2013-04-15

Number of pages: 170

ISO/TC 211
Geographic information/Geomatics
Secretariat: SN (Norway)

Document type:	Other draft
ISO reference number:	19157
Title:	**Text for ISO/FDIS 19157 Geographic information — Data quality, as sent to the ISO Central Secretariat for issuing as FDIS.**
Status:	The DIS was approved with one negative vote, see N 3273. The draft will be checked by ISO before being issued as FDIS and parallel voting with CEN.
Source:	ISO/TC 211 Secretariat
Expected action:	Info
Reference:	N 3273, 3521
Email to secretary:	bjs@standard.no
Committee URL:	http://isotc.iso.org/livelink/livelink/open/tc211 and http://www.isotc211.org

(Blank)

Geographic information — Data quality

Information géographique — Qualité des données

Contents

Page

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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 19157 was prepared by Technical Committee ISO/TC 211, *Geographic information/Geomatics*.

This second edition cancels and replaces ISO/TS 19138:2006, ISO 19114:2003 and ISO 19113:2002, which have been technically revised.

Introduction

Geographic data is 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 is directly related to its quality. A user of geographic data may 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 reports 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 International Standard standardizes the components and structures of data quality measures and defines commonly used data quality measures.

This International Standard recognizes that a data producer and a data user may 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 International Standard is to provide principles for describing the quality for geographic data and concepts for handling quality information for geographic data, and a consistent and standard manner to determine and report a dataset's quality information. It aims also to provide guidelines for evaluation procedures of quantitative quality information for geographic data.

Geographic information — Data quality

1 Scope

This International Standard establishes the principles for describing the quality of geographic data. It

— defines components for describing data quality;

— specifies components and content structure of a register for data quality measures;

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

— establishes principles for reporting data quality.

This International Standard also defines a set of data quality measures for use in evaluating and reporting data quality. It 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 is of sufficient quality for their particular application.

This International Standard does not attempt to define minimum acceptable levels of quality for geographic data.

2 Conformance

Any product claiming conformance to this International Standard shall pass all the requirements described in the abstract test suite presented in Annex A as follows:

a) A data quality evaluation process shall pass the tests outlined in A.1;

b) Data quality metadata shall pass the tests outlined in A.2 and A.3;

c) A standalone quality report shall pass the tests outlined in A.4;

d) A data quality measure shall pass the tests outlined in A.5.

3 Normative references

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

ISO/TS 19103:2005, *Geographic information — Conceptual schema language*

ISO 19108:2002, *Geographic information — Temporal schema*

ISO 19115-1:2013, *Geographic information — Metadata — Part 1: Fundamentals*

ISO 19115-2:2009, *Geographic information — Metadata — Part 2: Extensions for imagery and gridded data*

ISO 19135:2005, *Geographic information — Procedures for item registration*

4 Terms and definitions

4.1
accuracy
closeness of agreement between a test result or measurement result and the true value

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

[SOURCE: ISO 3534-2:2006, 3.3.1, modified – Note 1 to entry has been added.]

4.2
catalogue
collection of *items* (4.18) or an electronic or paper document that contains information about the collection of items

[SOURCE ISO 10303-227:2005, 3.3.10]

4.3
conformance
fulfilment of specified requirements

[SOURCE: ISO 19105:2000, 3.8]

4.4
conformance quality level
threshold value or set of threshold values for data *quality* (4.21) results used to determine how well a *dataset* (4.8) meets the criteria set forth in its *data product specification* (4.6) or user requirements

4.5
correctness
correspondence with the *universe of discourse* (4.24)

4.6
data product specification
detailed description of a *dataset* (4.8) or *dataset series* (4.9) together with additional information that will enable it to be created, supplied to and used by another party

[SOURCE: ISO 19131:2007, 4.7]

4.7
data quality basic measure
generic data *quality* (4.21) measure used as a basis for the creation of specific data quality measures

Note 1 to entry: Data quality basic measures are abstract data types. They cannot be used directly when reporting data quality.

4.8
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* (4.15), is located physically within a larger dataset. Theoretically, a dataset can be as small as a single *feature* (4.11) or *feature attribute* (4.12) contained within a larger dataset. A hardcopy map or chart can be considered a dataset.

[SOURCE: ISO 19115-1:2013, 4.3]

4.9
dataset series
collection of *datasets* (4.8) sharing common characteristics

[SOURCE: ISO 19115-1:2013, 4.4]

4.10
direct evaluation method
method of evaluating the *quality* (4.21) of a *dataset* (4.8) based on inspection of the *items* (4.18) within the dataset

4.11
feature
abstraction of real world phenomena

Note 1 to entry: A feature may occur as a type or an instance. *Feature type* (4.15) or *feature instance* (4.13) will be used when only one is meant.

[SOURCE: ISO 19101:2002, 4.11]

4.12
feature attribute
characteristic of a *feature* (4.11)

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* (4.13) also has an attribute value taken from the value domain.

[SOURCE: ISO 19101:2002, 4.12, modified – Note 1 to entry has been added.]

4.13
feature instance
individual of a given *feature type* (4.15) having specified *feature attribute* (4.12) values

4.14
feature operation
operation that every instance of a *feature type* (4.15) may perform

[SOURCE: ISO 19110:2005, 4.5]

4.15
feature type
class of *features* (4.11) having common characteristics

[ISO 19156:2011, 4.7]

4.16
geographic data
data with implicit or explicit reference to a location relative to the Earth

[SOURCE: ISO 19109:2005, 4.12]

4.17
indirect evaluation method
method of evaluating the *quality* (4.21) of a *dataset* (4.8) based on external knowledge

Note 1 to entry: Examples of external knowledge are dataset lineage, such as production method or source data.

4.18
item
anything that can be described and considered separately

Note 1 to entry: An item can be any part of a *dataset* (4.8), such as a *feature* (4.11), feature relationship, *feature attribute* (4.12), or combination of these.

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

4.19
metadata
information about a resource

[SOURCE: ISO 19115-1:2013, 4.10]

4.20
metaquality
information describing the *quality* (4.21) of data quality

4.21
quality
degree to which a set of inherent characteristics fulfils requirements

[SOURCE: ISO 9000:2005, 3.1.1]

4.22
register
set of files containing identifiers assigned to *items* (4.18) with descriptions of the associated items

[SOURCE: ISO 19135:2005, 4.1.9]

4.23
standalone quality report
free text document providing fully detailed information about data *quality* (4.21) evaluations, results and measures used

4.24
universe of discourse
view of the real or hypothetical world that includes everything of interest

[SOURCE: ISO 19101:2002, 4.29]

5   Abbreviated terms

5.1   Abbreviations

ADQR   aggregated data quality results

AQL   acceptance quality limit [ISO 3534-2:2006]

RMSE   root mean square error

UML   Unified Modelling Language

XML   Extensible Markup Language

5.2 Package abbreviations

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

CI   Citation [ISO 19115-1:2013]

CT   Catalogues [ISO/TS 19139:2007]

DQ   Data Quality [ISO 19157]

DQM   Data Quality Measure [ISO 19157]

EX   Extent [ISO 19115-1:2013]

GF   General Feature [ISO 19109:2005]

MD   Metadata [ISO 19115-1:2013]

QE   Quality Extended [ISO 19115-2:2009]

RE   Registration [ISO 19135:2005]

6   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 19131:2007;

— specifying quality aspects in application schemas;

— evaluating data quality;

— reporting data quality.

NOTE 1   The development of application schemas is described in ISO 19109:2005.

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 shall be described using the data quality elements. 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.

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, the data element may be evaluated in a non-quantitative subjective way as a descriptive result for each element.

Some quality related information is provided by purpose, usage and lineage. This information is reported as metadata in conformance with ISO 19115-1:2013.

NOTE 2   Purpose describes the rationale for creating a dataset and contains information about its intended use, which may 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 recount 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 9.2.3).

This International Standard recognizes that quantitative 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 7.5.

Figure 1 provides an overview of data quality information.
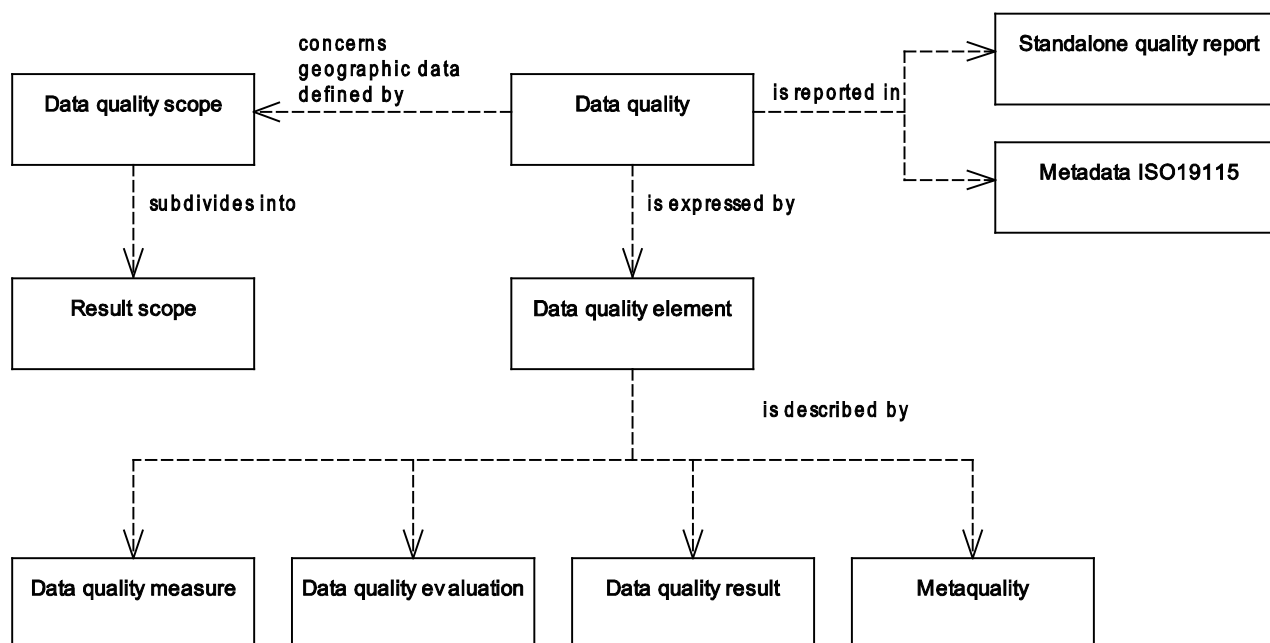
Figure 1 — Conceptual model of quality for geographic data

7   Components of data quality

7.1   Overview of the components

The components of data quality are described in this clause (Clause 7). Figure 2 presents an overview of the components and the connections between them. See the data dictionary defined in Annex C (normative) for more details about components and their attributes.
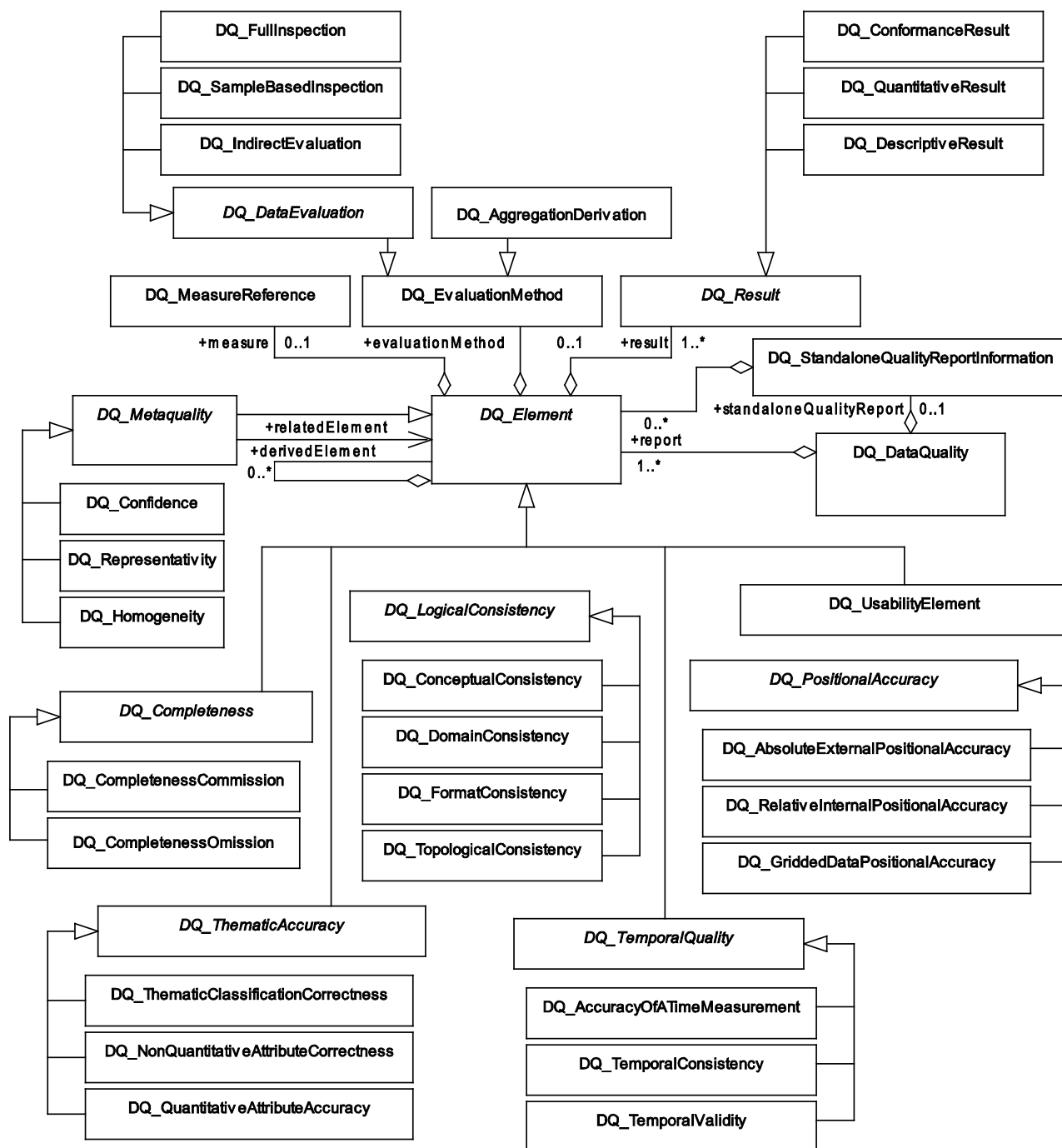
Figure 2 — Overview of the components of data quality

7.2 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.
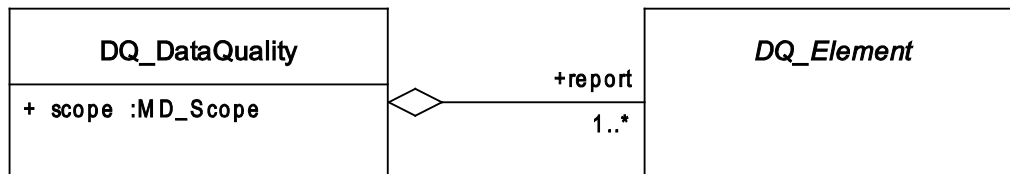
Figure 3 — Data quality unit

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 evaluated.

One data quality scope shall be specified for each data quality unit. One data quality report (metadata or standalone quality report) may encompass several data quality units, since scopes are often different for individual data quality elements. These different scopes may be, for example, spatially separate, overlapping or even sharing the same extents.

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 feature relationships);

   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).

