass
1	Commission	Excess item (2)	Path	c	2	1	7	Yes
2	Omission	Missing item (5)	Path	b	2	3	7	No
3	Commission	Excess item (2)	Road	c	2	2	5	Yes
4	Omission	Missing item (5)	Road	b	2	0	5	Yes
5	Commission	Excess item (1)	Tree	i	10 %	3	25	No
6	Omission	Missing item (5)	Tree	h	10 %	2	25	Yes
7	Commission	Excess item (2)	Industrial building	f	2	0	4	Yes
8	Omission	Missing item (5)	Industrial building	e	2	2	4	Yes
9	Commission	Excess item (2)	House	f	2	1	10	Yes
10	Omission	Missing item (5)	House	e	2	1	10	Yes

a AQL = acceptance quality limit

D.3.5.3.4 Aggregated conformance result

Conformance results regarding transport networks (paths and roads) and buildings (industrial and houses) are aggregated in Table D.5 using the following rule: if one of the original results is “No” the aggregated result will be “No”. (100 % pass fail, Annex G)

Table D.5— Aggregated completeness conformance

Scope	Quality element	Conformance requirements	Number of evaluations and ID (see Table D.4)	Counts yes/no	Pass
Transport network	Omission	b) Max two missing for each feature type	2 (evaluation ID b and d)	1/1	No
Transport network	Commission	c) Max two in excess for each feature type	2 (evaluation ID a and c)	2/0	Yes
Buildings	Omission	e) Max two missing for each feature type	2 (evaluation ID h and j)	2/0	Yes
Buildings	Commission	f) Max two in excess for each feature type	2 (evaluation ID g and i)	2/0	Yes

D.3.5.4 Thematic quality – classification correctness

D.3.5.4.1 General

Information about completeness can be further clarified by thematic quality information. For example, two of the three omitted paths are in fact classified as roads (see Table D.6). The results are shown in Tables D.6 to D.8.

D.3.5.4.2 Quantitative result

One way of depicting errors associated with thematic quality is by using the measure “misclassification matrix”.

Table D.6 is a misclassification matrix that shows errors by feature class. It explains how well the instances in the dataset are classified. The different percentages should always refer to the population in the dataset.

NOTE A misclassification matrix is a square matrix where the *i*, *j* element corresponds to the quantity classified as belonging to class *j* when it actually belongs to class *i*.

Table D.6— Feature misclassification matrix

Universe of discourse	Dataset					
	Path	Road	Tree	Industrial building	House	Sum
Path	4	2	0	0	0	6
Road	0	5	0	0	0	5
Tree	0	0	23	0	0	23
Industrial building	0	0	0	2	1	3
House	0	0	0	0	9	9
Sum	4	7	23	2	10	46

The discrepancy between the sum and the number of items in the universe of discourse and the dataset comes from the missing and excess items.

D.3.5.4.3 Derived conformance result

Table D.7 presents the conformance results derived from the quantitative results.

Table D.7— Thematic quality conformance

Evaluation id	Quality element	Measure	Feature type	Conformance requirements	AQL	Mis-classification Count	Pass
11	Thematic classification correctness	Number of incorrectly classified features	Path	d	2	2	Yes
12	Thematic classification correctness	Number of incorrectly classified features	Road	d	2	0	Yes
13	Thematic classification correctness	Number of incorrectly classified features	Industrial building	g	2	1	Yes
14	Thematic classification correctness	Number of incorrectly classified features	House	g	2	0	Yes
15	Thematic classification correctness	Number of incorrectly classified features	Tree	k	0	0	Yes

D.3.5.4.4 Aggregated conformance result

Conformance results regarding transport networks (paths and roads) and buildings (industrial and houses) are aggregated in Table D.8 using the following method: if one of the original results is “No” the aggregated result will be “No” (100 % pass fail, see Annex G).

Table D.8— Aggregated classification correctness conformance

Scope	Quality element	Conformance requirements	Number of evaluations and id (see Table D.7)	Counts yes/no	Pass
Transport network	Thematic classification correctness	d) Max two feature instances in each feature type misclassified as another of the Transport Network feature type	2 (evaluation No. 11 and 12)	2/0	Yes
Buildings	Thematic classification correctness	g) Max two feature instances misclassified as another of the Building feature types	2 (evaluation No. 13 and 14)	2/0	Yes

D.3.5.5 Thematic quality – quantitative attribute accuracy

D.3.5.5.1 General

The type of measure tested for in this example is quantitative attribute accuracy. In Table D.9, only features that have a homologue in the same feature type (“class”) are taken into account. The results are shown in Table D.9 and Table D.10.

D.3.5.5.2 Quantitative result

Attribute height of trees is shown in Table D.9

Table D.9 — Feature attribute height misclassification matrix – Tree height

Universe of discourse	Dataset				
	Class A 1 m to 3 m	Class B 3 m to 5 m	Class C 5 m to 10 m	Class D > 10 m	Sum
Class A	3	1	0	0	4
Class B	1	5	0	0	6
Class C	0	2	6	2	10
Class D	0	0	0	2	2
Sum	4	8	6	4	22

One tree is missing class code and is therefore not counted in the misclassification matrix. This error could be reported as a domain consistency error.

D.3.5.5.3 Derived conformance result

Table D.10 presents the conformance results derived from the quantitative results.

Table D.10 — Thematic quality conformance

Quality element	Measure and measure id	Feature type / attribute	Conformance requirements	AQL	Misclassification Count	Pop	Pass
Quantitative attribute accuracy	Misclassification matrix (62)	Tree / height Class	j	20 %	6	22	No

D.3.5.6 Aggregated conformance to data product specification

In Table D.11, all the conformance results for buildings, transport network and trees are aggregated together with the conformance to the conceptual schema to provide the conformance to the data product specification following the registered measure “data product specification passed”.

Table D.11 — Conformance to the data product specification

Scope	Conformance requirements	Number of evaluations	Counts yes/no	Conformant
Dataset	To be conformant with the data quality requirements, a dataset shall pass all the data quality requirements in the application schema.	11 requirements	8/3 (Not passed req. 2, 9 and 10)	Dataset NOT conformant

D.4 Reporting data quality

D.4.1 Reporting as metadata

D.4.1.1 General

D.4.1.2 to D.4.1.4 present examples of how to report the quality results as metadata, as described in this document (Clause 11 and Annex C) and in ISO 19115-1:2014. Indeed, one instance of MD_Metadata aggregates one or more instances of DataQuality.

In the examples, some instances of classes (DataQuality and QualityElement) have been given an identifier (id) according to XML principles. These identifiers are used when referencing to those instances within other classes.”

D.4.1.2 Reporting commission

Table D.12 presents an example of how to report the quantitative results, derived conformance result and aggregated conformance result for the transport network feature types.

The mechanism for reporting these results is similar for the others feature types of the dataset.

Table D.12 — Reporting commission as metadata

XML element	Example	Comment
DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Dataset	Scope of this data quality unit
QualityEvaluationReport: QualityEvaluationReportInformation		
reportReference: CI_Citation		Reference and abstract of the attached quality evaluation report.
title: CharacterString	Reporting as quality evaluation report, see D.4.2	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
abstract: CharacterString	The quality evaluation report attached to this quality evaluation is providing more details on the derivation and aggregation method.	
report: Commission ***id = quantitative_commission***		In this instance of commission, the quantitative result is provided for each feature type for the measure 2 (number of excess item)
measure: MeasureReference		
nameOfMeasure: CharacterString	Number of excess item	
measureIdentification: MD_Identifier		
code: CharacterString	2	
measureDescription: CharacterString	number of items within the dataset that should not have been in the dataset	
evaluation: FullInspection		
evaluationMethodType: EvaluationMethodTypeCode	directExternal	
evaluationMethodDescription: CharacterString	Compare count of items in the dataset against count of items in universe of discourse	
result: QuantitativeResult		For more readability, only commission for paths and roads are reported here, but every feature type shall be reported since the data quality scope is the dataset.
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Path	

XML element	Example	Comment
value: Record	0	
valueUnit: UnitOfMeasure	None	
result: QuantitativeResult		
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Road	
value: Record	2	
valueUnit: UnitOfMeasure	None	
report: Commission ***id = conformance_commission***		In this instance of commission, the derived conformance result is provided for each feature type for the measure 1 (excess item).
measure: MeasureReference		
nameOfMeasure: CharacterString	excess item	
measureIdentification: MD_Identifier		
code: CharacterString	1	
measureDescription: CharacterString	Indication that an item is incorrectly present in the data	
evaluation: AggregationDerivation		
evaluationMethodType: EvaluationMethodTypeCode	indirect	
evaluationMethodDescription: CharacterString	Derivation from quantitative result	
derivedElement: Element	***quantitative_commission***	Reference to the original results.
result: ConformanceResult		Derived conformance result for the path commission For more readability, only commission for paths and roads are reported here, but every feature type shall be reported since the data quality scope is the dataset.
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Path	
specification: CI_Citation		
title: CharacterString	Data product specification (see D.2.1.4) requirement b	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
pass: Boolean	True	
result: ConformanceResult		Derived conformance result for the road commission. For more readability, only commission for paths and roads are reported here,
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		

XML element	Example	Comment
features: GF_FeatureType	Road	but every feature type shall be reported since the data quality scope is the dataset.
specification: CI_Citation		
title: CharacterString	Data product specification (see D.2.1.4) requirement b	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Creation	
pass: Boolean	true	
DataQuality ***id = agg_commission1***		Aggregated conformance result for Transport Network.
scope: MD_Scope		The scope is now the feature types for Transport Network, which implies that the data quality unit changed. That is why a new instance of DataQuality was created.
level: MD_ScopeCode	FeatureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	TransportNetwork (road and path)	
report: Commission		
evaluation: AggregationDerivation		Aggregation method.
evaluationMethodType: EvaluationMethodTypeCode	indirect	
evaluationMethodDescription: CharacterString	100 % pass fail aggregation of the conformance commission result for roads and paths	
evaluationProcedure: CI_Citation		
title: CharacterString	Annex G	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Creation	
derivedElement: Element	***conformance_commission***	Reference to the original results.
result: ConformanceResult		
specification: CI_Citation		
title: CharacterString	Data product specification (see D.2.1.4), requirement b	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Creation	
Pass: Boolean	true	

D.4.1.3 Reporting classification correctness

Table D.13 presents an example of how to report the derived conformance results and aggregated conformance result for the Buildings feature types.

The mechanism for reporting these results is similar for the others feature types of the dataset.

Table D.13 — Reporting classification correctness as metadata

XML element	Example	Comment
DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Dataset	Scope of this data quality unit.
QualityEvaluationReport: QualityEvaluationReportInformation		
reportReference: CI_Citation		Reference and abstract of the attached quality evaluation report.
title: CharacterString	Reporting as quality evaluation report see D.4.2	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
abstract: CharacterString	The quality evaluation report attached to this quality evaluation is providing all the quantitative results which are not provided in the metadata, and more details on the derivation and aggregation method.	
report: ThematicClassificationCorrectness ***id = conformance_classification***		In this instance of classification correctness, the derived conformance result is provided for each feature type for the measure 60 (number of incorrectly classified features).
measure: MeasureReference		
nameOfMeasure: CharacterString	Number of incorrectly classified features.	
measureIdentification: MD_Identifier		
code: CharacterString	60	
evaluation: AggregationDerivation		
evaluationMethodType: EvaluationMethodTypeCode	Indirect	
evaluationMethodDescription: CharacterString	Derivation from quantitative results reported in the quality evaluation report.	
QualityEvaluationReportDetails: CharacterString	The original quantitative results are described in E.3.4.4.2 of the quality evaluation report.	Reference to the original results.
result: ConformanceResult		Derived conformance result for the industrial buildings classification. The original quantitative result is intentionally not provided in metadata. It is described in the quality evaluation report.
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Industrial Building	
specification: CI_Citation		

XML element	Example	Comment
title: CharacterString	Data product specification (see D.2.1.4), requirement g	The attribute QualityEvaluationReportDetails give the precise reference to the original result within the quality evaluation report.
date: CI_Date		For more readability, only classification for industrial buildings and houses are reported here, but every feature type shall be reported since the data quality scope is the dataset.
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
explanation: CharacterString	The original quantitative result is provided in D.3.5.5.2 of the quality evaluation report.	
pass: Boolean	True	
result: ConformanceResult		Derived conformance result for the industrial buildings classification. The original quantitative result is intentionally not provided in metadata. It is described in the quality evaluation report. The attribute QualityEvaluationReportDetails give the precise reference to the original result within the quality evaluation report. For more readability, only classification for industrial buildings and houses are reported here, but every feature type shall be reported since the data quality scope is the dataset.
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	House	
specification: CI_Citation		
title: CharacterString	Data product specification (see D.2.1.4), requirement g	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
explanation: CharacterString	The original quantitative result is provided in quality evaluation report.	
pass: Boolean	True	
DataQuality ***id = agg_classification2***		Aggregated classification correctness result for Buildings
Scope: MD_Scope		The scope is now the Buildings feature types, which implies that the data quality unit changed. That is why a new instance of DataQuality was created.
level: MD_ScopeCode	FeatureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Buildings (industrial building and house)	
report: ThematicClassificationCorrectness		
evaluation: AggregationDerivation		Aggregation method
evaluationMethodType: EvaluationMethodTypeCode	Indirect	
evaluationMethodDescription: CharacterString	100 % pass fail aggregation of the conformance classification correctness result for industrial buildings and houses	
evaluationProcedure: CI_Citation		

XML element	Example	Comment
title: CharacterString	Annex G	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
derivedElement: Element	***conformance_classification***	Reference to the original results
result: ConformanceResult		
specification: CI_Citation		
title: CharacterString	Data product specification (see D.2.1.4), requirement g	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
pass: Boolean	True	

D.4.1.4 Reporting conformance to the data product specification

Table D.14 presents an example of how to express the conformance to the data product specification by aggregating the results for the different requirements.

Table D.14 — Reporting conformance as metadata

XML element	Example	Comment
DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Dataset	
QualityEvaluationReport: QualityEvaluationReportInformation		Reference and abstract of the attached quality evaluation report.
reportReference: CI_Citation		
title: CharacterString	Reporting as quality evaluation report see E.4.2	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
abstract: CharacterString	The quality evaluation report attached to this quality evaluation is providing fully detailed information about the evaluation applied and results obtained.	
report: Quality Element		Combination of data quality elements is used to report the conformance of the dataset to the data product specification.
measure: MeasureReference		
nameOfMeasure: CharacterString	Data product specification passed.	
measureIdentification: MD_Identifier		

XML element	Example	Comment
code: CharacterString	101	
measureDescription: CharacterString	Indication that all requirements in the referred data product specification are fulfilled.	
evaluation: AggregationDerivation		
evaluationMethodType: EvaluationMethodTypeCode	indirect	
evaluationMethodDescription: CharacterString	100 % pass fail aggregation of each conformance results for the requirement expressed in the data product specification.	
evaluationProcedure: CI_Citation		
title: CharacterString	Annex G	
date: CI_Date		
date: Date	2010–07–05	
dateType: CI_DateTypeCode	Creation	
QualityEvaluationReportDetails: CharacterString	The original results are described in D.3.5.2, D.3.5.3.4, D.3.5.4.4 and D.3.5.5.3 of the quality evaluation report.	Reference to the original results in the quality evaluation report (conceptual consistency conformance result, quantitative attribute accuracy conformance result for tree heights…).
derivedElement: Element	***agg_commission1***	Reference to the aggregated commission conformance result for transport network described previously in the metadata.
derivedElement: Element	*(id)*	Reference to the aggregated commission conformance result for buildings described previously in the metadata.
derivedElement: Element	*(id)*	Reference to the commission conformance result for trees described previously in the metadata.
derivedElement: Element	*(id)*	Reference to the aggregated omission conformance result for transport network described previously in the metadata.
derivedElement: Element	*(id)*	Reference to the aggregated omission conformance result for buildings described previously in the metadata.
derivedElement: Element	*(id)*	Reference to the omission conformance result for trees described previously in the metadata.
derivedElement: Element	*(id)*	Reference to the aggregated classification correctness conformance result for transport network described previously in the metadata.

XML element				Example	Comment
	derivedElement: Element			***agg_classification2***	Reference to the aggregated classification correctness conformance result for buildings described previously in the metadata.
	derivedElement: Element			*(id)*	Reference to the classification correctness conformance result for trees described previously in the metadata.
	result: ConformanceResult				
		specification: CI_Citation			
			title: CharacterString	Data product specification (see D.2.1.4)	
			date: CI_Date		
			date: Date	2010–07–05	
			dateType: CI_DateTypeCode	Creation	
		explanation: CharacterString		3 requirements of 11 are not fulfilled: the dataset is not conformant	
		pass: Boolean		False	

D.4.2 Reporting in a quality evaluation report

The structure of the quality evaluation report is free.

D.5 Additional examples

D.5.1 General

Some concepts have not been described in the previous example. The additional examples in D.5.2 to D.5.4 show how to report descriptive result, metaquality and sampling evaluation procedures.

Some concepts have not been described in the examples in D.4.

D.5.2 Reporting descriptive results as metadata

Sometimes it can be impossible to express the evaluation of a data quality element in a quantitative way. Descriptive result could then be used. Table D.15 is an example of the reporting as metadata of descriptive results.

Table D.15 — Reporting descriptive result as metadata

XML element	Example	Comment
DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Dataset	The dataset describes archaeological objects.
report: RelativeInternalPositionalAccuracy		
evaluation: IndirectEvaluation		
evaluationMethodType: EvaluationMethodTypeCode	Indirect	
evaluationMethodDescription: CharacterString	Compares absolute positional accuracy of the archaeological objects and the absolute positional accuracy of the rivers.	
deductiveSource: CharacterString	Positional accuracy of the rivers nearby the archaeological camp.	
result: DescriptiveResult		
statement: CharacterString	Relative positional accuracy between archaeological objects and rivers is higher than the absolute positional accuracy of the archaeological objects (5 m).	

D.5.3 Reporting metaquality as metadata

The absolute positional accuracy of the topological survey on an archaeological site is evaluated: The result is 5 m accuracy.

An evaluation of the quality of the evaluation is then provided using the confidence metaquality element, for which a measure called “Safety Factor” is used.

Table D.16 describes how to report metaquality as metadata.