7.3   Data quality elements

7.3.1   General

A data quality element is a component describing a certain aspect of the quality of geographic data and these have been organized into different categories. These categories are shown in Figure 4.
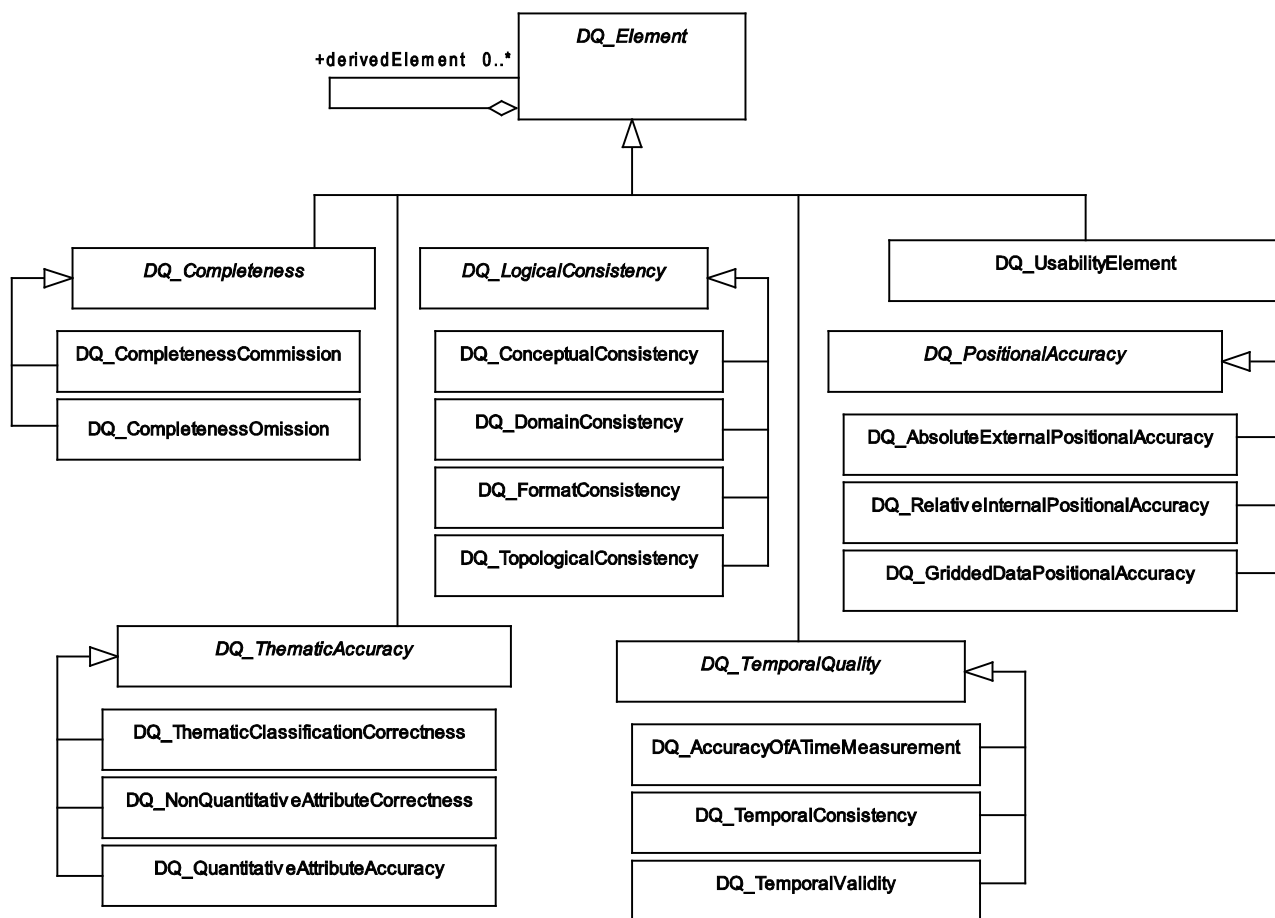
Figure 4 — Overview of the data quality elements

7.3.2 Completeness

Completeness is defined as the presence and absence of features, their attributes and relationships. It consists of two data quality elements:

— commission – excess data present in a dataset;

— omission – data absent from a dataset.

7.3.3 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). It consists of four data quality elements:

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

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

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

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

7.3.4 Positional accuracy

Positional accuracy is defined as the accuracy of the position of features within a spatial reference system. It consists of three data quality elements:

— absolute or external accuracy – closeness of reported coordinate values to values accepted as or being true;

— relative or internal accuracy – closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true;

— gridded data positional accuracy – closeness of gridded data spatial position values to values accepted as or being true.

7.3.5 Thematic accuracy

Thematic accuracy is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. It consists of three data quality elements:

— 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.

7.3.6 Temporal quality

Temporal quality is defined as the quality of the temporal attributes and temporal relationships of features. It consists of three data quality elements:

— accuracy of a time measurement – closeness of reported time measurements to values accepted as or known to be true;

— temporal consistency – correctness of the order of events;

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

NOTE 1   Time measurement can be either a defined point in time or a period.

NOTE 2   March 33 is an example of invalid data.

7.3.7 Usability element

Usability is based on user requirements. All quality elements may be used to evaluate usability. Usability evaluation may be based on specific user requirements that can not be described using the quality elements described above. In this case, the usability element shall be used to describe specific quality information about a dataset's suitability for a particular application or conformance to a set of requirements.

It is recommended when using the usability element, to use all applicable quality elements descriptors (see 7.4) and to define the quality measures applied in conformance with Clause 8 or Annex D, in order to provide precise details on the evaluation.

NOTE   For example, with this element, a data producer can show how a dataset, is suitable for various identified usages. This element can be used to declare the conformance of the dataset to a particular specification.

7.4 Descriptors of data quality elements

7.4.1 General

An evaluation of a data quality element is described by the following:

⎯ measure – the type of evaluation;

⎯ evaluation method – the procedure used to evaluate the measure;

⎯ result – the output of the evaluation.

These are shown in Figure 5, and are described in 7.4.2, 7.4.3 and 7.4.4.
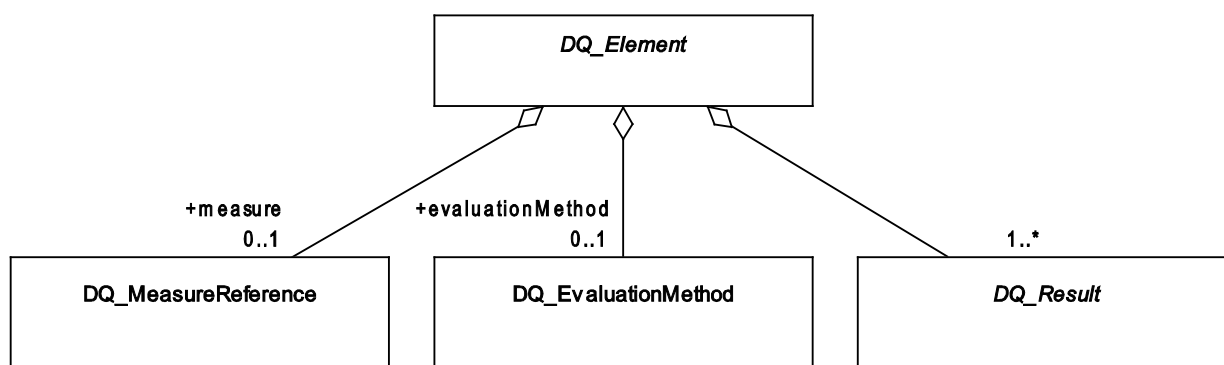
Figure 5 — Data quality element descriptors

7.4.2 Measure

A data quality element should refer to one measure only, by means of a measure reference (see Figure 6), providing an identifier of a measure fully described elsewhere (DQM_Measure.measureIdentifier, see 8.6.1) and/or providing the name and a short description of the measure.

NOTE   The whole description can be found within a measure register or catalogue, which may form part of a data product specification or a standalone quality report.
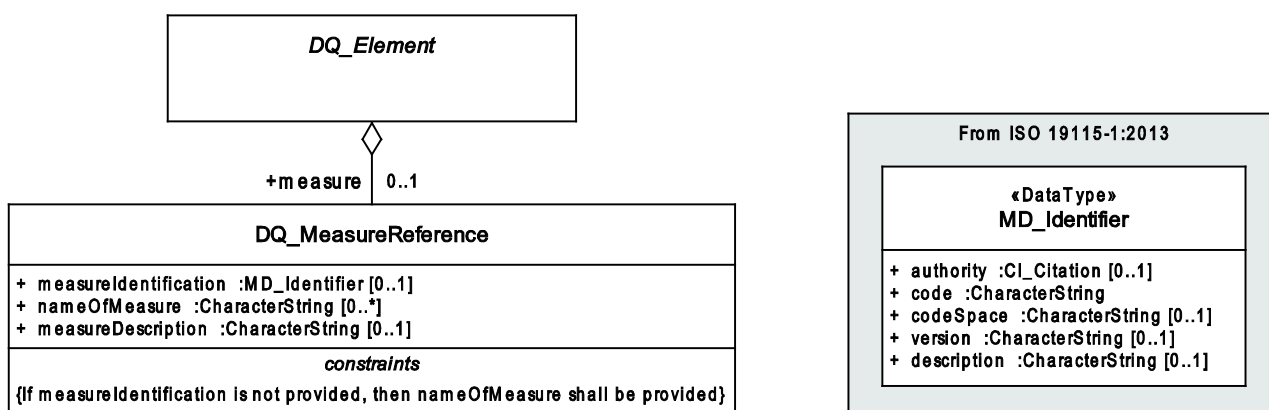
Figure 6 — Data quality measure reference

Data quality measures are further described in Clause 8 of this International Standard and Annex D contains a list of standardized data quality measures.

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

This International Standard recognizes that the quality of a dataset is measured using a variety of methods. A single data quality measure might 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 DQ_Element for each measure applied.

7.4.3   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 should be included for each applied data quality measure. 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 1   Data quality evaluation is further described in Clause 9.

NOTE 2   Examples of documentation are data product specifications, published articles or accepted industry standards.

One date or range of dates should be included for each evaluation. If the evaluation was carried out on non-consecutive dates, each single date should be included. The dates shall be in conformance with ISO 19108:2002.
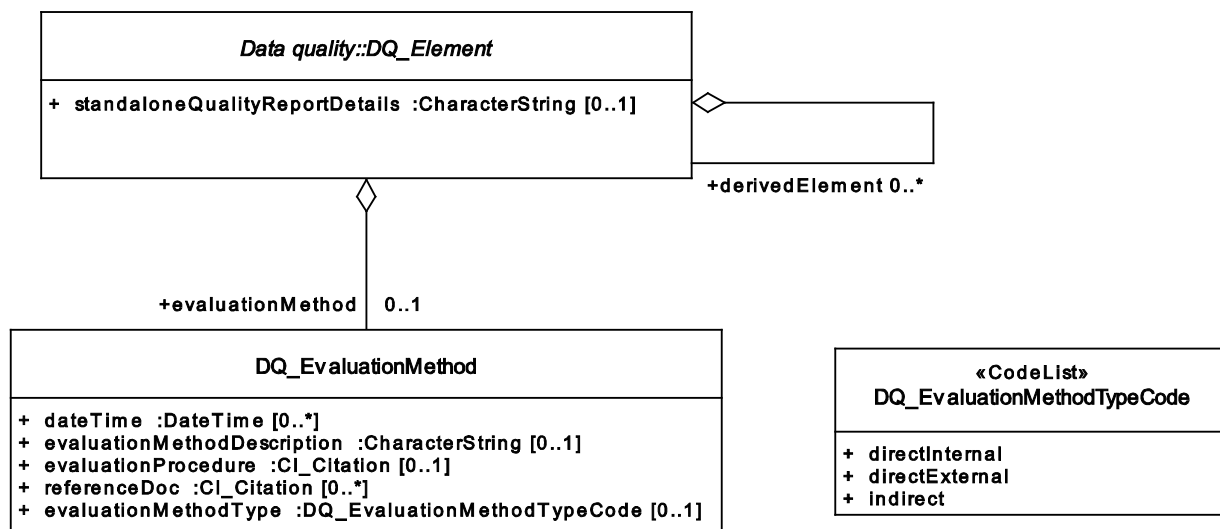
Figure 7 — Data quality evaluation method

7.4.4   Result

7.4.4.1   General

At least one data quality result shall be 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 1   Different types of results can be provided for the same data quality element.
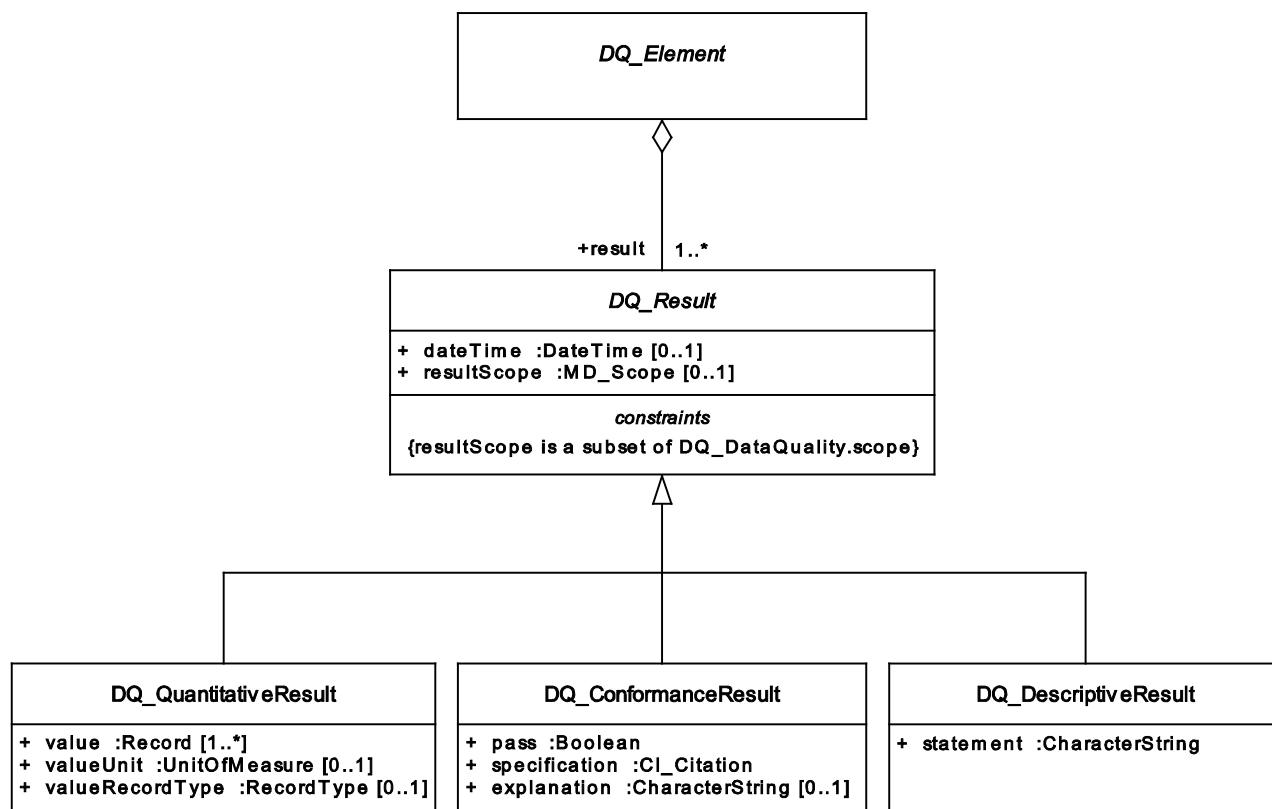
Figure 8 — Data quality result

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 (DQ_Element), several results with individual result scopes can be used.

NOTE 2   The result scope is a subset of the data quality scope (see 7.2).

EXAMPLE   A dataset contains features of identical type but whose positions have been established with separate 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 of 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).

7.4.4.2   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/TS 19103:2005. Its value changes depending on which implementation solution is used for providing the quantitative result. An example of XML implementation for recordType is provided in ISO/TS 19139:2007.

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: matrix (nxn). The value attribute of the quantitative result provide the result matrix itself, within a numeric encoding using a particular XML

type called MatrixType (for example). The attribute valueRecordType provide 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.

One value unit should be included for each result, if applicable.

EXAMPLE 3   Possible distance units: metre, centimetre, millimetre.

EXAMPLE 4   Measure "Rate of excess items" (see Table D.3) is used to evaluate the number of excess items in the dataset in relation to the number of items that should have been 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 with 100. In this example the value unit is "%".

7.4.4.3   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 acceptable conformance quality level.

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 may be specified in suitable reference documentation such as the 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.

More than one data quality conformance result may be provided for the same measure if evaluation has been performed against conformance levels originating from different sources.

7.4.4.4   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.

7.4.4.5   Coverage result

A coverage result is the result of a data quality evaluation, organized as a coverage. This is documented in ISO 19115-2:2009.

7.5   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 E.5.3 for an example of metaquality evaluation.

Metaquality may be described using the following elements, represented in Figure 9:

— 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 9.2.2 and Annex F.