Table D.16 — Reporting metaquality as metadata

XML element	Example	Comment
DataQuality		
scope: MD_Scope		
level: MD_ScopeCode	Dataset	
report: AbsolutExternalPositionalAccuracy ***id = positionalaccuracy1***		Absolute positional accuracy report. An id is provided to the data quality element in order to be able to reference it in the following metaquality element. All optional attributes have not been filled here.
measure: MeasureReference		
nameOfMeasure: CharacterString	Root mean square error planimetry	
measureIdentification: MD_Identifier		
code: CharacterString	47	
measureDescription: CharacterString	Standard deviation where the true value is not estimated from the observations but known a priori	
evaluation: FullInspection		

XML element	Example	Comment
evaluationMethodType: EvaluationMethodTypeCode	directExternal	
evaluationProcedure: CI_Citation		
title: CharacterString	IGN data quality evaluation procedure	
date: CI_Date		
date: Date	1995–02–09	
dateType: CI_DateTypeCode	Creation	
result: QuantitativeResult		
value: Record	5	
valueUnit: UnitOfMeasure	Metre	
report: Confidence		Metaquality report (confidence) related to the previous accuracy report.
relatedElement: Element	***positionalaccuracy1***	
measure: MeasureReference		
nameOfMeasure: CharacterString	Safety Factor	
measureIdentification: MD_Identifier		
code: CharacterString	1	
authority: CI_Citation		
title: CharacterString	IGN Measures	
date: CI_Date		
date: Date	1995–01–01	
dateType: CI_DateTypeCode	creation	
measureDescription: CharacterString	The ratio between the accuracy class of the evaluation elements and the accuracy class that has to be obtained in the dataset.	
evaluation: FullInspection		
evaluationMethodType: EvaluationMethodTypeCode	directExternal	
evaluationMethodDescription	The bigger the “Safety Factor” is the more trustful is the evaluation. The “Safety Factor“ has to be bigger than 2 to validate the evaluation	
evaluationProcedure: CI_Citation		
title: CharacterString	Arrêté 2003 (French legislation)	
date: CI_Date		
date: Date	2003	
dateType: CI_DateTypeCode	Publication	
result: QuantitativeResult		
value: Record	2.4	
valueUnit: UnitOfMeasure		

D.5.4 Reporting alternatives

D.5.4.1 Example of quality report using coverage result

The actuality of a dataset can be described by the latest control date. In this example a date is expressed as a year when an area was updated (e.g. from aerial photography, or for a theme updated from another dataset). Figure D.5 illustrates the example of using a coverage result to report evaluation of dataset's actuality. In example in Figure D.5 on the left, the updated part of the database is reported as a coverage, with pixel values representing the update year. The example in Figure D.5 on the right shows a date when the data has been controlled with respect to completeness and positional accuracy for a scope (e.g. buildings and roads).

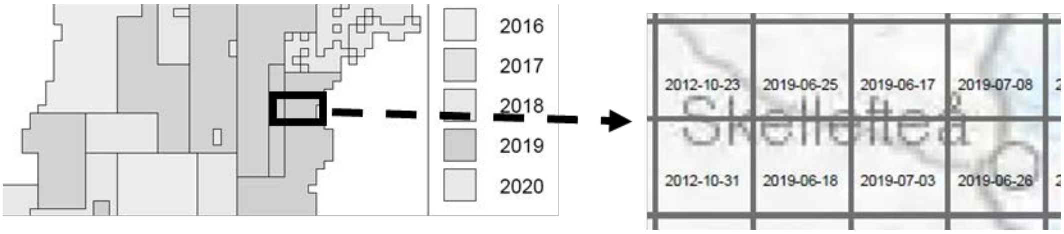


Figure D.5 Coverage result to report evaluation of dataset's actuality

D.5.4.2 Example of quality report for lidar data height model using free text

General information about quality

The measured or estimated quality applies to the entirety of the surface scanned. For the area in question, it is therefore important to refer to metadata for information on survey technique, time of scanning and point density. Good knowledge of the local terrain conditions is also valuable in the interpretation of data.

Completeness

The point density on ground varies with terrain type, type of vegetation, the season in which laser scanning was performed and a number of other factors. This variation means that in some areas there are deficiencies in the point density, while in other places the point density is high[34].

D.5.4.3 Example of quality report for INSPIRE Hydrography Positional accuracy

Geometrical requirements on positional accuracy depend on the object's distinctness within a geographically limited area. Concrete objects have higher requirements than objects with diffuse boundaries.

Objects in the water have very high positional accuracy.

The shoreline is measured at normal water level, except in regulated lakes and rivers where it is measured at the maximum damming limit. The position of the shoreline may vary due to different water levels. Updating is only carried out when it is clear that there has been a major lasting change.

The positional accuracy of streams is high in open surfaces but is varied in forest areas where it is difficult to see through the canopy. Large errors regarding streams are gradually being corrected using laser/height data in forest areas. Other hydrographic objects have very high positional accuracy [35].

D.5.5 How to report sampling procedure

This example is based upon a Topographic Database produced by a European national land survey. The quality conformance levels have been defined in the data product specification.

Road feature type is evaluated in this example through a sampling evaluation.

The sampling procedure is applied using the principles of ISO 2859-1, as described in Table D.17.

Table D.17 — Procedure for sampling

Process step	Example
Define a sampling method	Multistage sampling. Selecting enough sampling units so that sample ratio is fulfilled. Sampling is based on weighted features.
Define items	All features.
Divide the data quality scope (population) into lots	Number of datasets.
Divide lots into sampling units	N-number 1 km × 1 km squares.
Define the sampling ratio or the size of the sample	Sample size depends on the AQL value for that lot.
Select sampling units	Select required number of sampling units so that sampling ratio or sample size for items is fulfilled.
Inspect items in the sampling units	Inspect every item in the sampling units.

If the quality requirements for the feature is 1 nonconformity per 100 units (AQL = 1), then all features collected are checked from the data source. Inspection by sampling is done when the AQL = 4 or 15.

A lot used for testing should consist of datasets produced as far as possible at the same time and with the same methods. From the lot, sampling units of N-number 1 km × 1 km squares are selected so that the number of features in the sample is sufficient for an AQL = 4.

Table D.18 is an example of how to report sampling procedure information as metadata.

Table D.18 — Reporting results of sampling evaluation as metadata

XML element	Example
DataQuality	
scope: MD_Scope	
level: MD_ScopeCode	Feature Type
levelDescription: MD_ScopeDescription	
features: GF_FeatureType	Road
report: Commission	
measure: MeasureReference	
nameOfMeasure: CharacterString	Number of excess item
measureIdentification: MD_Identifier	
code: CharacterString	2
measureDescription: CharacterString	Number of items within the dataset that should not have been in the dataset.

XML element	Example
evaluation: SampleBasedInspection	
evaluationMethodType: EvaluationMethodTypeCode	directExternal
evaluationMethodDescription: CharacterString	Multistage sampling. Selecting enough sampling units so that sample ratio is fulfilled. Sampling is based on weighted features.
evaluationProcedure: CI_Citation	
title: CharacterString	Annex E
date: CI_Date	
date: Date	2010–07–05
dateType: CI_DateTypeCode	Publication
referenceDoc: CI_Citation	
title: CharacterString	ISO 2859-1
date: CI_Date	
date: Date	1999–11–18
dateType: CI_DateTypeCode	Publication
lotDescription: CharacterString	A lot is a group of databases (e.g. 1:10 000 map sheet) which are taken for inspection. The lot size is the number of features in the lot. All the roads in the dataset (one lot for the whole dataset).
samplingScheme: CharacterString	From the lot an area of so many 1km × 1 km squares are sampled that the number of roads in the sample is at least the same as AQL = 4 requires.
samplingRatio: CharacterString	On average an area comprising format sheets (16 databases) with 6 to 10 squares (1 km × 1 km) is recommended as a practical lot size.
result: ConformanceResult	
specification: CI_Citation	
title: CharacterString	Data product specification (see D.2.1.4), requirement b
date: CI_Date	
date: Date	2010-07-05
dateType: CI_DateTypeCode	Creation
Pass: Boolean	True

Annex E
(informative)

Sampling methods for evaluating data quality

E.1 Introduction

This annex provides guidelines for defining samples and devising sampling methods. For sampling for evaluating conformance to a data product specification, the ISO 2859 series and ISO 3951-1 may be applied. These standards were originally developed for non-spatial use. This annex describes how to apply the ISO 2859 series and ISO 3951-1 and other spatial sampling techniques to geographic data.

E.2 Lot and item

Lot and item are important concepts in the sampling inspection method specified in the ISO 2859 series and ISO 3951-1. A lot is the minimum unit for which quality may be evaluated. An item is the minimum unit to be inspected and should be defined by the data producer in accordance with the data product specification.

E.3 Sample size

The size of a population, and consequently the size of samples, may be defined according to different bases on items. The definition of a sample size requires an explicit indication of the items. Examples of different bases are presented in Table E.1.

The difference between the perspectives is illustrated in Figure E.1. The whole figure represents the data within the data quality scope. The figure depicts a possible sample area of approximately 15 % of the total data quality scope area, but only about 10 % of the curve length within the sample area, and 0 % of the vertices.

NOTE The data quality scope is the area in the outer box. The sample area is the shaded box.

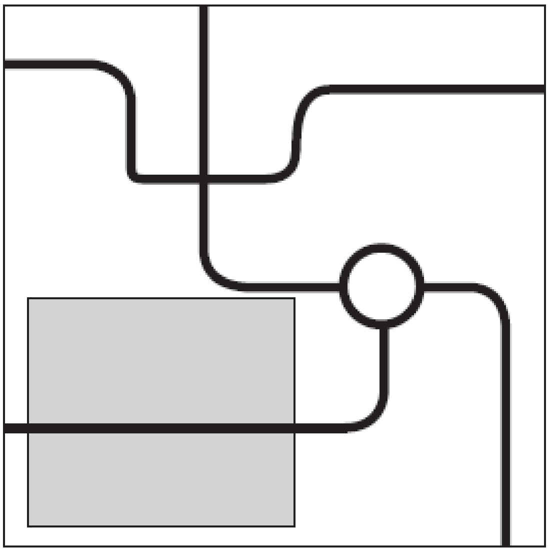
To help overcome sample difficulties such as those in Figure E.1, the size and location of a sample might be defined using a combination of different criteria, thus enforcing the representativity of the sample.

NOTE The data quality scope is the area in the outer box. The sample area is the shaded box.

EXAMPLE The sample should include 10 % of the area covered by the dataset and contain not less than 5 % of the total curve length describing the objects in the dataset.

Table E.1 — Different basis for defining population

Basis	Size of the dataset	Sample size
Features	Number of features of a given type.	Number of features of a given type expressed as percentage of the total number of objects.
Area covered	Area covered by the dataset.	Area covered by the sample expressed as percentage of the total area.
Curves	Total length of the curves in the dataset.	Length of the sampled curves expressed as a percentage of the total length.
Vertices	Total number of vertices describing curves or areas in the dataset.	Number of vertices in the sample expressed as a percentage of the total number of vertices.



NOTE The data quality scope is the area in the outer box. The sample area is the shaded box.

Figure E.1 — Effect of sample area location on representativity of items in the sample

E.4 Sampling strategies

E.4.1 Introduction

This clause provides guidelines for defining samples and sampling methods, considering particular aspects of geographic data. The sampling strategies described in this annex are shown graphically in Figure E.2. There are two aspects to a sampling strategy: the items to be sampled (area or feature), and the manner by which the items are selected (probability or judgement).

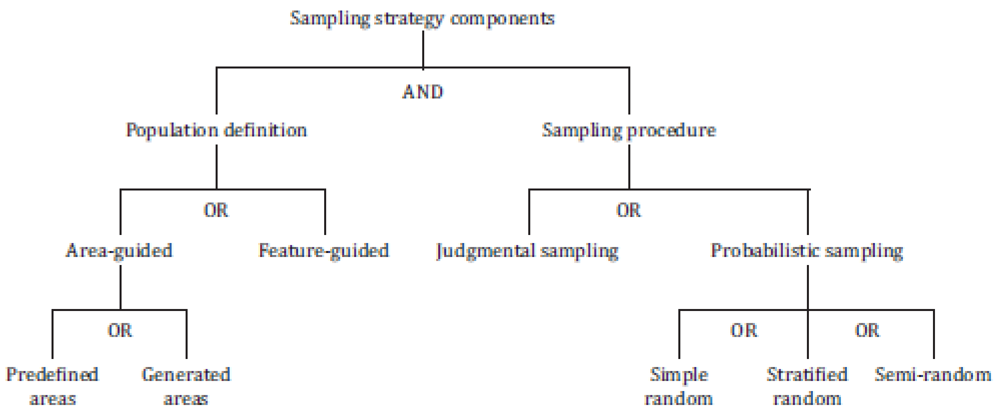


Figure E.2 — Sampling strategy relationships

E.4.2 Probabilistic versus judgemental sampling

E.4.2.1 Differences

Probabilistic sampling applies sampling theory and involves random selection of the sample items. The essential characteristic of probabilistic sampling is that each member of the population from which the sample is selected has a known probability of selection. When probabilistic sampling is used, statistical inferences may be made about the sampled population. Judgemental sample designs involve selection of samples based on expert knowledge or professional judgement.

E.4.2.2 Simple random sampling

Simple random sampling is probability-based and involves selection of samples randomly. The particular sample (e.g. features, location, time) is selected using random numbers to identify the items and all possible selections are equally likely. Simple random sampling is useful when the population of interest is relatively homogeneous in the characteristics being sampled, i.e. no major patterns and clusters. This method may not result in representative coverage of an area, i.e. it is possible that the sample selected will be only from a part of the area.

E.4.2.3 Stratified random sampling

Stratified sampling requires the population to be separated into non-overlapping strata or subpopulations that are more homogeneous among sample items in the same strata than among sample items in different strata. This sampling strategy has the potential for greater precision in estimates of mean and variance than that of a non-stratified strategy for the same population.

E.4.2.4 Semi-random sampling

Semi-random or systematic sampling applies random selection of the initial sample items (e.g. location, time, or feature) and rules for selection for all remaining items. An example of semi-random or systematic sampling is grid sampling where the initial position of a grid is randomly determined and samples are taken at regularly spaced intervals (grid cells) over space. Systematic grid sampling is used to search for clusters and to infer means, percentiles or other parameters, and is useful for estimating spatial trends or patterns. This method provides a practical and easy way to ensure coverage of an area.

E.4.3 Feature-guided versus area-guided sampling

E.4.3.1 Feature-guided sampling (non-spatial sampling)

A feature-guided sampling strategy selects sample items based on the non-spatial attributes of the features and not on their spatial location. A sample within a data quality scope can be selected randomly, assuming homogeneous production characteristics for the entire data quality scope. In some cases, simple random sampling may not produce a satisfactory sample because homogeneity may be found only for subsets and homogeneous distribution of samples may be required; i.e. major patterns or clusters occur in the characteristics being sampled. In that case, a stratified or semi-random sampling may give better results.

NOTE If the sampling method is defined by selecting features randomly, then there is the risk of the occurrence of a sample being concentrated in a small area (which can potentially not be acceptable).

Semi-random sampling may be used to ensure the verification of different criteria on the sample size and/or location, to satisfy supplementary constraints for the samples or to reduce costs of the inspection process.

EXAMPLE A power company needs to evaluate the correctness of the attributes surveyed for features of different types. Two methods were considered: a random selection and a semi-random selection (selecting randomly the features of one type and then collecting the objects of different types in the neighbourhood of the first one until the samples for each type become fulfilled) leading to a reduced field inspection cost.

E.4.3.2 Area-guided sampling (spatial sampling)

In an area-guided sampling strategy, selection of sampling units is based on spatial considerations. The sampling units may be existing geographic areas (e.g. political or statistical areas) or some other partitioning of the universe of discourse for which the inspection is conducted. This type of sampling may be used as a first stage of sampling, followed by a feature-guided sampling within each subarea.

EXAMPLE Random selection of UTM 1 × 1 km grid areas in order to evaluate the attributes of the objects contained in that area.

Figure E.3 illustrates the result of the definition of areas to be submitted for inspection, obtained by random generation of centre point coordinates of squares of equal area (constrained to be non-overlapping).

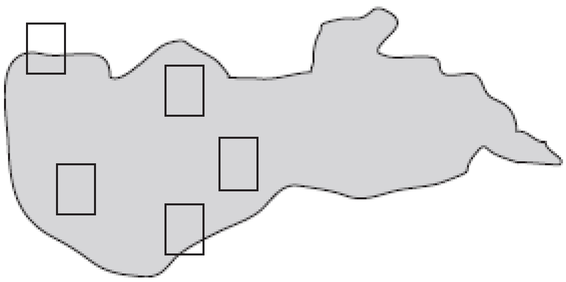
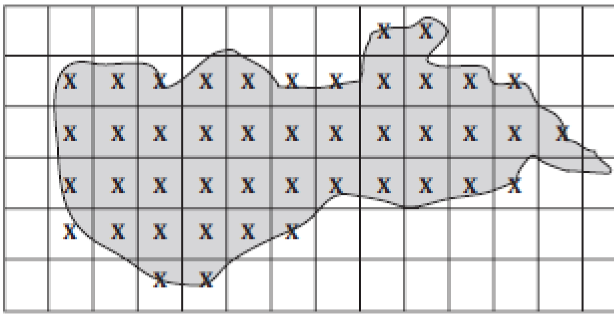


Figure E.2 — Example of area-guided random sampling

When coverage of the entire area is important, then the sample locations should be determined according to a regular or semi-regular pattern.

Figure E.4 illustrates an example of semi-random (systematic) sampling with the sampled features distributed along a regular pattern used to evaluate the positional accuracy of a dataset.



NOTE The “X” denotes the grid cells selected by rule for inclusion in the sample.

Figure E.3 — Example of area-guided regular and non-random sampling

Spatial partitioning with different sizes in different areas of the dataset may be needed in semi-random sampling, if the distribution of features is non-homogeneous. When using a grid of constant cell size, a rule is needed to include or exclude cells that are not completely inside the area of interest.

E.5 Probability-based sampling

E.5.1 General considerations

In applying sampling, the following points need to be taken into account:

a) The areas covered by a geographic dataset may form a continuous space. When splitting the dataset into lots, special attention should be paid to the omission or commission of items crossing over the lot boundaries;

b) A variety of factors, including the quality of source data and skill of operators, may affect the quality of geographic data. The data producer should be careful to define lots to achieve homogeneity in terms of quality.

Depending on the sample design, design weights (which in a probability sample corresponds to the reciprocal value of the probability of being included into the sample) may be used to define items in a sample lot.

E.5.2 Existing standard for inspection by sampling

E.5.2.1 General

Based on the characteristics of production and in accordance with the data product specification, suitable documents for inspection by sampling should be selected from the existing standards. ISO 2859-1 is primarily for the inspection of a continuing series of lots. ISO 2859-2 may be applied for individual or isolated lots, while ISO 2859-3 may be applied for skip-lot sampling procedures. ISO 3951-1 may be applied for the inspection by variables for percentage nonconforming items.

The conformance quality level of a dataset is specified as AQL (acceptance quality limit) in ISO 3534-2.

NOTE In earlier editions of ISO 2859-1, ISO 2859-3 and ISO 3951-1, the AQL was called Acceptable Quality Level, and Limiting Quality) in earlier editions of ISO 2859-2.

Specification limits for determining conformity of each item should be specified when applying the ISO 2859 series based on the data product specification. In applying ISO 3951-1 quality statistics should be specified based on the data product specification.

E.5.2.2 Useful tables based on these standards – sample size and rejection limits

E.5.2.2.1 Overview

When sampling is used, the estimated missing rate cannot be directly compared to the AQL. Table E.2 and Table E.4 provide guidelines on the sample size according to dataset size, and on the rejection level associated.