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

NOTE 3   Homogeneity consists in comparing the evaluation results of several segments of a global dataset. This comparison can be expressed using root mean square errors for example. 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.
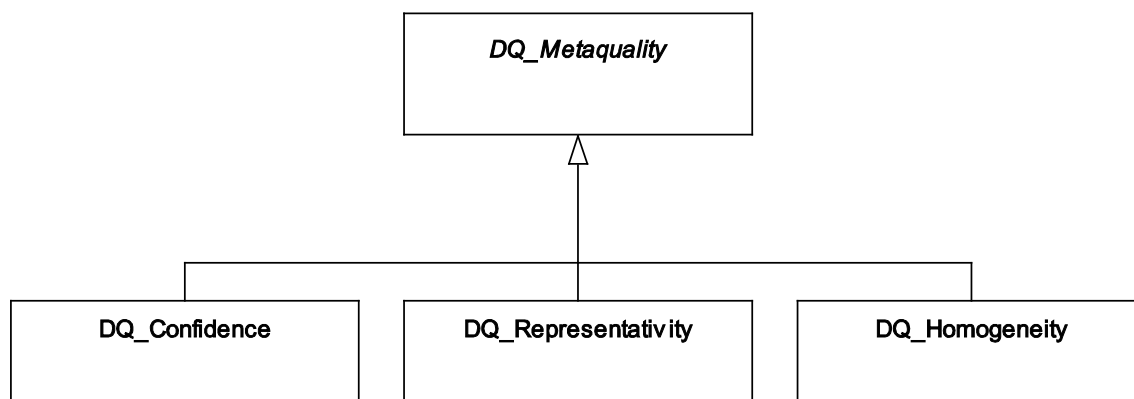
Figure 9 — Metaquality elements

7.6   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 7.4 and Figure 10). Additionally the following descriptor shall be used:

— related quality element.

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

See E.5.3 for an example of metaquality evaluation.
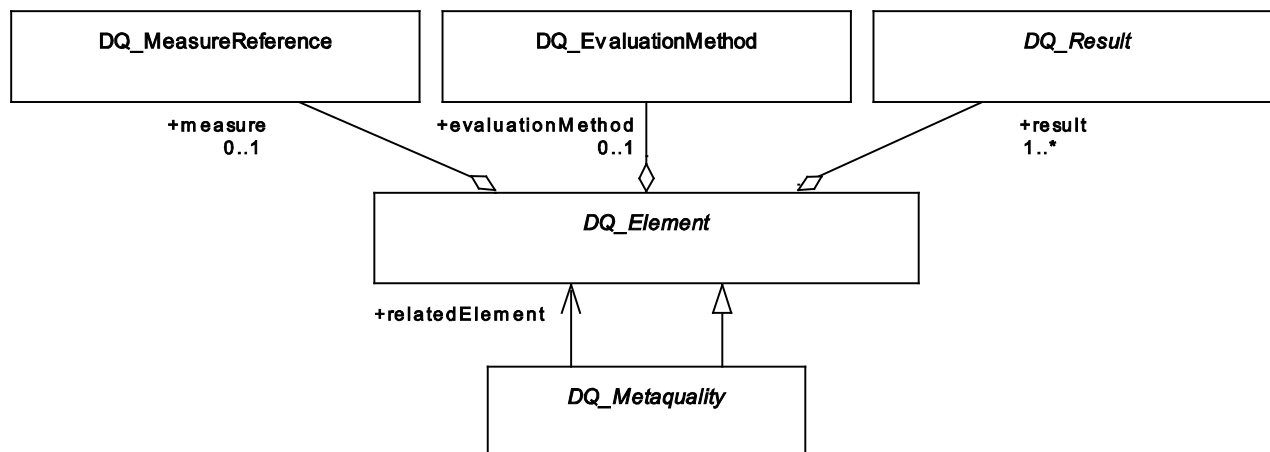
Figure 10 — Metaquality descriptors

8 Data quality measures

8.1 General

To facilitate dataset comparisons, it is necessary that the results in the data quality reports are expressed in a comparable way and that there is a common understanding of the data quality measures that have been used. In order to make evaluations and data quality reports (metadata or a standalone quality report) from different sources comparable, standardized data quality measures described in Annex D shall be used when possible.

8.2 Standardized data quality measures

A list of standardized data quality measures is given in Annex D. Each data quality measure of this list contains all the required components as specified in Clause 8. Multiple measures are 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. Measures from this list should be used when implementing the standard.

Any register established to manage standardized data quality measures, shall be in conformance with ISO 19135:2005.

8.3 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 may be cases where the user of this International Standard has to devise other data quality measures. When possible, these measures shall be defined using the data quality basic measures provided in Annex G and the measure shall be defined using the structure given in this clause, Clause 8.

8.4 Catalogue of data quality measures

Catalogues of data quality measures may be provided associated with metadata or made available online to fully describe the measures referenced in the data quality report of the data evaluated.

The catalogue may contain the set of measures used in one or several data quality reports with all required components for data quality measures as specified in this International Standard.

The catalogue (as a register) enables the user to describe the measure, and store the information in order to be able to refer to it each time needed, instead of re-describing the measure within a data quality report.

Annex H describes the structure of a measure catalogue. ISO/TS 19139:2007 provides an XML mechanism to associate the catalogue to a metadata set.

8.5 List of components

Each data quality measure is described by the following components:

⎯ measure identifier (8.6.1);

⎯ name (8.6.2);

⎯ alias (8.6.3);

⎯ element name (8.6.4);

⎯ basic measure (8.6.5);

⎯ definition (8.6.6);

⎯ description (8.6.7);

⎯ parameter (8.6.8);

⎯ value type (8.6.9);

⎯ value structure (8.6.10);

⎯ source reference (8.6.11);

⎯ example (8.6.12).

Figure 11 represents the components of data quality measures.
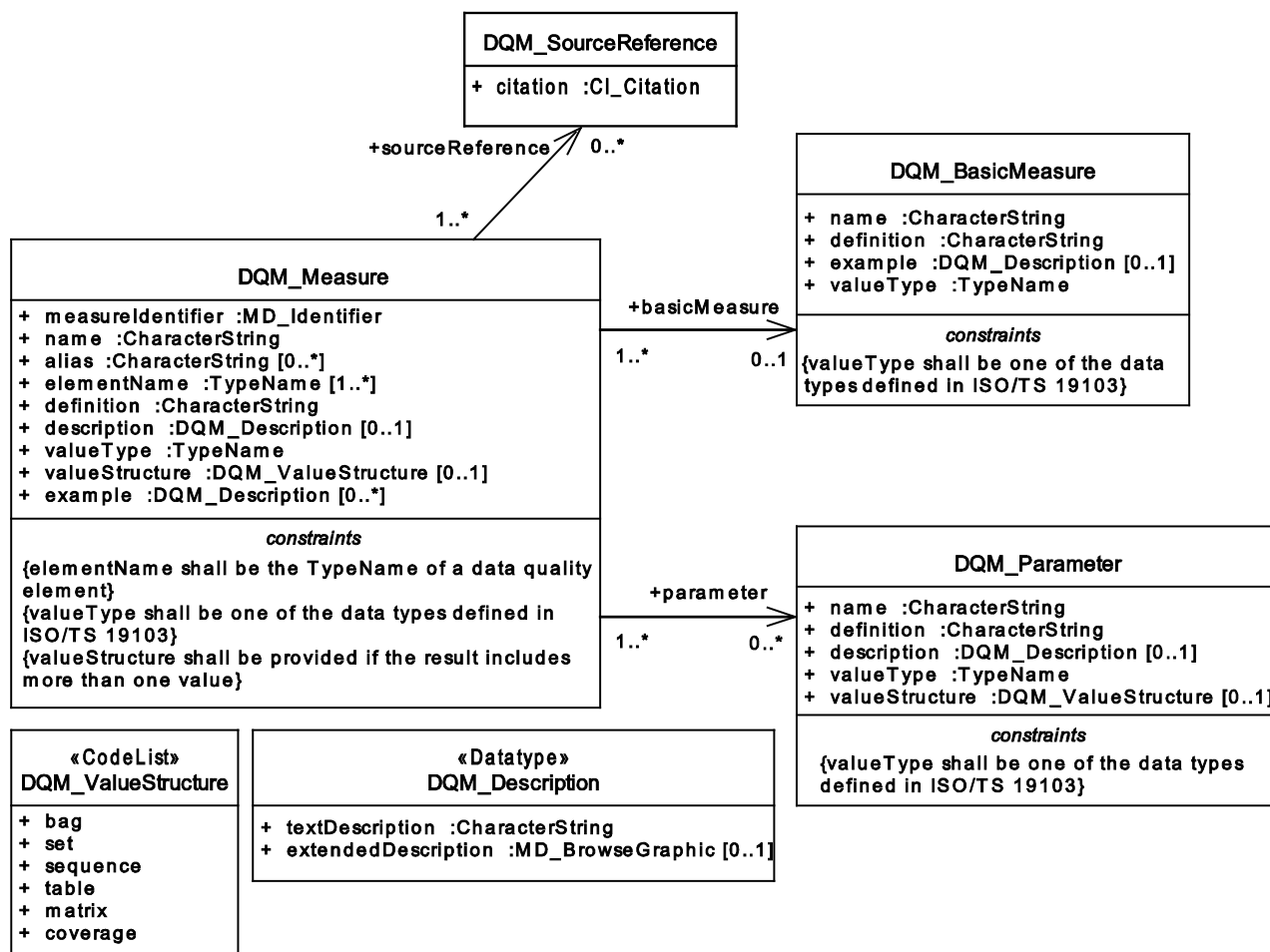
Figure 11 — Data quality measures

8.6 Component details

8.6.1 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 7.4.2)

8.6.2 Name

Name is the name of the measure.

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

8.6.3 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.

8.6.4 Element name

Element name is the name of the data quality element (see 7.3 and 7.5) to which a measure applies. More than one element name may be provided.

8.6.5 Basic measure

If a measure is based on one of the basic measures, it shall be described by its name, definition and value type. Basic measures are identified by their names.

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 should be defined in terms of basic measures.

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.

NOTE   The basic measures are defined in Annex G.

8.6.6 Definition

Definition is the fundamental concept of the measure.

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

8.6.7 Description

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 only can be reported as correct or incorrect.

8.6.8 Parameter

Parameter is an auxiliary variable used by the measure. It shall includes name, definition and value type, More than one measure parameter may be provided.

NOTE   See Table D.66 for an example of Parameter.

8.6.9 Value type

Value type is the data type used for reporting the result of the measure. The data types defined in ISO/TS 19103:2005 shall be used.

8.6.10 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.

8.6.11 Source reference

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

When a measure, for which additional information is provided in an external source, is added to the list of standardized measures, a reference to that source may be provided here.

8.6.12 Example

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

9 Data quality evaluation

9.1 The process for evaluating data quality

9.1.1 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.

9.1.2 The process flow

The quality evaluation process is a sequence of steps followed to produce a quality evaluation result. Figure 12 illustrates a possible workflow for evaluating data quality; see also Annex E for a description of the concepts for evaluating and reporting data quality.

When the geographic data evaluated is heterogeneous with different quality for different parts, tests should be applied to suitable parts of the data.
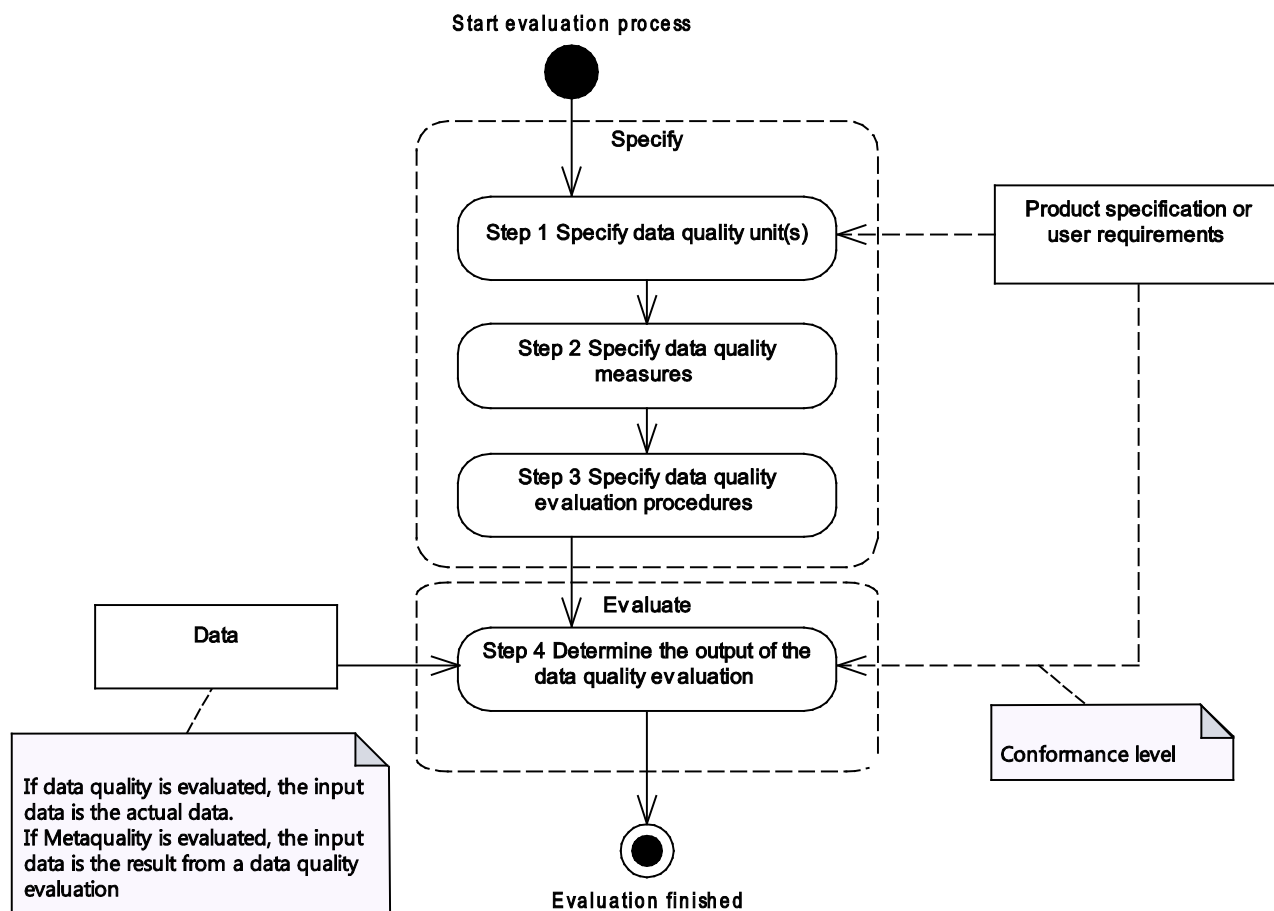
Figure 12 — Evaluating data quality

9.1.3 Process steps

Table 1 specifies the process steps.

Table 1 — 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 7.2. 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 7.3, and Annex I provides guidelines for the use of quality elements
2	Specify data quality measures	If applicable[a] a measure should be specified for each data quality element. Annex D contains a list of Data quality measures.
3	Specify data quality evaluation procedures	A data quality evaluation procedure consists of applying one or more evaluation methods
4	Determine the output of the data quality evaluation	A result is the output of applying the evaluation
[a] If no measure can be identified, a descriptive result may be provided.		

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.

9.2   Data quality evaluation methods

9.2.1   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. Direct evaluation methods 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 13 shows the classes used describing the evaluation methods.