E.5.2.2.2 Evaluating conforming/non-conforming items with samples

Table E.2 below presents the recommended sample size according to population size, and the rejection limit associated, for evaluating conforming/non-conforming items, e.g. for evaluating completeness. It is based on the hypergeometric distribution). It is assumed that the deviations fit this distribution.

How to use the table:

a) Decide the population size of the items to be checked;

b) Select the sample size (n) from the table;

c) Carry out the evaluation, and count number of “fail items”;

d) The whole population is rejected if the number of fails is equal or higher than the rejection limit for the actual n and p_0 (AQL).

Table E.2 — Statistical values for testing of number of conforming/non-conforming items Significance level 95 %

Population size		p_0 =	0,5 %	1,0 %	2,0 %	3,0 %	4,0 %	5,0 %
From	**To**	**Sample size (n)**	**Rejection limit**					
1	8	All	1	1	1	1	1	1
9	50	8	1	1	1	2	2	2
51	90	13	1	1	2	2	2	3
91	150	20	1	2	2	3	3	4
151	280	32	1	2	3	3	4	4
281	400	50	2	3	3	4	5	6
401	500	60	2	3	4	5	6	7
501	1 200	80	3	3	5	6	7	8
1 201	3 200	125	3	4	6	8	10	11
3 201	10 000	200	4	6	8	11	14	16
10 001	35 000	315	5	7	12	16	20	23
35 001	150 000	500	6	10	16	23	28	34
150 001	500 000	800	9	14	24	33	42	51
> 500 000		1 250	12	20	34	49	63	76

NOTE 1 If sample size is higher than the minimum size given in the table, the rejection limit should be calculated individually. This test is valid for situations where the quality evaluation is based on a pass/fail evaluation of items.

NOTE 2 There exist other statistical value ranges than the one presented in Table E.2.

EXAMPLE Testing for missing houses (completeness/omission) in a defined area.

First a sample area is selected, and every house in the sample area is checked, to decide if it is present in the dataset or not. Then number of missing houses and the total number of houses is estimated (by counting). The question is:

Is the result significantly higher than the Acceptance Quality Limit (AQL)? If so, the dataset can be rejected. If not, the dataset is accepted.

The dataset to be checked consists of 2 440 buildings.

Sample size (from Table E.2) is n = 125. Field check shows that 2 buildings are missing, giving an estimated missing rate of:. 2/(125+2)x100%=1,6%.

AQL (from the data product specification for the dataset) is p_0 = 0,5 %.

1,6 % is higher than 0,5 %, a statistical test will be used to evaluate if the data can be rejected. As sampling is used, the estimated missing rate cannot be directly compared to the AQL. A single-sided hypothesis testing is performed, and Table E.2 helps with this. The rejection level (n = 125, p_0 = 0,5 %) is 3. In the field check 2 missing items were found.

Conclusion: As 2 is lower than 3 (rejection limit), the dataset cannot be rejected, and is accepted.

E.5.2.2.3 Standard deviation

Table E.4 presents the recommended sample size according to population size, and the rejection limit associated, when measuring a standard deviation.

To decide if the estimated standard deviation for a sample size is significantly higher than the AQL, this statistical method can be used. Table E.4 below is based on normal distribution and assumes normal distribution of deviations.

The symbols and formulas connected to the Table E.4 are presented in Table E.3.

Table E.3 — Symbols and Formulas

Standard deviation estimated based on sample	s
Sample size	n
AQL for the standard deviation	σ
F (from the F-distribution)	$F_{0.05,n-1,\infty}$
Confidence interval	$\frac{s}{\sqrt{F}}.s\times\sqrt{F}$
Standard deviation too high if:	$\sigma<\frac{s}{\sqrt{F}}$

The dataset is not good enough (i.e. can be rejected with 95 % significance) if the estimated standard deviation divided by the F-value (taken from Table E.4) is higher than the AQL.

Table E.4 — Statistical numbers for testing standard deviation. 95 % significance level

Population size		Sample size (n)	$\sqrt{F_{0.05,n-1,\infty}}$
From	**To**		
26	50	5	1,54
51	90	7	1,45
91	150	10	1,37
151	280	15	1,30
281	400	20	1,26
401	500	25	1,23
501	1200	35	1,20
1201	3200	50	1,16
3201	10000	75	1,13
10001	35000	100	1,12
35001	150000	150	1,09
150001	500000	200	1,08
> 500000		200	1,08

EXAMPLE

Positional accuracy/absolute accuracy for manhole covers is evaluated.

From a dataset containing 450 manhole covers, 25 manhole covers are measured (sample size n = 25). Estimated standard deviation s = 21 cm, AQL = 19 cm.

Lower limit for confidence interval = 21 cm/1,23 (from Table E.4) = 17,1 cm. The AQL (19 cm) is within the confidence interval of the estimated standard deviation.

Conclusion: The standard deviation from the control is not significantly higher than AQL, and the dataset cannot be rejected.

E.5.3 Sampling process

E.5.3.1 Define items

Items should be defined according to the data product specification or requirements. If nonconforming items are statistically highly correlated, they are handled as a single item.

E.5.3.2 Define data quality scopes of a dataset to be inspected

If the data quality scope is not homogeneous, it should be divided into homogeneous subsets. These homogeneous subsets should be treated as separate data quality scopes.

Homogeneity can be deduced where the following conditions occur:

— source data of production have almost the same quality;

— production systems (hardware, software, skill of operator) are essentially the same;

— other factors which may affect the likelihood of occurrence of nonconformities, such as complexity and density of features, are essentially the same.

E.5.3.3 Divide the data quality scope into lots

Lots are generated by dividing the data quality scope. When there is a strong positive spatial auto-correlation of the occurrence of nonconformity, a smaller lot size is desirable.

E.5.3.4 Divide the lot into sampling units

A sampling unit may be an existing geographic area or some other partitioning of the universe of discourse for which the inspection is conducted. When the sampling unit is a geographic area, rules should be provided for the inclusion of items partially in a sampling unit.

E.5.3.5 Select sampling units by simple random sampling for inspection

The total number of items which belong to selected sampling units should be as specified in relevant documents.

NOTE If lots are statistically heterogeneous, simple random sampling with the same level of sampling cannot be applied. The ISO 2859 series additionally allows for stratified sampling.

E.5.3.6 Inspection of selected sampling units

All items which belong to the selected sampling units are inspected. The items in the dataset are compared with the universe of discourse according to the chosen quality measure.

Annex F
(informative)

Guidelines for the use of quality elements

F.1 Overview

In some cases, there may be several possible quality elements for one specific quality requirement and one detected error in a quality evaluation. This annex provides guidelines for which quality element to use.

NOTE The quality elements are described in 8.3.

F.2 Data quality element

F.2.1 General

These different quality element categories are defined in 8.3:

— Completeness (8.3.2);

— Logical consistency (8.3.3);

— Positional accuracy (8.3.4);

— Thematic quality (8.3.5);

— Temporal quality (8.3.6);

Of the remaining elements, logical consistency is the only one that can be fully evaluated without ground truth knowledge. The logical consistency requirements and evaluations handle the “internal relationships” in the data, and how the data fits the rules set up in specifications.

The three categories completeness, positional accuracy and thematic quality are used to describe how the dataset relates to the universe of discourse.

Temporal Quality consists of a mix of data quality elements that partly is dependent upon logical rules (comparable to logical consistency) and partly needs ground truth knowledge to be evaluated (in similar way as completeness and the accuracy categories).

F.2.2 Other candidate quality elements

The accepted data quality elements can all be evaluated and the result of the evaluation will not become out of date just because the time is passing by. A possible candidate “up-to-dateness” and “timeliness” both describe how well the data represent the current real world situation. When measuring up-to-dateness (how well the dataset represents today's real world), the result will only be valid for a short time. After e.g. a year, the result of a possible stored up-to-dateness-measurement will be wrong, i.e. not telling how well the dataset represents today’s world situation, but rather how well it represented the situation one year ago.

Completeness and accuracy both describe how well the dataset represents real world at a given point of time.

F.2.3 Ordering in data quality evaluation

When evaluating geographic data, one individual error may influence several data quality elements. For measurements resulting in rates (e.g. percentage rates of aspects of completeness) the use of proper denominators describing the total population is important; see Figure F.1.

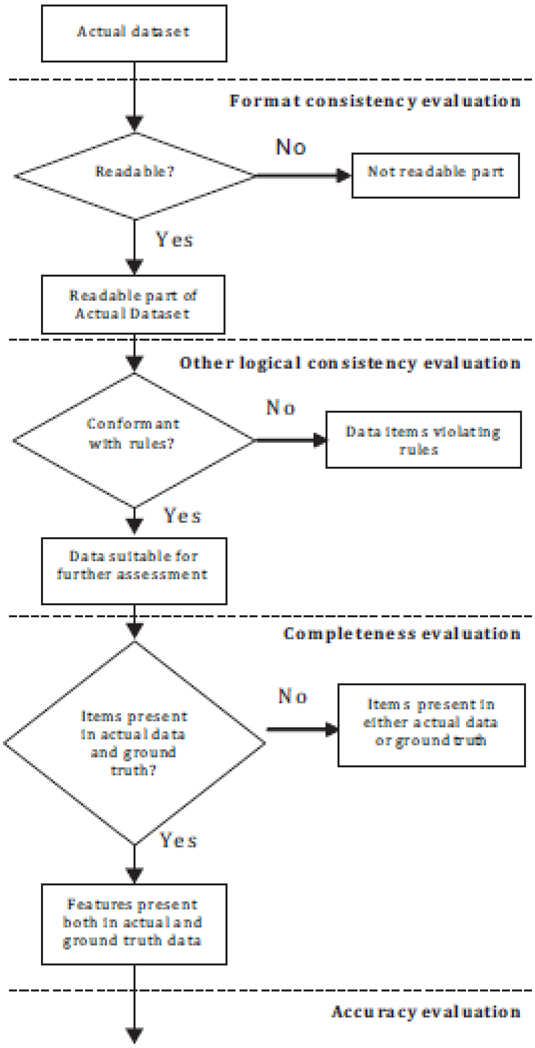


Figure F.1 — Ordering in data quality evaluation

When evaluating data quality, the usual ordering is:

a) Logical consistency/Format consistency: The very first to be evaluated is the readability (or interpretability) of the data to decide whether it is possible to decode/read/understand the data or not. Not interpretable data should be reported and ignored in the further evaluation. The result of the format consistency should describe which parts of the data are not readable.

b) Logical consistency: Decide if the rules set up for the dataset are followed. Parts of the dataset not conforming to the rules should be ignored in the further evaluation.

c) Completeness: The next step in the evaluation is the feature existence aspect covered by completeness. To evaluate this, the features in the actual dataset and the ground truth data are compared, and commissions and omissions reported. This is in most cases not possible to do for the complete dataset, a sample of the dataset will be tested according to Annex E.

d) Accuracy (positional, thematic and temporal aspects): The last step in the evaluation covers the accuracy aspect, measuring the deviation between actual and ground truth feature properties. These measurements can be based only on parts of the dataset present in both the actual dataset and the universe of discourse.

F.3 The relationships between the data quality elements

F.3.1 General

Many data quality elements are related to each other. In some cases this may lead to uncertainty about how identified deviations/errors in the data should be reported. This section discusses the relationship between the data quality elements.

F.3.2 Data quality elements related to missing attribute values

At least three different values should be considered to indicate "no value available". The way these three are used may influence the data quality element selected for reporting the missing value. The three values have different semantics:

— The empty value. In this case, the attribute has no value at all;

— The not applicable value. This indicates that for this specific feature the attribute is not valid, i.e. have no meaning;

EXAMPLE 1 Date of death for living persons.

— The unknown value. In this case, the attribute is valid i.e. there should have been a value, but the value is not known.

Mandatory attributes with empty values should be reported as logical consistency errors. Not applicable mandatory attributes should not be counted when evaluating attribute completeness. The amount of unknown occurrences should be reported as attribute completeness.

A way of increasing the attribute completeness is to add artificial values to a dataset. By doing so, the dataset will become better from an attribute consistency point-of-view, but the attribute accuracy will decrease.

EXAMPLE 2 A dataset has 50 feature instances of feature type Tree. 45 of them have a stored attribute value for the attribute HeightOfTree. The accuracy of this attribute (the 45 instances) is estimated to ± 1 m (standard deviation), and the attribute completeness is 45/50, i.e. 90 %. If however these missing HeightOfTree-values were given wrong (dummy) values of e.g. 10 m, then the attribute completeness would become better (100 %), but the attribute accuracy will probably become poorer.

F.3.3 Relationships between the different aspects of accuracy

Deviations of actual data from the universe of discourse can be measured using positional accuracy, time (temporal) accuracy and attribute (thematic) accuracy. Examples of alternative ways of expressing the deviation are:

— Attribute versus space: For attributes where the geographical distribution is known, a deviation can be expressed either by the theme or the positional component. The height value of a contour line can be considered as an attribute of the contour line. The deviation of the current position from the true position can be measured either by the attribute component ("half a metre too high") or by the space component ("the contour line has an offset of 10 m in north direction").

— Space versus time: If the movement of a feature is known, a difference between measured and real position can be expressed either by the time component or by the positional component, for example the positional error for a car moving along a road can be expressed either as “The position given would have been correct 20 sec ago” or “the position is now out by 400 m”.

— Attribute versus time: “The price ($\$/m^2$) for the specific parcel is wrong by $20”, or “this was the correct price 10 years ago”

F.3.4 Dependency between completeness and accuracy

Evaluation of completeness usually is based on comparison of the dataset and the universe of discourse. The critical operation is the linking between features in the dataset and the universe of discourse. When a unique identifier exists the linking is usually based on this.

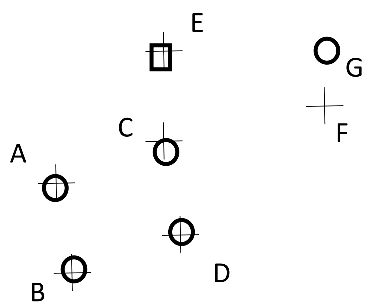
When handling features without this kind of identification of the individuals, methods based on closeness of attributes and attribute values have to be used. When linking geographical features two aspects have to be considered:

a) the thematic closeness (usually expressed as feature type);

b) the geographical closeness of the features.

When two features (a pair with one in the dataset and the other in the ground truth) are decided to be representations of the same real-world phenomenon, the deviations between the two are handled as accuracy. If the pair of features is decided to represent different phenomena, the deviation between the two is reported using completeness (omission and/or commission).

For example when evaluating completeness and accuracy for feature type 1, see Figure F.2, there is no problem in positions A, B, C and D. Here the classification is identical (thematic deviation equal to zero) and the geographical deviations between actual and real position are within the accepted level. The features are linked, and the deviations are described by positional accuracy. In position E, the two instances have different thematic classifications but are located very close to each other. A decision has to be made whether the difference in classification is within the level of acceptance for linking. If yes, the two instances will contribute to the accuracy evaluation (positional and/or thematic), if not it is a question of completeness (one point missing and one in excess). In positions F and G, the two instances have the same classification, but differ in position. If this geographical deviation is considered to be within the level of acceptance for linking, the deviation will contribute to positional accuracy (probably an outlier), if not it is a question of completeness (omission and commission).



Key

Features classified to feature type 1

SGround true position

actual dataset position

Features classified to feature type 1

actual dataset positioning

Figure F.2 — Accuracy versus completeness

F.4 Data quality elements – example of use

F.4.1 Completeness

F.4.1.1 General

The presence and absence of features may be described by the data quality elements commission and omission. Completeness should mainly be used on the feature type level, describing whether the features in the universe of discourse are found in the dataset or not.

Completeness may also be relevant for feature properties ("attribute completeness" and "relationship completeness"). Before using completeness for this, the logical consistency/conceptual consistency should be carefully considered.

F.4.1.2 Commission – excess data present in a dataset

This may be applied at the feature instance level. This means that data are considered to be in "excess" if it is a whole feature instance. If there is non-required data within a feature instance or attribute of a feature instance then this is not considered commission.

This definition incorporates feature instances which are present in the dataset but which are not within the scope (as defined in the specification).

The rule for the examples below is defined as: "Only features present in the universe of discourse shall be included in the dataset."

EXAMPLE 1 Presence of data from Scotland as this is excluded from the scope of the dataset (England).

EXAMPLE 2 Only buildings that are bigger than 5 m^2 should be included in the dataset. Presence of buildings under 5 m^2 are reported as commission.

F.4.1.3 Omission – data absent from a dataset

Similarly to commission, this may be applied at the feature instance level. In practice this refers to the absence of feature instances whose inclusion is specified in the specification.

Omission should mainly be used when a "whole item", e.g. a feature instance is missing. If a mandatory part of an item, e.g. a mandatory attribute of a feature instance, is missing, this should be reported as a conceptual consistency error.

The rule for the example below is defined as: "All residential property within England and Wales shall be included in the dataset."

EXAMPLE Absence of a residential property within England or Wales in the dataset.

F.4.2 Logical consistency

F.4.2.1 General

The degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical) may be described by the following data quality elements.

F.4.2.2 Conceptual consistency – adherence to rules of the conceptual schema

Applications usually have a conceptual schema describing the requirements to the data structure. This conceptual schema may include:

— the name of all classes (feature types, data types, etc),

— the attribute names for all classes, and also the multiplicity limitations,

— the domains for all attributes,

— the relationships between the classes,

— the topological relationships between feature types, e.g. the relationship between an area and the border lines.

— the relationship between feature type attributes for different feature types, e.g. the relationship between the height-above-sea value from a contour line and the same from a road in the geographical crossing point for the two feature instances.

Conceptual consistency may cover all these aspects of data quality.

Others logical consistency elements (domain consistency, topological consistency) may also be considered for some of the aspects listed above if conceptual consistency is used only to ensure that the correct feature properties are present for each feature instance.

F.4.2.3 Domain consistency – adherence of values to the value domains

Domains of values are usually described by the conceptual schema of the application and may be reported as part of the conceptual consistency or as domain consistency. If the domain definitions are not existing or not valid in the conceptual schema then only the quality element domain consistency can be used.

EXAMPLE 1 An organization defines the valid value domains for each field in terms of length, data type and content. Domain consistency is used to ensure compliance to these conditions with the following exceptions:

— Where the field contains position data (i.e. Easting and Northing), in which case it is considered as positional accuracy;

— Where the field contains date/time data, in which case it is considered as temporal quality;

— Where the field contains a primary key, in which case it is considered under logical consistency.

— The rule for the example below is defined as: The LANGUAGE field shall contain either “ENG” or “CYM”

EXAMPLE 2 Domain consistency error example: “COR”

F.4.2.4 Format consistency – degree to which data are stored in accordance with the physical structure of the dataset

Format consistency should mainly be used as the first quality evaluation testing whether the dataset is in the correct format according to the (product) specification.