NOTE   Lineage is described in ISO 19115-1:2013.
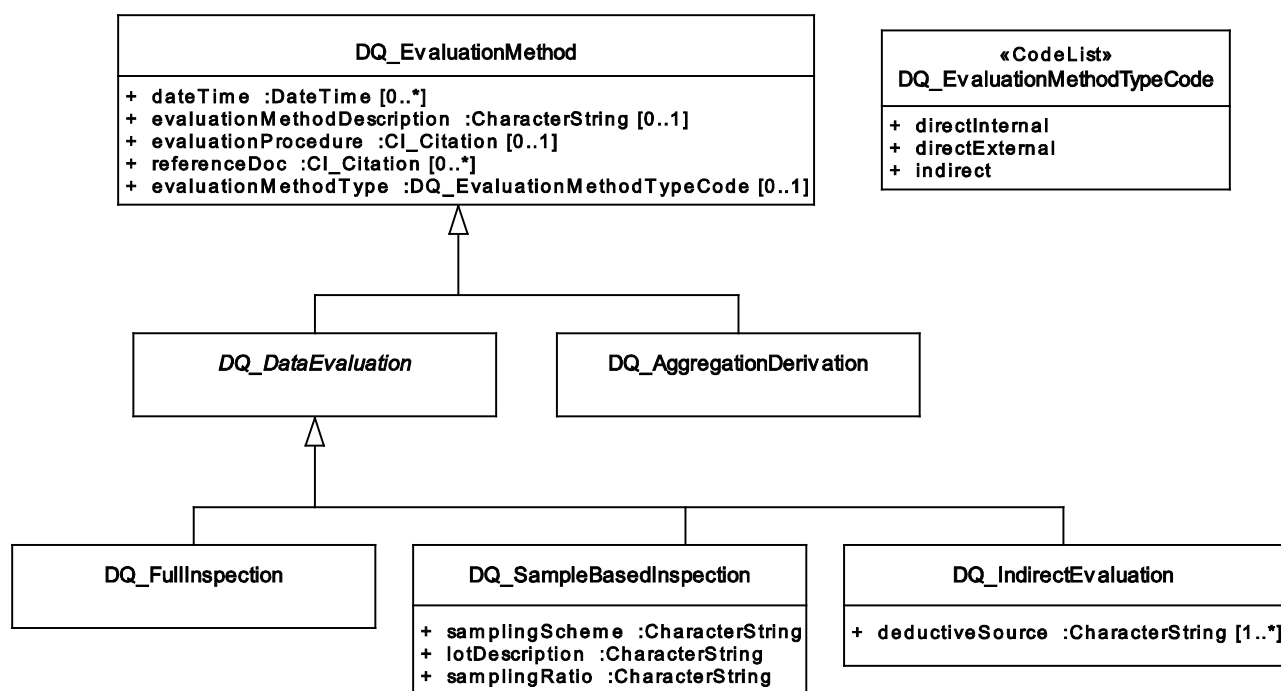
Figure 13 — Data quality evaluation methods

9.2.2   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 is data accepted as representing the universe of discourse.

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

⎯ full inspection;

⎯ sampling.

Full inspection tests every item in the population specified by the data quality scope.

NOTE 2   Full inspection is most appropriate for small populations or for tests that can be accomplished by automated means.

Sampling means that tests are performed on subsets of the geographic data defined by the data quality scope.

NOTE 3   Examples of sampling methods are given in Annex F.

9.2.3   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, but is not limited to one or more non-quantitative quality information usage, lineage and purpose (see ISO 19115-1:2013) 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. If indirectly evaluated data quality has been reported, it should be accompanied by a description on how it was determined.

In some cases it might 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 7.4.4.4.

9.3   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 J. How to report Aggregation is described in 10.2.1 and Annex E.

NOTE 3   How to report Derivation is described in 10.2.2 and Annex E.

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.

10 Data quality reporting

10.1 General

Data quality shall be reported as metadata in compliance with Clause 7, Clause 10, Annex C, ISO 19115-1:2013 and ISO 19115-2:2009.

In order to provide more details than reported as metadata, a standalone quality report may additionally be created. Its structure is free. However, the standalone quality report shall not replace the metadata. The metadata should provide a reference to the standalone quality report when it exists (see Figure 14).

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

NOTE 2   See E.4 for examples of how to report data quality.
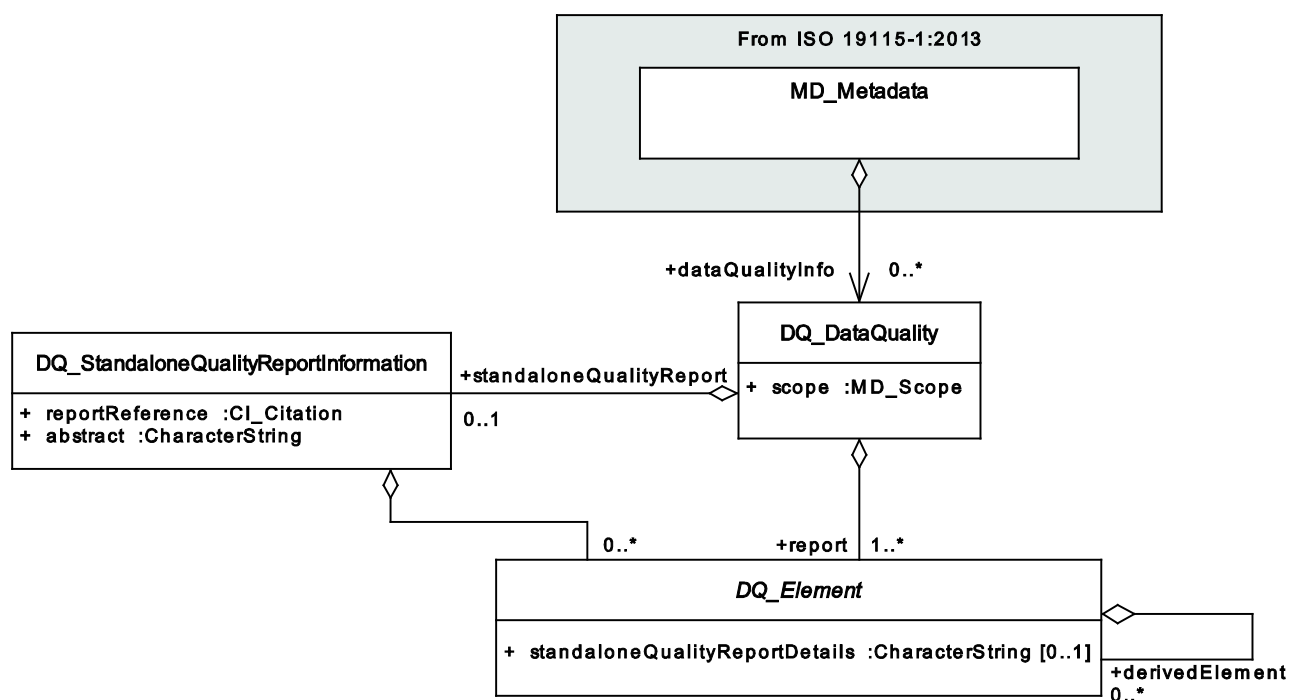
Figure 14 — Reporting data quality

10.2 Particular cases

10.2.1 Reporting Aggregation (aggregated results)

Where the result has been aggregated, a standalone quality report should be provided to complete the information provided in the metadata. Within this standalone quality report, fully detailed information on the original result (with measure(s) and evaluation procedure(s)), aggregated result and aggregation method should be provided.

Within the metadata:

1) 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 E.4.1.2 and E.4.1.3 for examples.

2) When several quality results for different data quality elements are aggregated into a single result, this should be reported in metadata as a result for the usability element (DQ_UsabilityElement), see E.4.1.4 for an example.

In both cases, in metadata, at least a reference to the original data quality results shall be provided for an aggregated result, and information on the aggregation measure and aggregation method may be provided.

10.2.2 Reporting Derivation (derived results)

When derived results only are reported in metadata, a standalone quality 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 standalone quality report and the original data quality result.

EXAMPLE   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 standalone quality report.

10.2.3 Reference to the original data quality result

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 standaloneQualityReportDetails references the part of the standalone quality report where the original result(s) are described.

Annex A
(normative)

Abstract test suites

A.1 Test case identifier: Quality evaluation process

a) Test purpose: To validate the data quality evaluation process.

b) Test method: Check whether the quality evaluation process includes all of the steps specified in 9.1.3. This implies:

    1) Identify the data product specification statements or the user requirements relevant to data quality and use them to identify the applicable data quality elements and their appropriate scope. Compare the applicable data quality elements with the data quality elements evaluated to ensure that all applicable data quality elements have been identified and evaluated on the appropriate scope.

    2) Check that the data quality measure applied for each data quality evaluation is appropriate regarding the data product specification statement or the user requirements.

    3) Check that the data quality evaluation procedure applied for each data quality evaluation is appropriate regarding the data product specification statement or the user requirements.

c) Reference: 9.1.

d) Test type: Basic.

A.2 Test case identifier: Data quality metadata

a) Test purpose: To verify that the data quality metadata is modelled according to the UML models and the data dictionary.

b) Test method: Check whether the metadata contains the appropriate data quality components and follows the occurrences rules for each component.

c) Reference: Clause 7, Clause 10 and Annex C.

d) Test type: Basic.

A.3 Test case identifier: Metadata conformity

a) Test purpose: To verify that the data quality metadata is reported in conformance with ISO 19115-1:2013 and ISO 19115-2:2009.

b) Test method: Check abstract test suites provided in ISO 19115-1:2013, A.2.1, A.2.2, A.2.3, A.2.4, A.2.5

c) Reference: ISO 19115-1:2013, A.2.1, A.2.2, A.2.3, A.2.4, A.2.5

d) Test type: Basic.

A.4 Test case identifier: Standalone quality report

a) Test purpose: To verify that the standalone quality report includes sections on all appropriate aspects of quality and that the description of all components of data quality follows the rules defined in this International Standard.

b) Test method: Check whether the standalone quality report contains all the relevant components.

c) Reference: Clause 7 and Clause 10.

d) Test type: Basic.

A.5 Test case identifier: Data quality measures

a) Test purpose: To verify that a data quality measure is structurally and semantically well-defined.

b) Test method: Check whether the data quality measures used are described as specified in Clause 8, and modelled according to the UML model and the data dictionary.

c) Reference: Clause 8 and Annex C.

d) 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 19115-1:2013). 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. If these are changed then quality evaluation should be repeated against the changed specification or requirements. Care should be taken when comparing quality results where the universe of discourse is different. 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.
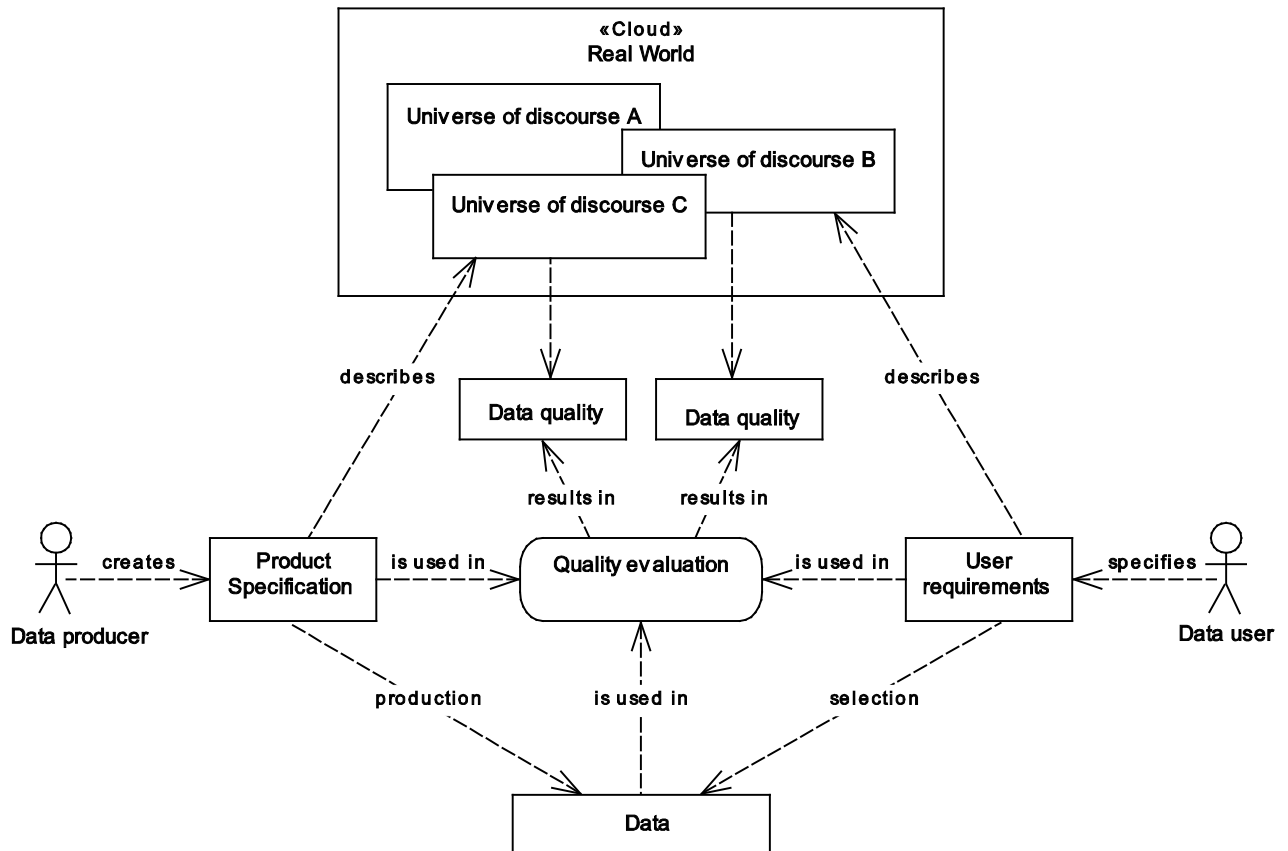
Figure B.1 — Framework of data quality concepts

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

A dataset may belong to a dataset series meaning that all of this 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 data 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.

NOTE   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 the lifecycle of a dataset. Data quality concepts recognize that 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 International Standard 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 I. 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.

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 lifecycle 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 10, see also the example in Annex E 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 have 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 exist 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 data set;

— 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, 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 International Standard 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

B.4.3.1 Hierarchy principle

This International Standard 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 B.1. If the data quality differs between upper and lower level, then supplemental information should be provided at lower level.

Table B.1 — Hierarchical levels

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

NOTE   Quality for an instance of feature, feature attribute or associations between features can be reported as an attribute for that instance as defined in ISO 19109:2005.

B.4.3.2   Metadata and standalone quality report

B.4.3.2.1   General

Quality information may be reported as metadata and as a standalone quality 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 standalone quality report may be used to provide fully detailed information about the data quality evaluation. The standalone quality report is to be provided attached to the dataset or product for direct human reading.

For example, in the case of aggregation of different quality results, the standalone quality 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 standalone quality report.

B.4.3.2.2   Reporting quality information as metadata

The class MD_Metadata, defined in ISO 19115-1:2013, aggregates zero, one or several data quality units (instances of the class DQ_DataQuality, as specified in this International Standard), see Figure B.2.
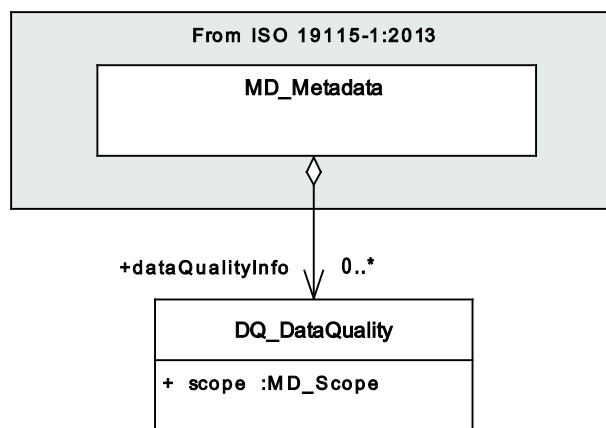
Figure B.2 — Data quality information

B.4.3.2.3   Reporting quality information within a standalone quality 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 standalone quality 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 catalogue.

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

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 7, 8, 9 and 10. 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 subword starting with a capital letter (example: XnnnYmmm). Class names are unique within the entire data dictionary of this International Standard. Class attribute names are unique within a class, not the entire data dictionary of this International Standard. 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 International Standard.

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 need not 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/TS 19103:2005, 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 or the use of free text. "Free text" indicates that no restrictions are placed on the content of the field. Integer-based codes shall be used to represent values for domains containing codelists.

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.