If certain rules are defined for defining the format of specific attributes, e.g. for generated IDs, format consistency can also be relevant for single attribute values. If attributes values are checked compared to a list of legal values (a domain), the domain consistency should be used.

EXAMPLE 1 The data product specification of a product specifies Geography Markup Language (GML) as the distribution format. If the dataset is not a GML file, then this error should be reported as format consistency error. If one single item in the GML file is “in wrong format”, e.g. text instead of number, this can be reported as conceptual consistency error or domain consistency error.

EXAMPLE 2 Within an organization this classification is used to describe tests that ensure adherence to the rules of the data product specification and includes:

— Presence, validity and uniqueness of primary key values. Example rule: Each feature instance will have a unique identifier. Format consistency error example: “NULL”.

— Foreign keys which reference an identifier for another feature instance not present in the dataset. Example rule: The PARENT_UPRN field shall contain an ID linked to an existing feature instance.

F.4.2.5 Topological consistency – correctness of the explicitly encoded topological characteristics of a dataset

Topological characteristics of the dataset describe the geometric relationships between dataset items unchanged by “rubber-sheet transformations”. The main parts of the topological constraints are supposed to be described in the conceptual schema and may be reported as conceptual consistency or topological consistency. In the case when the relevant topological requirements are not part of the conceptual schema, only topological consistency could be used.

EXAMPLE 1 For a dataset with feature types defined to be located on the shoreline of water bodies (feature types like shore line, harbour, boathouse), and also feature types for water bodies (lakes, seas, etc.). The topological relationships between the feature types are well defined in the conceptual schema, and the quality element conceptual consistency is used to report whether shorelines (1 dimension) geometry coincide with the water body (2 dimensions) geometry.

EXAMPLE 2 In a network dataset, with vague requirement in the conceptual schema for a “clean network”, the “dirty parts” (undershoot, overshoot, overlapping, self-intersecting, etc.) are reported as topological consistency errors.

F.4.3 Positional accuracy

Accuracy of the position of features in relation to Earth may be described using the data quality elements in this subclause.

Measuring positional accuracy using ground truth implies establishing “correspondence pairs” with one feature instance from the dataset and the corresponding one in the control (ground truth) dataset. If the features have unique identifiers (e.g. as for cadastral parcels) this correspondence can be set up using the identifiers. Gross errors, bias, standard deviation can be estimated and reported as positional accuracy.

With no available identifiers the correspondence has to be established using the positions. A “correspondence distance limit” shall be defined. This makes it impossible to compute gross errors. This “correspondence distance limit” shall be documented in the report. In this case:

— the feature instances in the dataset with no corresponding control dataset feature instance should be reported as completeness/commission,

— the control dataset feature instances with no corresponding dataset feature instance should be reported as completeness/omission.

F.4.4 Temporal quality

F.4.4.1 General

Accuracy of the temporal attributes and temporal relationships of features may be described using the following data quality parameters.

F.4.4.2 Accuracy of a time measurement – closeness of reported time measurements to values accepted as or known to be true

EXAMPLE Within a certain organization accuracy of a time measurements is used to ensure that:

— the value does not contravene a specific condition imposed on the field (over and above the conditions imposed by the nature of date/time data).

Example rule: The START_DATE field cannot contain a value in the future.

F.4.4.3 Temporal consistency – correctness of the order of events

The rules describing the "correctness of the order of events" may be part of the conceptual schema. It might be reported either as temporal consistency or as conceptual consistency if the rules are part of the conceptual schema.

EXAMPLE Within a certain organization temporal consistency is used to:

— confirm the consistency between date/time values relating to the life cycle of the real-world object,

— ensure the consistency of date/time values used in the management of the feature instances in the dataset.

Example rule: The END_DATE shall be the same as or after START_DATE.

Temporal consistency error example: START_DATE = "2010-02-02", END_DATE = "2000-01-01".

F.4.4.4 Temporal validity – validity of data with respect to time

The rules describing the "validity of data with respect to time" may be part of the conceptual schema. It might be reported either as temporal validity or as conceptual consistency if the rules are part of the conceptual schema.

EXAMPLE Within a certain organization accuracy of a time measurements is used to:

— ensure that the content of a date or time field is in the correct format and uses the calendar defined in the specification.

Example rule: The date value shall be in ISO 8601 format – "CCYY-MM-DD".

Temporal validity error example: "01-01-2010" or "2010-51-15".

F.4.5 Thematic quality

F.4.5.1 General

The accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships may be described using the following data quality elements.

F.4.5.2 Classification correctness – comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference dataset)

Within a certain organization, this definition is used strictly. Classifications which are not defined within the dataset specification are not considered in the evaluation of classification correctness (these are considered in the evaluation of domain consistency.

F.5Discussions on difficult cases

F.5.1 Relation between misclassification and completeness at feature type level

At feature type level, completeness and thematic quality/classification correctness are strongly related to each other. Indeed the misclassification of one feature instance to the wrong feature type will appear in the evaluation of completeness for both feature types (one commission and one omission).

Therefore it is recommended when evaluating completeness at feature level to be aware that some of commission or omission error may come from misclassification issues. It could then be useful to provide classification correctness information, but the error will then be reported twice.

To avoid reporting errors twice, it is possible to report completeness at one upper level (dataset, grouping of feature type, etc.), and misclassification at feature level.

EXAMPLE All living houses shall have an address. This requirement cannot be correctly evaluated if some living houses are classified as storehouse or garage.

Another example of this is provided in Annex D.

F.5.2 Quality elements related to unique identifiers

Some use cases are presented below associated with relevant data quality elements for describing issues with unique identifiers, see Table F.1.

Table F.1 — Quality elements related to unique identifiers

Use case	Data quality element to consider
All the unique identifiers shall have a format that fits the rules for defining them.	format consistency domain consistency
All the unique identifiers used are valid according to a list of reserved unique identifiers.	domain consistency
The same feature instance is present twice with the same unique identifier.	completeness conceptual consistency (unique identifiers shall be unique)
The same feature instance is present twice with different unique identifiers. NOTE The challenge here is to be sure that the two feature instances are really two representations of the same real world object.	commission

Annex G
(informative)

Aggregation of data quality results

G.1 Introduction

An evaluation based on a single data quality element is usually not sufficient for a user to be satisfied. The data producer will usually (and hopefully in cooperation with potential users of the product) set up a data product specification giving all the requirements set up for the product.

For a potential user, it will be of great advantage to find a statement describing that the product is evaluated based on a specification. Such a statement is an aggregated data quality result, and may be useful also in other situations than reporting conformance to a specification.

The quality of a dataset may be represented by one or more aggregated data quality results (ADQR). The ADQR combines quality results from data quality evaluations based on different data quality elements or different data quality scopes.

Examples of methods that may be used for producing an ADQR are given in G.2 to G.4 A dataset may be deemed to be of an acceptable aggregate quality even though one or more individual data quality results fails acceptance. Aggregation should therefore only be used when compelling reasons exist. The meaning of the aggregate data quality result should always be made clear.

As the ADQR may be difficult to fully understand, the meaning of the aggregate data quality result should be understood before drawing conclusions based on aggregate data quality results for the quality of the dataset.

How to report aggregated data quality results is described in 11.2.1.

G.2 100 % pass/fail

Each data quality result involved in the computation is given a Boolean value of one (1) if it passed and zero (0) if it failed. The aggregate quality (Q_{ADQR}) is determined by the equation,

$Q_{ADQR} = v_1 * v_2 * v_3 * \ldots * v_n$, where n is the number of data quality measurement frames.

If ADQR = 1, then the overall dataset quality is deemed to be fully conformant, hence pass. If ADQR = 0, then it is deemed non-conformant, hence fail. The technique does not provide a result that indicates location or magnitude of the non-conformance.

G.3 Weighted pass/fail

Each data quality result involved in the computation is given a Boolean value of one (1) if it passed and a zero (0) if it failed. Based on the significance for the purpose of the product, a weight value between 0 and 1, inclusive, is assigned to each data quality result. The total of all the weights should equal 1. The choice of weights is a subjective decision made by the data producer or user. The reason for the data producer's decision should be reported as part of the result. The aggregated quality (Q_{ADQR})is determined by the equation,

$Q_{ADQR} = v_1*w_1 + v_2*w_2 + v_3*w_3 + \ldots + v_n*w_n$, where n is the number of data quality measurement frames.

This technique does provide a magnitude value indicating how close a dataset is to full conformance as measured. It does not provide a quantitative value that indicates where conformance or non-conformance occurs.

EXAMPLE An error table (see Table G.1) is prepared to show the number of errors encountered and how they are classified according to a typical procedure used for road databases. This particular example procedure assigns weights to each error type. The sum of the weights equals 100 percent. The resulting weighted value is considered to represent the quality of the dataset.

Table G.1 — Example of computation of an aggregated quality evaluation result

Feature	**Number of items in lot**	**Number of non-conforming items**	**Ratio of non-conforming**	**Accuracy proportion** (defined as 1-ratio)	**Weights**	**Weighted value** (accuracy proportion * weight)
Road segment	19					
Incorrect		1				
Missing		0	4 / 19	0,79	50 %	0,3950
Excess		3				
Street Name						
Base name	19	5	5 / 19	0,74	15 %	0,1110
Direction-of-travel	19	1	1 / 19	0,95	25 %	0,2375
Hydrography	1	0	0 / 1	1,00	10 %	0,1000
Total accuracy	(defined as the sum of weighted accuracy proportion * 100)					84,35 %

NOTE 1 An item is defined as a road segment which is bounded by intersection points with the other roads or boundaries of sample unit.

NOTE 2 Aggregation of data quality information especially using weights does not carry great meaning for end-users and can be misleading depending on which weights the data producer has used.