UML model shown in Figure 3 and Figure 15.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
1.	DQ_DataQuality	quality information for the data specified by a data quality scope	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	quantitative quality information for the data specified by the scope	M	N	Association	DQ_Element (Abstract class) (C.2.1.2)
4.	Role name: standaloneQualityReport	reference to external standalone quality report	O	1	Association	DQ_StandaloneQualityReportInformation (C.2.1.6)

C.2.1.2   Data quality element

UML model shown in Figure 4, Figure 5, Figure 10 and Figure 14.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
5.	DQ_Element	aspect of quantitative quality information	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (DQ_DataQuality) (Abstract class)	Lines 6-10

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
6.	standaloneQualityReportDetails	Clause in the standaloneQualityReport where this data quality element or any related data quality element (original results in case of derivation or aggregation) is described	O	1	CharacterString	Free Text
7.	*Role name:* measure	reference to measure used	O	1	Association	DQ_MeasureReference (C.2.1.3)
8.	*Role name:* evaluationMethod	evaluation information	O	1	Association	DQ_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 a specified acceptable conformance quality level	M	N	Association	DQ_Result (Abstract class) (C.2.1.5)
10.	*Role name:* derivedElement	In case of aggregation or derivation, indicates the original element	O	N	Association	DQ_Element (Abstract class) (C.2.1.2)
11.	*DQ_Completeness*	presence and absence of features, their attributes and their relationships	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Element) (Abstract class)	Lines 6-10
12.	DQ_Completeness 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 (DQ_Completeness)	Lines 6-10
13.	DQ_CompletenessOmission	data absent from the dataset, as described by the scope	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Completeness)	Lines 6-10
14.	*DQ_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 (DQ_Element) (Abstract class)	Lines 6-10
15.	DQ_ConceptualConsistency	adherence to rules of the conceptual schema	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Logical Consistency)	Lines 6-10
16.	DQ_DomainConsistency	adherence of values to the value domains	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Logical Consistency)	Lines 6-10

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
17.	DQ_FormatConsistency	degree to which data is 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 (DQ_Logical Consistency)	Lines 6-10
18.	DQ_TopologicalConsistency	correctness of the explicitly encoded topological characteristics of the dataset as described by the scope	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Logical Consistency)	Lines 6-10
19.	*DQ_PositionalAccuracy*	accuracy of the position of features	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Element) (Abstract class)	Lines 6-10
20.	DQ_AbsoluteExternal PositionalAccuracy	closeness of reported coordinate values to values accepted as or being true	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Positional Accuracy)	Lines 6-10
21.	DQ_RelativeInternalPosition alAccuracy	closeness of the relative positions of features in the scope to their respective relative positions accepted as or being true	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Positional Accuracy)	Lines 6-10
22.	DQ_GriddedDataPositional Accuracy	closeness of gridded data position values to values accepted as or being true	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Positional Accuracy)	Lines 6-10
23.	*DQ_TemporalQuality*	accuracy of the temporal attributes and temporal relationships of features	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Element) (Abstract class)	Lines 6-10
24.	DQ_AccuracyOfATime Measurement	correctness of the temporal references of an item (reporting of error in time measurement)	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Temporal Quality)	Lines 6-10
25.	DQ_TemporalConsistency	correctness of ordered events or sequences, if reported	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Temporal Quality)	Lines 6-10
26.	DQ_TemporalValidity	validity of data specified by the scope with respect to time	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Temporal Quality)	Lines 6-10
27.	DQ_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 (DQ_Element) (Abstract class)	Lines 6-10

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
28.	DQ_ThematicClassification Correctness	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 (DQ_Thematic Accuracy)	Lines 6-10
29.	DQ_NonQuantitativeAttribute Correctness	correctness of non-quantitative attributes	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Thematic Accuracy)	Lines 6-10
30.	DQ_QuantitativeAttribute Accuracy	accuracy of quantitative attributes	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Thematic Accuracy)	Lines 6-10
31.	DQ_UsabilityElement	degree of adherence of a dataset to a specific set of requirements	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Element)	Lines 6-10
32.	*DQ_Metaquality*	information about the reliability of data quality results	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Element) (Abstract class)	Lines 33 and 6-10
33.	*Role name*: derivedElement	derived element	M	1	Association	DQ_Element (Abstract class) (C.2.1.2)
34.	DQ_Confidence	trustworthiness of a data quality result	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Metaquality)	Lines 33 and 6-10
35.	DQ_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 (DQ_Metaquality)	Lines 33 and 6-10
36.	DQ_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 (DQ_Metaquality)	Lines 33 and 6-10

C.2.1.3   Measure reference

UML model shown in Figure 6.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
37.	DQ_MeasureReference	reference to the measure used	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (DQ_Element)	Lines 38-40
38.	measureIdentification	Identifier of the measure, value uniquely identifying the measure within a namespace	O	1	Class	MD_Identifier <<DataType>> (see ISO 19115-1:2013, B.3.3.3)
39.	nameOfMeasure	name of the test applied to the data	C/ if measureIdentification not documented	N	CharacterString	Free text
40.	measureDescription	description of the measure	O	1	CharacterString	Free text

C.2.1.4   Data quality evaluation

UML model shown in Figure 7 and Figure 13.

	Name / Role Name	Definition	Obligation / Condition	Occurrence	Data type	Domain
41.	DQ_EvaluationMethod	Description of the evaluation method and procedure applied	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (DQ_Element)	Lines 42-46
42.	evaluationMethodType	type of method used to evaluate quality of the data	O	1	Class	DQ_EvaluationMethodType Code <<CodeList>> (C.3.2)
43.	evaluationMethodDescription	description of the evaluation method	O	1	CharacterString	Free text
44.	evaluationProcedure	reference to the procedure information	O	1	Class	CI_Citation <<DataType>> (see ISO 19115-1:2013, B.3.2.1)

	Name / Role Name	Definition	Obligation / Condition	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 19115-1:2013, B.3.2.1)
46.	dateTime	date or range of dates on which a data quality measure was applied	O	N	Class	DateTime (see ISO/TS 19103:2005)
47.	*DQ_DataEvaluation*	data evaluation method	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_EvaluationMethod) (Abstract class)	Lines 42-46
48.	DQ_FullInspection	full inspection	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_DataEvaluation)	Lines 42-46
49.	DQ_IndirectEvaluation	indirect evaluation	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_DataEvaluation)	Lines 42-46 and 50
50.	deductiveSource	information on which data are used as sources in deductive evaluation method	M	1	CharacterString	Free text
51.	DQ_SampleBasedInspection	sample based inspection	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_DataEvaluation)	Lines 42-46 and 52-54
52.	samplingScheme	information of the type of sampling scheme and description of the sampling procedure	M	1	CharacterString	Free text
53.	lotDescription	information of how lots are defined	M	1	CharacterString	Free text
54.	samplingRatio	information on how many samples on average are extracted for inspection from each lot of population	M	1	CharacterString	Free text
55.	DQ_AggregationDerivation	Aggregation or derivation method	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Evaluation)	Lines 42-46

C.2.1.5 Data quality result

UML model shown in Figure 8.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
56.	*DQ_Result*	generalization of more specific result classes	Use obligation from referencing object	Use maximum occurrence from referencing object	Aggregated Class (DQ_Element) (Abstract class)	Lines 57-58
57.	resultScope	scope of the result	O	1	Class	MD_Scope (ISO 19115-1)
58.	dateTime	date when the result was generated	O	1	Class	DateTime (see ISO/TS 19103:2005)
59.	DQ_ConformanceResult	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 (DQ_Result)	Lines 60-62 and 57-58
60.	specification	citation of data product specification or user requirement against which data is being evaluated	M	1	Class	CI_Citation <<DataType>> (see ISO 19115-1:2013, B.3.2.1)
61.	explanation	explanation of the meaning of conformance for this result	O	1	CharacterString	Free text
62.	pass	indication of the conformance result where 0 = fail and 1 = pass	M	1	Boolean	1 = yes 0 = no
63.	DQ_QuantitativeResult	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 (DQ_Result)	Lines 64-66 and 57-58
64.	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/TS 19103:2005)
65.	valueUnit	value unit for reporting a data quality result	O	1	Class	UnitOfMeasure (see ISO/TS 19103:2005)
66.	valueRecordType	value type for reporting a data quality result, depends of the implementation	O	1	Class	RecordType <<Metaclass>> (see ISO/TS 19103:2005)

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
67.	DQ_DescriptiveResult	data quality descriptive result	Use obligation from referencing object	Use maximum occurrence from referencing object	Specified Class (DQ_Result)	Lines 68 and 57-58
68.	statement	textual expression of the descriptive result	M	1	CharacterString	Free text

C.2.1.6   Standalone quality report information

UML model shown in Figure 14.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
69.	DQ_StandaloneQualityReportInformation	reference to an external standalone quality report	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 70-71
70.	reportReference	reference to the associated standalone quality report	M	1	Class	CI_Citation <<DataType>> (see ISO 19115-1:2013, B.3.2.1)
71.	abstract	abstract for the associated standalone quality report	M	1	CharacterString	FreeText

C.2.2 Data quality measure

C.2.2.1   General

The UML model for measures information is shown in Figure 11.

C.2.2.2 Data quality measures

UML model shown in Figure 11.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
72.	DQM_Measure	Data quality measure	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 73-84
73.	measureIdentifier	value uniquely identifying the measure within a namespace	M	1	Class	MD_Identifier <<DataType>> (see ISO 19115-1:2013, B.3.3.3)
74.	Name	name of the data quality measure applied to the data	M	1	CharacterString	Free text
75.	alias	another recognized name, an abbreviation or a short name for the same data quality measure	O	N	CharacterString	Free text
76.	elementName	name of the data quality element for which quality is reported	M	N	Class	TypeName <<type>> (see ISO/TS 19103:2005)
77.	definition	definition of the fundamental concept for the data quality measure	M	1	CharacterString	Free text
78.	description	description of the data quality measure, including all formulae and/or illustrations needed to establish the result of applying the measure	C/if the definition is not sufficient for the understanding of the data quality measure concept	1	Class	DQM_Description <<Datatype>> (C.2.2.5)
79.	valueType	value type for reporting a data quality result (shall be one of the data types defined in ISO/TS 19103:2005)	M	1	Class	TypeName <<type>> (see ISO/TS 19103:2005)
80.	valueStructure	structure for reporting a complex data quality result	O	1	Class	DQM_ValueStructure << CodeList>> (C.3.3)
81.	example	illustration of the use of a data quality measure	O	N	Class	DQM_Description (C.2.2.5)
82.	*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	DQM_BasicMeasure (C.2.2.3)

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
83.	*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	DQM_SourceReference (C.2.2.6)
84.	*Role name:* parameter	auxiliary variable used by the data quality measure, including its name, definition and optionally its description	C/if required	N	Association	DQM_Parameter (C.2.2.4)

C.2.2.3   Data quality basic measure

UML model shown in Figure 11.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
85.	DQM_BasicMeasure	data quality basic measure	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 86-89
86.	name	name of the data quality basic measure applied to the data	M	1	CharacterString	Free text
87.	definition	definition of the data quality basic measure	M	1	CharacterString	Free text
88.	example	illustration of the use of a data quality measure	O	1	Class	DQM_Description <<Datatype>> (C.2.2.5)
89.	valueType	value type for the result of the basic measure (shall be one of the data types defined in ISO/TS 19103:2005)	M	1	Class	TypeName <<type>> (see ISO/TS 19103:2005)

C.2.2.4   Data quality parameter

UML model shown in Figure 11.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
90.	DQM_Parameter	data quality parameter	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 91-95
91.	name	name of the data quality parameter	M	1	CharacterString	Free text
92.	definition	definition of the data quality parameter	M	1	CharacterString	Free text
93.	description	description of the data quality parameter	O	1	Class	DQM_Description <<Datatype>> (C.2.2.5)
94.	valueType	value type of the data quality parameter (shall be one of the data types defined in ISO/TS 19103:2005)	M	1	Class	TypeName <<type>> (see ISO/TS 19103:2005)
95.	valueStructure	structure of the data quality parameter	O	1	Class	DQM_ValueStructure << CodeList>> (C.3.3)

C.2.2.5   Data quality measure description

UML model shown in Figure 11.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
96.	DQM_Description	data quality measure description	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Lines 97-98
97.	textDescription	text description	M	1	CharacterString	Free text
98.	extendedDescription	illustration	O	1	Class	MD_BrowseGraphic (see ISO 19115-1:2013, B.2.3.2)

C.2.2.6   Data quality measure source reference

UML model shown in Figure 11.

	Name / Role Name	Definition	Obligation / Condition	Maximum occurrence	Data type	Domain
99.	DQM_SourceReference	reference to the source of the data quality measure	Use obligation from referencing object	Use maximum occurrence from referencing object	Class	Line 100
100.	citation	reference to the source	M	1	Class	CI_Citation <<DataType>> (see ISO 19115-1:2013 B.3.2.1)

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

	Name	Domain code	Definition
1.	DQ_EvaluationMethodType Code	EvalMethTypeCd	type of method for evaluating an identified data quality measure
2.	directInternal	001	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	002	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	003	method of evaluating the quality of a dataset based on external knowledge

C.3.3 Value structure

	Name	Domain code	Definition
1.	DQM_ValueStructure	ValueStructureCd	
2.	bag	001	finite, unordered collection of related items (objects or values) that may be repeated (ISO 19107:2003)
3.	set	002	unordered collection of related items (objects or values) with no repetition (ISO 19107:2003)
4.	sequence	003	finite, ordered collection of related items (objects or values) that may be repeated (ISO 19107:2003)
5.	table	004	an arrangement of data in which each item may be identified by means of arguments or keys (ISO/IEC 2382-4:1999)
6.	matrix	005	rectangular array of numbers (ISO/TS 19129:2009)
7.	coverage	006	feature that acts as a function to return values from its range for any direct position within its spatial, temporal or spatiotemporal domain (ISO 19123:2005)

Annex D
(normative)

List of standardized data quality measures

D.1 Introduction

This Annex provides and defines a list of standardized data quality measures.

This Annex defines data quality measures. In order to achieve well defined and comparable quality information, it is strongly recommended to carry out the evaluation and reporting of data quality using these data quality measures.

D.2 Completeness

D.2.1 Commission

The data quality measures for the data quality element commission are provided in Tables D.1 to D.4.

Table D.1 — Excess item

Line	Component	Description
1	Name	excess item
2	Alias	–
3	Element name	commission
4	Basic measure	error indicator
5	Definition	indication that an item is incorrectly present in the data
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates that the item is in excess)
9	Value structure	–
10	Source reference	–
11	Example	True (In a dataset, more items are classified as houses than in the universe of discourse)
12	Identifier	1

Table D.2 — Number of excess items

Line	Component	Description
1	Name	number of excess items
2	Alias	–
3	Element name	commission
4	Basic measure	error count
5	Definition	number of items within the dataset or sample that should not have been present
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	2 (12 houses are in the dataset although only 10 exist within the universe of discourse)
12	Identifier	2

Table D.3 — Rate of excess items

Line	Component	Description
1	Name	rate of excess items
2	Alias	–
3	Element name	commission
4	Basic measure	error rate
5	Definition	number of excess items in the dataset or sample in relation to the number of items that should have been present
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	10% (The dataset has 10% more houses than the universe of discourse)
12	Identifier	3

Table D.4 — Number of duplicate feature instances

Line	Component	Description
1	Name	number of duplicate feature instances
2	Alias	–
3	Element name	commission
4	Basic measure	error count
5	Definition	total number of exact duplications of feature instances within the data
6	Description	count of all items in the data that are incorrectly extracted with duplicate geometries
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	Features with identical attribution and identical coordinates: ▲ two (or more) points collected on top of each other; — two (or more) curves collected on top of each other; ▢ two (or more) surfaces collected on top of each other.
12	Identifier	4

D.2.2 Omission

The data quality measures for the data quality element omission are provided in Tables D.5 to D.7.

Table D.5 — Missing item

| Line	Component	Description
1	Name	missing item
2	Alias	–
3	Element name	omission
4	Basic measure	error indicator
5	Definition	indicator that shows that a specific item is missing in the data
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates that an item is missing)
9	Value structure	–
10	Source reference	–
11	Example	A data product specification requires all towers higher than 300 m to be captured. The data quality measure "missing item" allows a data quality evaluator or a data user to report that a specific item, in this case a feature of type "tower" (name depends on the application schema), is missing. Data quality scope: all towers with height > 300 Example result of a completeness evaluation of a particular data set: missing item = true for • tower.name = "Eiffel Tower, Paris, France" • tower.name = "Beijing Tower, Beijing, China"
12	Identifier	5

Table D.6 — Number of missing items

Line	Component	Description
1	Name	number of missing items
2	Alias	–
3	Element name	omission
4	Basic measure	error count
5	Definition	count of all items that should have been in the dataset or sample and are missing
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	2 (10 houses are in the dataset although 12 exist within the universe of discourse)
12	Identifier	6

Table D.7 — Rate of missing items

Line	Component	Description
1	Name	rate of missing items
2	Alias	–
3	Element name	omission
4	Basic measure	error rate
5	Definition	number of missing items in the dataset or sample in relation to the number of items that should have been present
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	10% (The dataset has 10% less houses than the universe of discourse)
12	Identifier	7

D.3 Logical consistency

D.3.1 Conceptual consistency

The data quality measures for the data quality element conceptual consistency are provided in Tables D.8 to D.13.

Table D.8 — Conceptual schema non-compliance

Line	Component	Description
1	Name	conceptual schema non-compliance
2	Alias	–
3	Element name	conceptual consistency
4	Basic measure	error indicator
5	Definition	indication that an item is not compliant to the rules of the relevant conceptual schema
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates that an item is not compliant with the rules of the conceptual schema)
9	Value structure	–
10	Source reference	–
11	Example	True (One feature relationship exists which is not defined in the conceptual schema)
12	Identifier	8

Table D.9 — Conceptual schema compliance

Line	Component	Description
1	Name	conceptual schema compliance
2	Alias	–
3	Element name	conceptual consistency
4	Basic measure	correctness indicator
5	Definition	indication that an item complies with the rules of the relevant conceptual schema
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates that an item is in compliance with the rules of the conceptual schema)
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	9

Table D.10 — Number of items not compliant with the rules of the conceptual schema

Line	Component	Description
1	Name	Number of items not compliant with the rules of the conceptual schema
2	Alias	–
3	Element name	conceptual consistency
4	Basic measure	error count
5	Definition	count of all items in the dataset that are not compliant with the rules of the conceptual schema
6	Description	If the conceptual schema explicitly or implicitly describes rules, these rules shall be followed. Violations against such rules can be, for example, invalid placement of features within a defined tolerance, duplication of features and invalid overlap of features.
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–

Table D.10 *(continued)*
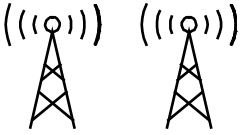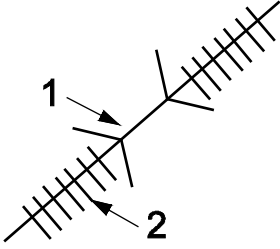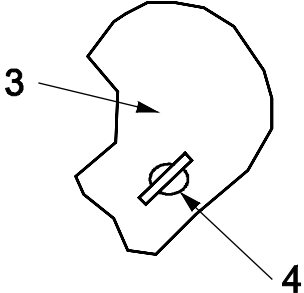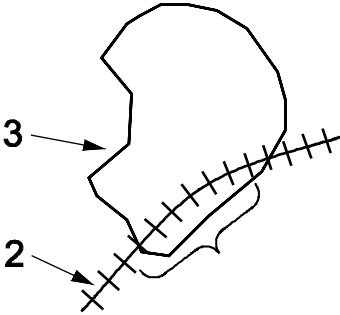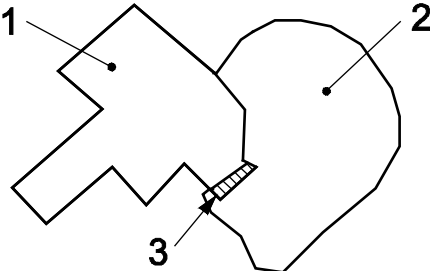
Line	Component	Description
11	Example	**Example 1**: Towers with identical attribution and within search tolerance (search tolerance = 10 m) **Example 2**: Bridge has invalid Transportation. Use Category of Road **Example 3**: Invalid placement of Airport inside a Lake **Example 4**: Invalid overlap of area feature Lake within line feature Railroad **Key** 1 Bridge 2 Railroad 3 Lake 4 Airport
12	Identifier	10

Table D.11 — Number of invalid overlaps of surfaces

Line	Component	Description
1	Name	number of invalid overlaps of surfaces
2	Alias	overlapping surfaces
3	Element name	conceptual consistency
4	Basic measure	error count
5	Definition	total number of erroneous overlaps within the data
6	Description	Which surfaces may overlap and which shall not is application dependent. Not all overlapping surfaces are necessarily erroneous. When reporting this data quality measure, the types of feature classes corresponding to the illegal overlapping surfaces shall be reported as well.
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	 **Key** 1 Surface 1 2 Surface 2 3 Overlapping Area
12	Identifier	11

Table D.12 — Non-compliance rate with respect to the rules of the conceptual schema

Line	Component	Description
1	Name	non-compliance rate with respect to the rules of the conceptual schema
2	Alias	–
3	Element name	conceptual consistency
4	Basic measure	error rate
5	Definition	number of items in the dataset that are not compliant with the rules of the conceptual schema in relation to the total number of these items supposed to be in the dataset
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	2%
12	Identifier	12

Table D.13 — Compliance rate with the rules of the conceptual schema

Line	Component	Description
1	Name	compliance rate with the rules of the conceptual schema
2	Alias	–
3	Element name	conceptual consistency
4	Basic measure	correct items rate
5	Definition	number of items in the dataset in compliance with the rules of the conceptual schema in relation to the total number of items
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	90%
12	Identifier	13

D.3.2 Domain consistency

The data quality measures for the data quality element domain consistency are provided in Tables D.14 to D.18.

Table D.14 — Value domain non-conformance

Line	Component	Description
1	Name	value domain non-conformance
2	Alias	–
3	Element name	domain consistency
4	Basic measure	error indicator
5	Definition	indication of if an item is not in conformance with its value domain
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates that an item is not in conformance with its value domain)
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	14

Table D.15 — Value domain conformance

Line	Component	Description
1	Name	value domain conformance
2	Alias	–
3	Element name	domain consistency
4	Basic measure	correctness indicator
5	Definition	indication that an item is conforming to its value domain
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates that an item is conforming to its value domain)
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	15

Table D.16 — Number of items not in conformance with their value domain

Line	Component	Description
1	Name	number of items not in conformance with their value domain
2	Alias	–
3	Element name	domain consistency
4	Basic measure	error count
5	Definition	count of all items in the dataset that are not in conformance with their value domain
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	16

Table D.17 — Value domain conformance rate

Line	Component	Description
1	Name	value domain conformance rate
2	Alias	–
3	Element name	domain consistency
4	Basic measure	correct items rate
5	Definition	number of items in the dataset that are in conformance with their value domain in relation to the total number of items in the dataset
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	17

Table D.18 — Value domain non-conformance rate

Line	Component	Description
1	Name	value domain non-conformance rate
2	Alias	–
3	Element name	domain consistency
4	Basic measure	error rate
5	Definition	number of items in the dataset that are not in conformance with their value domain in relation to the total number of items
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	18

D.3.3 Format consistency

The data quality measures for the data quality element format consistency are provided in Tables D.19 to D.21.

Table D.19 — Physical structure conflicts

Line	Component	Description
1	Name	physical structure conflicts
2	Alias	–
3	Element name	format consistency
4	Basic measure	error indicator
5	Definition	indication that items are stored in conflict with the physical structure of the dataset
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates physical structure conflict)
9	Value structure	–
10	Source reference	–
11	Example	True (dataset is stored in wrong fileformat, shapefile instead of gml)
12	Identifier	119

Table D.20 — Physical structure conflicts number

Line	Component	Description
1	Name	number of physical structure conflicts
2	Alias	–
3	Element name	format consistency
4	Basic measure	error count
5	Definition	count of all items in the dataset that are stored in conflict with the physical structure of the dataset
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	5 (5 living quarters type code is coded on more than 3 characters although the requirement in data product specification is 3)
12	Identifier	19

Table D.21 — Physical structure conflict rate

Line	Component	Description
1	Name	physical structure conflict rate
2	Alias	–
3	Element name	format consistency
4	Basic measure	error rate
5	Definition	number of items in the dataset that are stored in conflict with the physical structure of the dataset divided by the total number of items
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	20

D.3.4 Topological consistency

The data quality measures in Tables D.22 to D.28 are designed to test the topological consistency of geometric representations of features. They will not serve as measures of the consistency of explicit descriptions of topology using the topological objects specified in ISO 19107:2003.

Table D.22 — Number of faulty point-curve connections
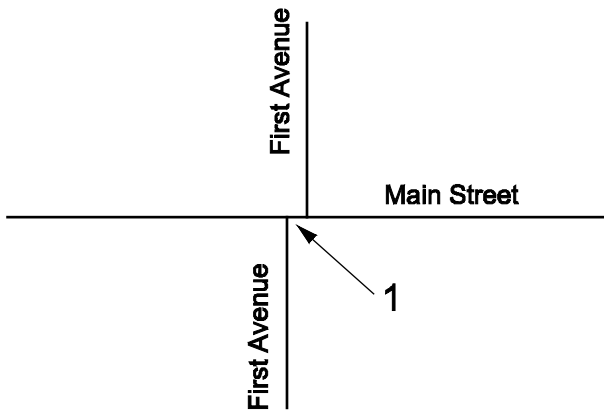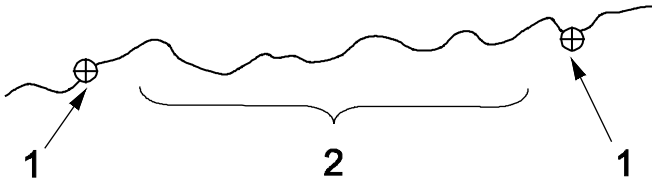
Line	Component	Description
1	Name	number of faulty point-curve connections
2	Alias	extraneous nodes
3	Element name	topological consistency
4	Basic measure	error count
5	Definition	number of faulty point-curve connections in the dataset
6	Description	A point-curve connection exists where different curves touch. These curves have an intrinsic topological relationship that shall reflect the true constellation. If the point-curve connection contradicts the universe of discourse, the point-curve connection is faulty with respect to this data quality measure. The data quality measure counts the number of errors of this kind.
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	**Example 1**: Two-point curve connections exist where only one should be present **Key** 1 Junction of two roads should be at a "+" intersection **Example 2**: System automatically places point-curve based on vertices limitation built into software code where no spatial justification for point-curve exists. **Key** 1 Link-node 2 500 vertices limit
12	Identifier	21

Table D.23 — Rate of faulty point-curve connections

Line	Component	Description
1	Name	rate of faulty point-curve connections
2	Alias	–
3	Element name	topological consistency
4	Basic measure	error rate
5	Definition	number of faulty link node connections in relation to the number of supposed link node connections
6	Description	A point-curve connection exists where different curves touch. These curves have an intrinsic topological relationship that shall reflect the true constellation. If the point-curve connection contradicts the universe of discourse, the point-curve connection is faulty with respect to this data quality measure. This data quality measure gives the erroneous point-curve connections in relation to the total number of point-curve connections.
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	22

Table D.24 — Number of missing connections due to undershoots
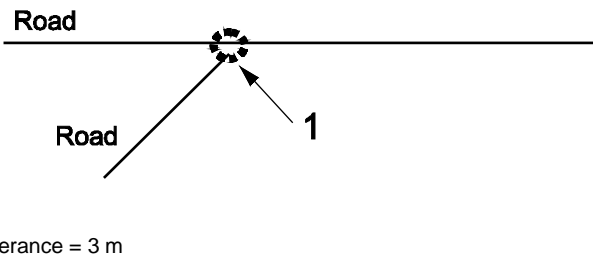
Line	Component	Description
1	Name	number of missing connections due to undershoots
2	Alias	undershoots
3	Element name	topological consistency
4	Basic measure	error count
5	Definition	count of items in the dataset, within the parameter tolerance, that are mismatched due to undershoots
6	Description	–
7	Parameter	search distance from the end of a dangling line
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	**Key** 1 Search tolerance = 3 m
12	Identifier	23

Table D.25 — Number of missing connections due to overshoots
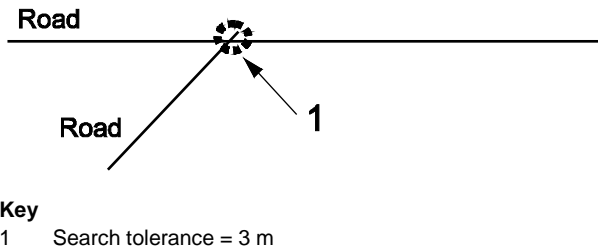
Line	Component	Description
1	Name	number of missing connections due to overshoots
2	Alias	overshoots
3	Element name	topological consistency
4	Basic measure	error count
5	Definition	count of items in the dataset, within the parameter tolerance, that are mismatched due to overshoots
6	Description	–
7	Parameter	search tolerance of minimum allowable length in the dataset
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	Road Road 1 **Key** 1 Search tolerance = 3 m
12	Identifier	24

Table D.26 — Number of invalid slivers

Line	Component	Description
1	Name	number of invalid slivers
2	Alias	slivers
3	Element name	topological consistency
4	Basic measure	error count
5	Definition	count of all items in the dataset that are invalid sliver surfaces
6	Description	A sliver is an unintended area that occurs when adjacent surfaces are not digitized properly. The borders of the adjacent surfaces may unintentionally gap or overlap by small amounts to cause a topological error.
7	Parameter	This data quality measure has 2 parameters: Parameter 1 Name: maximum sliver area size Definition: The maximum area determines the upper limit of a sliver. This is to prevent surfaces with sinuous perimeters and large areas from being mistaken as slivers. Value Type: Real Parameter 2 Name: thickness quotient Definition: The thickness quotient shall be a real number between 0 and 1. This quotient is determined by the following formula: T is the thickness quotient $T = 4\pi\,[\text{area}]/[\text{perimeter}]^2$ $T = 1$ value corresponds to a circle that has the largest area/perimeter^2 value. $T = 0$ value corresponds to a line that has the smallest area/perimeter^2 value. Description: The thickness quotient is independent of the size of the surface, and the closer the value is to 0, the thinner the selected sliver surfaces shall be. Value Type: Real
8	Value type	Integer
9	Value structure	–
10	Source reference	Source reference Environmental Systems Research Institute, Inc. (ESRI) GIS Data ReViewer 4.2 User Guide

Table D.26 *(continued)*
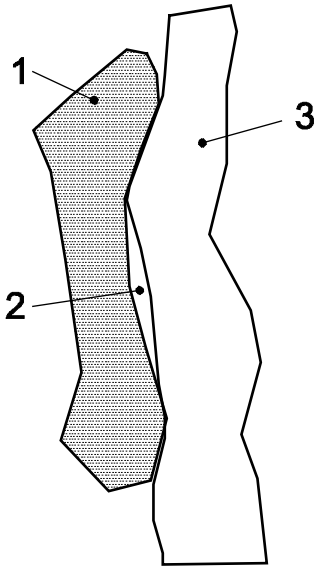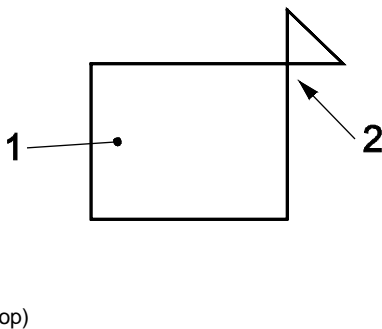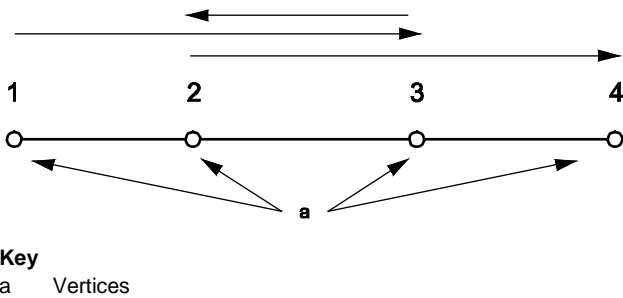
Line	Component	Description
11	Example	 **Key** 1 Single line drain 2 Double line drain a) Maximum area parameter prevents correct double line drain portrayal from being flagged as an error. **Key** 1 Sand 2 Sliver 3 Double line drain b) Sliver is less than the maximum parameter and is flagged for evaluation of possible error.
12	Identifier	25

Table D.27 — Number of invalid self-intersect errors

Line	Component	Description
1	Name	number of invalid self-intersect errors
2	Alias	loops
3	Element name	topological consistency
4	Basic measure	error count
5	Definition	count of all items in the data that illegally intersect with themselves
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	 **Key** 1 Building 1 2 Illegal intersection (loop)
12	Identifier	26

Table D.28 — Number of invalid self-overlap errors

Line	Component	Description
1	Name	number of invalid self-overlap errors
2	Alias	kickbacks
3	Element name	topological consistency
4	Basic measure	error count
5	Definition	count of all items in the data that illegally self overlap
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	 **Key** a Vertices
12	Identifier	27

D.4 Positional accuracy

D.4.1 Absolute or external accuracy

D.4.1.1 General measures for positional uncertainties

The data quality measures for positional uncertainty in general of the data quality element absolute or external accuracy are provided in Tables D.29 to D.34.