G.4 Maximum/minimum value

Each data quality result is given a value v based on the significance of a data quality result for the purpose of the product. The reason for the data producer's decision should be reported as part of the dataset's quality result. The aggregated quality (Q_{ADQR}) is determined by either of the two equations,

$Q_{ADQR} = \mathrm{MAX}(v_i,\ i n = 1...n)$ or $Q_{ADQR} = \mathrm{MIN}(v_i,\ i n = 1...n)$ where n is the number of data quality measurement frames measured.

This technique provides a magnitude value indicating how close a dataset is to full conformance as measured, but only in terms of the data quality measurement frame represented by the maximum or minimum. It does provide a quantitative value that indicates where conformance or non-conformance occurs when the selected data quality measurement frame is reported along with the ADQR. However, this type of ADQR tells little about the magnitude of the other data quality results.

Annex H
(normative)

XML Encoding description

H.1 Introduction

The XML schema implementation of this document follows the rules in ISO/TS 19139-1:2019.

The XML schema also uses the patterns for decoupling XML namespaces outlined in ISO/TS 19115-3:2016, Clause 8.

NOTE The previous edition of this document provided an XML schema implementation described in ISO/TS 19157-2:2016. However, this encoding description is not a revision of ISO/TS 19157-2: 2016 or any of the previously published XML schema implementation. After publication of this document, ISO/TS 19157-2:2016 will be withdrawn

H.2 XML namespaces

The XML schema definitions pertain to the following namespace: https://schemas.isotc211.org/19157/-1/dqc/1.0

This namespace is abbreviated "dqc", which stands for data quality concepts.

This XML schema implements all the UML classes defined in this document and imports all relevant classes from other standards.

H.3 XML schema

The XML schema location for the namespace "dqc" is https://schemas.isotc211.org/19157/-1/dqc/1.0.0/dqc.xsd

Annex I
(informative)

Backward compatibility with ISO 19157:2013

The data quality model, its concepts and measures conformant with ISO 19157:2013 are not necessarily conformant with this document (ISO 19157-1:2022). This annex lists in detail the changes that have been made during the revision of ISO 19157:2013.

NOTE The heading titles in this annex correspond to heading titles in ISO 19157:2013.

- 1 Scope
    - Minor revisions to the text.
- 2 Conformance
    - Conformance requirements updated.
    - Clause has been renumbered as Clause 5.
- 3 Normative references
    - List of normative references updated.
- 4 Terms and definitions
    - This clause has been substantially revised:
        - Following terms have been removed: ‘catalogue’, ‘direct evaluation method’, ‘indirect evaluation method’ and ‘usability’.
        - Term ‘data quality basic measure’ has been moved to a new standard (currently under development) on data quality measures register (ISO/AWI 19157-3).
        - Following terms new terms have been added: ‘coverage’, ‘data quality’, ‘data quality measure’, ‘data quality unit’, ‘lineage’, ‘quality evaluation’, ‘requirement’ and ‘uncertainty’.
    - This standard’s terminology has been revised in compliance with existing standards for data quality.
- 5 Abbreviated terms
    - List of abbreviations updated and reformatted.
- General requirements for geographic information quality
    - New section with summary of all conformance requirement, recommendation and permission classes has been defined in this standard. Following ISO/TC211 guidelines for modular standards development, requirements, recommendation and permissions that were written directly into the clause paragraphs of ISO 19157:2013 were identified assigned an ISO/TC211 compliant URI (regulated by N5034).
- 6 Overview of data quality
    - The data quality model has been updated and the change is reflected in respective figures of its parts inserted throughout the standard.
    - Dropped usage of package prefixes for type name.
    - Aligned with ISO/FDIS 19131 and ISO 19115-1:2014.
    - Introduced set of requirements and recommendations to the standard according to the ISO/TC211 Best practice.
    - Clause has been renumbered as Clause 7.
- 7 Components of data quality
    - Dropped *UsabilityElement* from the quality model (Figure 2).
    - Added Clause 8.4 on extending standard quality model.

  - Moved *Metaquality* element to 8.3.7.
  - Revised Figure 8
  - Clause has been renumbered as Clause 8.
- 8 Data quality measures
  - Added Clause 9.3 on user defined data quality measures.
  - Clause has been renumbered as Clause 9.
- 9 Data quality evaluation
  - Minor revision to Figure 11.
  - Clause has been renumbered as Clause 10.
- 10 Data quality reporting
  - Revised Figure 13.
  - Clause has been renumbered as Clause 11.
- New Clause 12 Requirements for XML encoding
  - This is a new section defining requirements for XML encoding of data quality definition
- Annex A
  - New conformance tests replaced the previous tests
- Annex B
  - Revised Figure B.1.
  - Minor revisions in the text.
- Annex C
  - Revised tables to match diagrams in Clause 7 to 11.
- Annex D
  - Moved to ISO/AWI 19157-3.
- Annex E
  - Renamed to Annex D.
- Minor revisions in the text.
  - References to 'Usability Element' removed.
- Annex F
  - Renamed to Annex E.
  - Minor revisions in the text.
- Annex G
  - Moved to ISO/AWI 19157-3.
- Annex H
  - Moved to ISO/AWI 19157-3.
- Annex I
  - Renamed to Annex F.
  - Minor revisions in the text.
  - New examples added.
- Annex J
  - Renamed to Annex G.
- New normative Annex (Annex H in this standards) added detailing the XML schema implementation of this standard.
- New informative Annex (Annex I in this standard) added detailing this standard's backwards compatibility with ISO 19157:2013.
- Bibliography
  - Bibliography updated.

Bibliography

[1] ISO 2859 (all parts), *Sampling procedures for inspection by attributes*

[2] ISO 3534-2:2006, *Statistics — Vocabulary and symbols — Part 2: Applied statistics*

[3] ISO 3951-1:2006, *Sampling procedures for inspection by variables — Part 1: Specification for single sampling plans indexed by acceptance quality limit (AQL) for lot-by-lot inspection for a single quality characteristic and a single AQL*

[4] ISO 8000-2, *Data quality — Part 2: Vocabulary*

[5] ISO 8601-1, *Date and time — Representation for information interchange — Part: 1 Basic rules*

[6] ISO 9000:2015, *Quality management systems — Fundamentals and vocabulary*

[7] ISO 19101-1:2014, *Geographic information* — Part 1: Fundamentals

[8] ISO 19103:2015, *Geographic information — Conceptual schema language*

[9] ISO 19104:2016, *Geographic information — Terminology*

[10] ISO 19105:2022, *Geographic information — Conformance and testing*

[11] ISO 19108, *Geographic information — Temporal schema*

[12] ISO 19109:2015, *Geographic information — Rules for application schema*

[13] ISO 19110:2016, *Geographic information — Methodology for feature cataloguing*

[14] ISO 19115-1:2014, *Geographic information — Metadata — Part 1: Fundamental*

[15] ISO 19116: 2019 *Geographic information — Positioning services*

[16] ISO/DIS 19123-1, *Geographic information — Schema for coverage geometry and functions – Part 1: Fundamentals*

[17] ISO/FDIS 19131, *Geographic information — Data product specifications*

[18] ISO 19135-1:2015, *Geographic information — Procedures for item registration — Part 1: Fundamentals*

[19] ISO 19156:2011, *Geographic information — Observations and measurements*

[20] ISO 19160-3:2020, *Addressing — Part 3: Address data quality*

[21] ISO/IEC 2382-4:1999, *Information technology — Vocabulary — Part 4: Organization of data*

[22] ISO/IEC Directives, Part 1, Procedures for the technical work

[23] ISO/IEC GUIDE 98-3, Uncertainty of measurement — Part 3: Guide to expression of *uncertainty in measurement (GUM:1995)*

[24] ISO/IEC GUIDE 99, International vocabulary of metrology — Basic and general concepts and associated terms (VIM)

[25] ISO/IEC 25012:2008, Software engineering — Software product Quality Requirements and *Evaluation (SQuaRE) — Data quality model*

[26] ISO/IEC 25020:2019, Systems and software engineering — Systems and software Quality *Requirements and Evaluation (SQuaRE) — Quality measurement framework*

[27] ISO/IEC 25022:2016, Systems and software engineering — Systems and software Quality *Requirements and Evaluation (SQuaRE) — Measurement of quality in use*

[28] ISO/IEC 25023:2016, Systems and software engineering — Systems and software Quality *Requirements and Evaluation (SQuaRE) — Measurement of system and software product quality*

[29] ISO/IEC 25024:2015, Systems and software engineering — Systems and software Quality *Requirements and Evaluation (SQuaRE) — Measurement of data quality*

[30] ISO/IEC/IEEE 24765:2017, *Systems and software engineering — Vocabulary*

[31] ISO/TS 19115-3:2016, *Geographic information — Metadata — Part 1:* XML schema implementation for fundamental concepts

[32] ISO/TS 19139-1, *Geographic information* — XML schema implementation — Part 1: Encoding rules

[33] ISO/TS 19129:2009, *Geographic information — Imagery, gridded and coverage data framework*

[34] Lantmäteriet (2020), Quality description Laser data, version 1.3, Available at: https://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/hojddata/quality_description_lidar.pdf

[35] Lantmäteriet (2019), Hydrography download service, product description version 1.1, Available at: : https://www.lantmateriet.se/globalassets/kartor-och-geografisk-information/geodatatjanster/e_pb_hydrografi_nedladdning_v1.1.pdf

[36] OGC® Geospatial User Feedback Standard: Conceptual Model, Available at: http://docs.opengeospatial.org/is/15-097r1/15-097r1.html