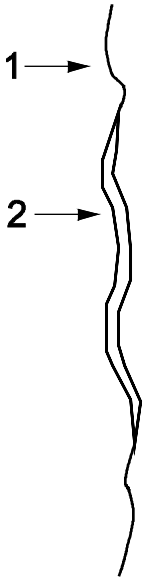**Table D.29 — Mean value of positional uncertainties**

Line	Component	Description
1	Name	mean value of positional uncertainties (1D, 2D and 3D)
2	Alias	–
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	mean value of the positional uncertainties for a set of positions where the positional uncertainties are defined as the distance between a measured position and what is considered as the corresponding true position
6	Description	For a number of points (N), the measured positions are given as xmi, ymi and zmi coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, xti, yti and zti, are considered to represent the true positions. The errors are calculated as 1D: $\quad e_i = \left\| x_{\mathrm{m}i} - x_{\mathrm{t}i} \right\|$ 2D: $\quad e_i = \sqrt{(x_{\mathrm{m}i} - x_{\mathrm{t}i})^2 + (y_{\mathrm{m}i} - y_{\mathrm{t}i})^2}$ 3D: $\quad e_i = \sqrt{x_{\mathrm{m}i} - x_{\mathrm{t}i})^2 + (y_{\mathrm{m}i} - y_{\mathrm{t}i})^2 + (z_{\mathrm{m}i} - z_{\mathrm{t}i})^2}$ The mean positional uncertainties of the horizontal absolute or external positions are then calculated as $$\bar{e} = \frac{1}{N} \sum_{i=1}^{N} e_i$$ A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result. This data quality measure is different from the standard deviation.
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	28

Table D.30 — Bias of positions

Line	Component	Description
1	Name	bias of positions (1D, 2D and 3D)
2	Alias	–
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	bias of the positions for a set of positions where the positional uncertainties are defined as the deviation between a measured position and what is considered as the corresponding true position
6	Description	For a number of points (N), the measured positions are given as x_{mi}, y_{mi} and z_{mi} coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, x_{ti}, y_{ti} and z_{ti}, are considered to represent the true positions. The deviation and biases are calculated as Single deviations: $$e_{xi} = x_{mi} - x_{ti}$$ $$e_{yi} = y_{mi} - y_{ti}$$ $$e_{zi} = z_{mi} - z_{ti}$$ Bias: $$a_x = \frac{\sum e_{xi}}{N_x}$$ $$a_y = \frac{\sum e_{yi}}{N_y}$$ $$a_z = \frac{\sum e_{zi}}{N_z}$$ $$a_p = \sqrt{a_x^2 + a_y^2}$$ $$a_{3D} = \sqrt{a_x^2 + a_y^2 + a_z^2}$$ A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result.
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	128

Table D.31 — Mean value of positional uncertainties excluding outliers

Line	Component	Description
1	Name	mean value of positional uncertainties excluding outliers (2D)
2	Alias	–
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	for a set of points where the distance does not exceed a defined threshold, the arithmetical average of distances between their measured positions and what is considered as the corresponding true positions
6	Description	For a number of points (N), the measured positions are given as x_{mi}, y_{mi} and z_{mi} coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, x_{ti}, y_{ti} and z_{ti}, are considered to represent the true positions. All positional uncertainties above a defined threshold e_{max} are then removed from the set. The positional uncertainties are calculated as $$e_i' = \begin{cases} e_i, & if & e_i \leq e_{max} \\ 0, & if & e_i > e_{max} \end{cases}$$ The calculation of e_i is given by the data quality measure "mean value of positional uncertainties" in one, two and three dimensions. For the remaining number of errors (N_R), the mean of the horizontal absolute positions is calculated as $$\overline{e}_{\text{excluding outliers}} = \frac{1}{N_R} \sum_{i=1}^{N} e_i'$$ A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines). The criteria for finding the corresponding points shall be reported with the data quality evaluation result.
7	Parameter	Name: e_{max} Definition: is the threshold for accepted positional uncertainties Value type: Number
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	29

Table D.32 — Number of positional uncertainties above a given threshold

Line	Component	Description
1	Name	number of positional uncertainties above a given threshold
2	Alias	–
3	Element name	absolute or external accuracy
4	Basic measure	error count
5	Definition	number of positional uncertainties above a given threshold for a set of positions The errors are defined as the distance between a measured position and what is considered as the corresponding true position.
6	Description	For a number of points (N), the measured positions are given as x_{mi}, y_{mi} and z_{mi} coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, x_{ti}, y_{ti} and z_{ti}, are considered to represent the true positions. The calculation of e_i is given by the data quality measure "mean value of positional uncertainties" in one, two and three dimensions. All positional uncertainties above a defined threshold e_{max} ($e_i > e_{max}$) are then counted as error. A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result.
7	Parameter	Name: e_{max} Definition: is the threshold for accepted positional uncertainties Value type: Number
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	30

Table D.33 — Rate of positional errors above a given threshold

Line	Component	Description
1	Name	rate of positional uncertainties above a given threshold
2	Alias	–
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	number of positional uncertainties above a given threshold for a set of positions in relation to the total number of measured positions The errors are defined as the distance between a measured position and what is considered as the corresponding true position
6	Description	For a number of points (N), the measured positions are given as x_{mi}, y_{mi} and z_{mi} coordinates depending on the dimension in which the position of the point is measured. A corresponding set of coordinates, x_{ti}, y_{ti} and z_{ti}, are considered to represent the true positions. The calculation of e_i is given by the data quality measure "mean value of positional uncertainties" in one, two and three dimensions. All positional uncertainties above a defined threshold $e_{\max}$ ($e_i > e_{\max}$) are then counted as error. The number of errors is set in relation to the total number of measured points. A criterion for the establishing of correspondence should also be stated (e.g. allowing for correspondence to the closest position, correspondence on vertices or along lines). The criterion/criteria for finding the corresponding points shall be reported with the data quality evaluation result.
7	Parameter	Name: $e_{\max}$ Definition: is the threshold above which the positional uncertainties are counted Value type: Number
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	25% of the nodes within the data quality scope have error distance greater than 1 metre
12	Identifier	31

Table D.34 — Covariance matrix

Line	Component	Description
1	Name	covariance matrix
2	Alias	variance-covariance matrix
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	symmetrical square matrix with variances of point coordinates on the main diagonal and covariance between these coordinates as off-diagonal elements
6	Description	The covariance matrix generalizes the concept of variance from one to n dimensions, i.e. from scalar-valued random variables to vector-valued random variables (tuples of scalar random variables).

(1)   1D coordinates (e.g. height data)

Vector-valued random variable: $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

Its covariance matrix: $\Sigma_{xx} = \begin{bmatrix} \sigma_{x_1}^2 & \cdots & \sigma_{x_1 x_n} \\ \vdots & \ddots & \vdots \\ \sigma_{x_n x_1} & \cdots & \sigma_{x_n}^2 \end{bmatrix}$, with $\sigma_{x_1 x_n} = \sigma_{x_n x_1}$

$\sigma_{x_1}^2$ denotes the variance of the element x_1, its square root gives the standard deviation of this element $\sigma_{x_1} = \sqrt{\sigma_{x_1}^2}$.

The correlation between 2 elements can be calculated by

$\rho_{x_i x_j} = \dfrac{\sigma_{x_i x_j}}{\sigma_{x_i} \sigma_{x_j}}$ .If the coordinates are uncorrelated, the off-diagonal elements are of value 0.

(2)   2D coordinates

Vector-valued random variable: $x = \begin{bmatrix} x_1 \\ y_1 \\ \vdots \\ x_n \\ y_n \end{bmatrix}$

Its covariance matrix: $\Sigma_{xx} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 y_1} & \cdots & \sigma_{x_1 y_n} \\ \sigma_{y_1 x_1} & \sigma_{y_1}^2 & \cdots & \sigma_{y_1 y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{y_n x_1} & \sigma_{y_n y_1} & \cdots & \sigma_{y_n}^2 \end{bmatrix}$,

Table D.34 *(continued)*

Line	Component	Description
		(3) 3D coordinates Vector-valued random variable: $x = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ \vdots \\ y_n \\ z_n \end{bmatrix}$ Its covariance matrix: $\Sigma_{xx} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 y_1} & \sigma_{x_1 z_1} & \cdots & \sigma_{x_1 y_n} & \sigma_{x_1 z_n} \\ \sigma_{x_1 y_1} & \sigma_{y_1}^2 & \sigma_{y_1 z_1} & \cdots & \sigma_{y_1 y_n} & \sigma_{y_1 z_n} \\ \sigma_{x_1 z_1} & \sigma_{y_1 z_1} & \sigma_{z_1}^2 & \cdots & \sigma_{z_1 y_n} & \sigma_{z_1 z_n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \sigma_{x_1 y_n} & \sigma_{y_1 y_n} & \sigma_{z_1 y_n} & \cdots & \sigma_{y_n}^2 & \sigma_{y_n z_n} \\ \sigma_{x_1 z_n} & \sigma_{y_1 z_n} & \sigma_{z_1 z_n} & \cdots & \sigma_{y_n z_n} & \sigma_{z_n}^2 \end{bmatrix}$, (4) arbitrary observables Vector-valued random variable: $x = \begin{bmatrix} a \\ b \\ \vdots \\ z \end{bmatrix}$ Its covariance matrix: $\Sigma_{xx} = \begin{bmatrix} \sigma_a^2 & \sigma_{b_a} & \cdots & \sigma_{z_a} \\ \sigma_{a_b} = \sigma_{b_a} & \sigma_b^2 & \cdots & \sigma_{z_b} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{a_z} = \sigma_{z_a} & \sigma_{b_z} = \sigma_{z_b} & \cdots & \sigma_z^2 \end{bmatrix}$
7	Parameter	–
8	Value type	Measure
9	Value structure	Matrix
10	Source reference	–
11	Example	–
12	Identifier	32

D.4.1.2   Vertical positional uncertainties

Height measurements are position observations in one dimension. The height may therefore be treated as a one-dimensional random variable. The data quality measures for positional uncertainties are therefore based on the data quality basic measure "one-dimensional random variable".

The data quality measures for vertical positional uncertainty of the data quality element absolute or external accuracy are provided in Tables D.35 to D.43.

Table D.35 — Linear error probable

Line	Component	Description
1	Name	linear error probable
2	Alias	LEP
3	Element name	absolute or external accuracy
4	Basic measure	LE50 or LE50(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value lies with probability 50 %
6	Description	See G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	33

Table D.36 — Standard linear error

Line	Component	Description
1	Name	standard linear error
2	Alias	SLE
3	Element name	absolute or external accuracy
4	Basic measure	LE68.3 or LE68.3(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value lies with probability 68,3 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	34

Table D.37 — Linear map accuracy at 90 % significance level

Line	Component	Description
1	Name	linear map accuracy at 90 % significance level
2	Alias	LMAS 90 %
3	Element name	absolute or external accuracy
4	Basic measure	LE90 or LE90(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value lies with probability 90 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	35

Table D.38 — Linear map accuracy at 95 % significance level

Line	Component	Description
1	Name	linear map accuracy at 95 % significance level
2	Alias	LMAS 95 %
3	Element name	absolute or external accuracy
4	Basic measure	LE95 or LE95(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value lies with probability 95 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	36

Table D.39 — Linear map accuracy at 99 % significance level

Line	Component	Description
1	Name	linear map accuracy at 99 % significance level
2	Alias	LMAS 99 %
3	Element name	absolute or external accuracy
4	Basic measure	LE99 or LE99(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value lies with probability 99 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	37

Table D.40 — Near certainty linear error

Line	Component	Description
1	Name	near certainty linear error
2	Alias	–
3	Element name	absolute or external accuracy
4	Basic measure	LE99.8 or LE99.8(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value lies with probability 99,8 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	38

Table D.41 — Root mean square error

Line	Component	Description
1	Name	root mean square error
2	Alias	RMSE
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	standard deviation, where the true value is not estimated from the observations but known *a priori*
6	Description	The true value of an observable Z is known as z_t. From this, the estimator $$\sigma_z = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_{mi} - z_t)^2}$$ yields to the linear root mean square error RMSE = α_z.
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	39

Table D.42 — Absolute linear error at 90 % significance level of biased vertical data (NATO)

Line	Component	Description
1	Name	absolute linear error at 90 % significance level of biased vertical data (Alternative 1)
2	Alias	LMAS
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	absolute vertical accuracy of the data's coordinates, expressed in terms of linear error at 90 % probability given that a bias is present
6	Description	A comparison of the data (source) and the control (reference) is calculated in the following manner: 1. Calculate the absolute error in the vertical dimension at each point: $\delta V_i = \text{source} V_i - \text{reference} V_i$ for $i = 1 \dots N$ 2. Calculate absolute value of the bias: $$\left| \overline{\delta V} \right| = \left| \frac{1}{N} \sum_{i=1}^{N} \delta V_i \right|$$ 3. Calculate the linear standard deviation of measured differences between the tested product and the reference source: $$\sigma_{\text{M}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \delta V_i^2}$$ 4. Calculate the linear standard deviation of errors in the reference source: σ_{R} 5. Calculate the linear standard deviation of errors in the tested product: $$\sigma_V = \sqrt{\sigma_{\text{M}}^2 + \sigma_{\text{R}}^2}$$ 6. Calculate the ratio of the absolute value of the mean error to the standard deviation: $$\text{ratio} = \frac{\left| \overline{\delta V} \right|}{\sigma_V}$$ 7. If $\text{ratio} > 1,4$, then $\text{LMAS} = \sigma_V \cdot \left[1,282 + \text{ratio} \right]$ 8. If ratio $\leq 1,4$ then $\text{LMAS} = \sigma_V \cdot \left[1,6435 + 0,92 \times \text{ratio}^2 - 0,28 \times \text{ratio}^3 \right]$
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	NATO STANAG 2215 IGEO (Reference [19])
11	Example	–
12	Measure identifier	40

Table D.43 — Absolute linear error at 90 % significance level of biased vertical data

Line	Component	Description
1	Name	Absolute linear error at 90 % significance level of biased vertical data (Alternative 2)
2	Alias	ALE
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	absolute vertical accuracy of the data's coordinates, expressed in terms of linear error at 90 % probability given that a bias is present
6	Description	A comparison of the data (source) and the control (reference) is calculated in the following manner: 1. Calculate the absolute error in the vertical dimension at each point: $\delta V_i = \text{source}V_i - \text{reference}V_i \ \text{ for } i = 1 \dots N$ 2. Calculate the mean vertical error: $$\left|\overline{\delta V}\right| = \left|\frac{1}{N}\sum_{i=1}^{N}\delta V_i\right|$$ 3. Calculate the standard deviation of the vertical errors: $$\sigma_V = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta V_i^{\,2}}$$ 4. Calculate the ratio of the absolute value of the mean error to the standard deviation: $$\text{ratio} = \left|\overline{\delta V}\right|/\sigma_V$$ 5. If $\text{ratio} > 1{,}4$, then $\quad k = 1{,}2815$ 6. If ratio $\leq$ 1,4, then calculate k based on the ratio of the vertical bias to the standard deviation of the heights using a cubic polynomial fit through the tabular values as defined in the *Handbook of Tables for Probability and Statistics* (Reference [17]). $$k = 1{,}643\,5 - \left(0{,}999\,556 \times \text{ratio}\right) + \left(0{,}923\,237 \times \text{ratio}^2\right) - \left(0{,}282\,533 \times \text{ratio}^3\right)$$ 7. Compute LE90 for the source: $$\text{LE90}_{\text{source}} = \left|\overline{\delta V}\right| + \left(k \times \sigma_V\right)$$ 8. Compute absolute LE90: $$\text{LE90}_{\text{abs}} = \sqrt{\text{LE90}_{\text{reference}}{}^2 + \text{LE90}_{\text{source}}{}^2}$$
7	Parameter	Name: Sample size Definition: minimum of 30 points is normally used but may not always be possible depending on identifiable control points. For feature level attribution sample 10 % of the feature population. Value Type: Real
8	Value type	Measure
9	Value structure	–

Table D.43 *(continued)*

Line	Component	Description
10	Source reference	1. *Mapping, Charting and Geodesy, Accuracy* (Reference [18])
		2. *Handbook of Tables for Probability and Statistics* (Reference [17])
		3. NATO STANAG 2215 IGEO (Reference [19])
11	Example	–
12	Measure identifier	41

D.4.1.3 Horizontal positional uncertainties

Horizontal point locations are defined by a 2D coordinates. The uncertainty of any point location can be described using the data quality basic measures for 2D random variables as described in G.3.3. The data quality measures for horizontal positional uncertainty of the data quality element absolute or external accuracy are provided in Tables D.44 to D.53.

Table D.44 — Circular standard deviation

Line	Component	Description
1	Name	circular standard deviation
2	Alias	circular standard error, Helmert's point error, CSE
3	Element name	absolute or external accuracy
4	Basic measure	CE39.4
5	Definition	radius describing a circle, in which the true point location lies with the probability of 39,4 %
6	Description	see G.3.3
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	42

Table D.45 — Circular error probable

Line	Component	Description
1	Name	circular error probable
2	Alias	CEP
3	Element name	absolute or external accuracy
4	Basic measure	CE50
5	Definition	radius describing a circle, in which the true point location lies with the probability of 50 %
6	Description	see G.3.3
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	43

Table D.46 — Circular map accuracy standard

Line	Component	Description
1	Name	circular error at 90% significant level
2	Alias	circular map accuracy standard (CMAS)
3	Element name	absolute or external accuracy
4	Basic measure	CE90
5	Definition	radius describing a circle, in which the true point location lies with the probability of 90 %
6	Description	see G.3.3
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	44

Table D.47 — Circular error at 95 % significance level

Line	Component	Description
1	Name	circular error at 95 % significance level
2	Alias	navigation accuracy
3	Element name	absolute or external accuracy
4	Basic measure	CE95
5	Definition	radius describing a circle, in which the true point location lies with the probability of 95 %
6	Description	see G.3.3
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	45

Table D.48 — Circular near certainty error

Line	Component	Description
1	Name	circular near certainty error
2	Alias	CNCE
3	Element name	absolute or external accuracy
4	Basic measure	CE99.8
5	Definition	radius describing a circle, in which the true point location lies with the probability of 99,8 %
6	Description	see G.3.3
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	46

Table D.49 — Root mean square error of planimetry

Line	Component	Description
1	Name	root mean square error of planimetry
2	Alias	RMSEP
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	radius of a circle around the given point, in which the true value lies with probability P
6	Description	The true values of the observed coordinates X and Y are known as x_t and y_t. From this the estimator $$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[(x_{mi} - x_t)^2 + (y_{mi} - y_t)^2\right]}$$ yields to the linear root mean square error of planimetry RMSEP = σ
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	47

Table D.50 — Absolute circular error at 90 % significance level of biased data

Line	Component	Description	
1	Name	absolute circular error at 90 % significance level of biased data (Alternative 2)	
2	Alias	absolute horizontal accuracy measure at the 90% significance level of biased data	CMAS
3	Element name	absolute or external accuracy	
4	Basic measure	not applicable	
5	Definition	absolute horizontal accuracy of the data's coordinates, expressed in terms of circular error at 90 % probability given that a bias is present	
6	Description	A comparison of the data (source) and the control (reference) is calculated in the following manner: 1. Calculate the absolute error in the horizontal dimension at each point and each coordinate Xi and Y_i: $$\delta X_i = \left(\text{source}X_i - \text{reference}X_i\right) \text{ and } \delta Yi = \left(\text{source}Y_i - \text{reference}Y_i\right) \text{ for } i = 1\ldots N$$ 2. Calculate the mean horizontal error of each coordinate: $$\overline{\delta X} = \frac{1}{N}\sum_1^N \delta Xi \text{ and } \overline{\delta Y} = \frac{1}{N}\sum_1^N \delta Yi$$ 3. Calculate the circular standard deviation of measured differences between the tested product and the reference source: $$\sigma_{CM} = \sqrt{\frac{1}{2(N-1)}\left(\sum_{i=1}^N\left(\delta Xi - \overline{\delta X}\right)^2 + \sum_{i=1}^N\left(\delta Xi - \overline{\delta X}\right)^2\right)}$$ 4. Calculate the circular standard deviation of errors in the reference source: σ_{CR} 5. Calculate the circular standard deviation of errors in the tested product: $$\sigma_C = \sqrt{\sigma_{CM}^2 + \sigma_{CR}^2}$$ 6. Compute absolute circular error at 90 % confidence level of biased data (CMAS): $$\text{CMAS} = \sigma_C \cdot \left[1{,}294\,3 + \sqrt{\left(\frac{\overline{\delta X}^2 + \overline{\delta Y}^2}{\sigma_C}\right)} + 0{,}725\,4\right]$$	
7	Parameter	–	
8	Value type	Measure	
9	Value structure	–	
10	Source reference	NATO STANAG 2215 IGEO (Reference [19])	
11	Example	–	
12	Measure identifier	48	

Table D.51 — Absolute circular error at 90 % significance level of biased data (NATO)

Line	Component	Description
1	Name	absolute circular error at 90 % significance level of biased data
2	Alias	ACE
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	absolute horizontal accuracy of the data's coordinates, expressed in terms of circular error at 90% probability given that a bias is present
6	Description	A comparison of the data (source) and the control (reference) is calculated in the following manner: 1. Calculate the absolute error in the horizontal dimension at each point: $$\Delta H_i = \sqrt{\left(\text{source}X_i - \text{reference}X_i\right)^2 + \left(\text{source}Y_i - \text{reference}Y_i\right)^2} \quad \text{for } i = 1 \ldots N$$ 2. Calculate the mean horizontal error: $$\mu_{\mathsf{H}} = \left(\sum \Delta H_i\right) \Big/ N$$ 3. Calculate the standard deviation of the horizontal errors: $$\sigma_{\mathsf{H}} = \sqrt{\sum \left(\Delta H_i - \mu_{\mathsf{H}}\right)^2 \Big/ (N-1)}$$ 4. Calculate the ratio of the absolute value of the mean error to the standard deviation: $$\text{ratio} = \left|\mu_H\right| / \sigma_H$$ 5. If $\text{ratio} > 1{,}4$, then $k = 1{,}2815$ 6. If ratio $\leq 1{,}4$, then calculate k, the ratio of the mean to the standard deviation, using a cubic polynomial fit through the tabular values as defined in the *CRC Handbook of Tables for Probability and Statistics* [17] $$k = 1{,}6435 - \left(0{,}999\,556 \times \text{ratio}\right) + \left(0{,}923\,237 \times \text{ratio}^2\right) - \left(0{,}282\,533 \times \text{ratio}^3\right)$$ 7. Compute CE90 for the source: $$\text{CE90}_{\text{source}} = \left|\mu_{\mathsf{H}}\right| + \left(k \times \sigma_{\mathsf{H}}\right)$$ 8. Compute absolute CE90: $$\text{CE90}_{\text{abs}} = \sqrt{\text{CE90}_{\text{reference}}{}^2 + \text{CE90}_{\text{source}}{}^2}$$
7	Parameter	Name: Sample size Definition: minimum of 30 points is normally used but may not always be possible depending on identifiable control points. For feature level attribution sample 10 % of the feature population. Value Type: Real
8	Value type	Measure
9	Value structure	–

Table D.51 *(continued)*

Line	Component	Description
10	Source reference	1. *Mapping, Charting and Geodesy Accuracy* (Reference [18])
		2. *Handbook of Tables for Probability and Statistics* (Reference [17])
11	Example	–
12	Measure identifier	49

Table D.52 — Uncertainty ellipse

Line	Component	Description
1	Name	uncertainty ellipse
2	Alias	standard point error ellipse
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	2D ellipse with the two main axes indicating the direction and magnitude of the highest and the lowest uncertainty of a 2D point
6	Description	From a given covariance matrix (data quality measure Table D.34) of 2D point coordinates, the elements describing the uncertainty ellipse can be determined by its eigenvalues. For a single point k, the covariance matrix is given by $$\Sigma^k{}_{xx} = \begin{bmatrix} \sigma^2_{x_k} & \sigma_{x_k y_k} \\ \sigma_{y_k x_k} & \sigma^2_{y_k} \end{bmatrix}, \text{ with } \sigma_{x_k y_k} = \sigma_{y_k x_k}$$ The direction α (bearing) of the major semi-axis of the uncertainty ellipse can be computed by $$\varphi = \frac{1}{2}\arctan\frac{2\sigma_{x_k y_k}}{\sigma^2_{x_k} - \sigma^2_{y_k}}$$ and $$a = \sqrt{\frac{1}{2}\left(\sigma^2_{x_k} + \sigma^2_{y_k} + \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$$ $$b = \sqrt{\frac{1}{2}\left(\sigma^2_{x_k} + \sigma^2_{y_k} - \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$$
7	Parameter	–
8	Value type	Measure
9	Value structure	Sequence (a, b, φ)
10	Source reference	–
11	Example	–
12	Identifier	50

Table D.53 — Confidence ellipse

Line	Component	Description
1	Name	confidence ellipse
2	Alias	confidence point error ellipse
3	Element name	absolute or external accuracy
4	Basic measure	not applicable
5	Definition	2D ellipse with the two main axes indicating the direction and magnitude of the highest and the lowest uncertainty of a 2D point
6	Description	From a given covariance matrix (data quality measure Table D.34), the elements describing the uncertainty ellipse can be determined by its eigenvalues. For a single point k, the covariance matrix is given by $$\Sigma^k{}_{xx} = \begin{bmatrix} \sigma^2_{x_k} & \sigma_{x_k y_k} \\ \sigma_{y_k x_k} & \sigma^2_{y_k} \end{bmatrix}, \text{ with } \sigma_{x_k y_k} = \sigma_{y_k x_k}$$ The direction α (bearing) of the major semi-axis of the uncertainty ellipse can be computed by $$\varphi = \frac{1}{2}\arctan\frac{2\sigma_{x_k y_k}}{\sigma^2_{x_k} - \sigma^2_{y_k}}$$ and $$a = \sqrt{\frac{1}{2}\chi^2_{1-\alpha}(2)\left(\sigma^2_{x_k} + \sigma^2_{y_k} + \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$$ $$b = \sqrt{\frac{1}{2}\chi^2_{1-\alpha}(2)\left(\sigma^2_{x_k} - \sigma^2_{y_k} - \sqrt{\left(\sigma^2_{x_k} - \sigma^2_{y_k}\right)^2 + 4\sigma^2_{x_k y_k}}\right)}$$ With values for the $\chi^2_{1-\alpha}(2)$-distribution of a 2D-confidence ellipse $$\chi^2_{1-\alpha}(2)$$ $P = 1 - \alpha = 95\ \%$ 5,99 $P = 1 - \alpha = 99\ \%$ 9,21
7	Parameter	Name: significance level Definition: $1 - \alpha$ Value Type: Number
8	Value type	Measure
9	Value structure	Sequence (a, b, φ)
10	Source reference	–
11	Example	–
12	Identifier	51

D.4.1.4 Relative or internal accuracy

This data quality element uses the same set of data quality measures as absolute or external accuracy. The difference is only in the method of evaluation.

The relative accuracy between features can be expressed using the data quality measures Relative vertical error and Relative horizontal error. They are defined in Tables D.54 and D.55.

Table D.54 — Relative vertical error

Line	Component	Description
1	Name	relative vertical error
2	Alias	Rel LE90
3	Element name	relative or internal accuracy
4	Basic measure	not applicable
5	Definition	evaluation of the random errors of one relief feature to another in the same dataset or on the same map/chart It is a function of the random errors in the two elevations with respect to a common vertical datum.
6	Description	A comparison of the data (measured) and the control (true) is calculated in the following manner: 1. Determine all possible point pair combinations: Point Pair Combinations = $m = n(n-1)/2$ 2. Calculate the absolute vertical error at each point: ΔZ_i = Measured Height$_i$ – True Height$_i$ for $i = 1…n$ 3. Calculate the relative vertical error for all point pair combinations: $\Delta Z_{\text{rel } kj} = \Delta Z_k - \Delta Z_j$ for $k = 1…m - 1$, $j = k + 1, … m$ 4. Calculate the relative vertical standard deviation: $$\sigma_{Z\,\text{rel}} = \sqrt{\dfrac{\Sigma \Delta Z_{\text{rel}}{}^2}{m-1}}$$ 5. Calculate the Relative LE by converting the sigma to a 90 % statistic: Rel LE90 = 1,645 $\sigma_{Z\,\text{rel}}$
7	Parameter	Name: n Definition: Sample size Value Type: Integer
8	Value type	Measure
9	Value structure	–
10	Source reference	*Mapping, Charting and Geodesy Accuracy* (Reference [18])
11	Example	–
12	Measure identifier	52

91

Table D.55 — Relative horizontal error

Line	Component	Description
1	Name	relative horizontal error
2	Alias	Rel CE90
3	Element name	relative or internal accuracy
4	Basic measure	not applicable
5	Definition	evaluation of the random errors in the horizontal position of one feature to another in the same dataset or on the same map/chart
6	Description	A comparison of the data (measured) and the control (true) is calculated in the following manner: 1. Determine all possible point pair combinations: Point Pair Combinations = $m = n(n-1)/2$ 2. Calculate the absolute error in the X and Y dimensions at each point: ΔX_i = Measured X_i – True X_i for $i = 1…n$ ΔY_i = Measured Y_i – True Y_i for $i = 1…n$ 3. Calculate the relative error in X and Y for all point pair combinations: $\Delta X_{\text{rel } kj} = \Delta X_k - \Delta X_j$ for $k = 1…m-1, j = k+1, … m$ $\Delta Y_{\text{rel } kj} = \Delta Y_k - \Delta Y_j$ for $k = 1…m-1, j = k+1, … m$ 4. Calculate the relative standard deviations in each axis: $\sigma_{X \text{ rel}} = \sqrt{\dfrac{\Sigma \Delta X_{\text{rel}}^2}{m-1}}$ $\sigma_{Y \text{ rel}} = \sqrt{\dfrac{\Sigma \Delta Y_{\text{rel}}^2}{m-1}}$ 5. Calculate the relative horizontal standard deviation: $\sigma_{H \text{ rel}} = \sqrt{\dfrac{\sigma_{X \text{ rel}}^2 + \sigma_{Y \text{ rel}}^2}{2}}$ 6. Calculate the Relative CE by converting the sigma to a 90 % significance level: Rel CE90 = 2,146 $\sigma_{H \text{ rel}}$
7	Parameter	Name: n Definition: Sample size Value Type: Integer
8	Value type	Measure
9	Value structure	–
10	Source reference	*Mapping, Charting and Geodesy Accuracy* (Reference [18])
11	Example	–
12	Measure identifier	53

D.4.2 Gridded data position accuracy

The accuracy of gridded data may be described using the same data quality measures as for the horizontal positional uncertainty, as specified in D.4.1.3. The band values in rasters may be described using the quantitative attribute accuracy as specified in D.6.3.

D.5 Temporal quality

D.5.1 Accuracy of a time measurement

Time measurements can be treated as 1-dimensional random variables. Using the data quality basic measures as described in G.3.2 leads to the data quality measures as provided in Tables D.56 to D.61.

Table D.56 — Time accuracy at 68,3 % significance level

Line	Component	Description
1	Name	time accuracy at 68,3 % significance level
2	Alias	–
3	Element name	accuracy of a time measurement
4	Basic measure	LE68.3 or LE68.3(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the time instance lies with probability 68,3 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	54

Table D.57 — Time accuracy at 50 % significance level

Line	Component	Description
1	Name	time accuracy at 50 % significance level
2	Alias	–
3	Element name	accuracy of a time measurement
4	Basic measure	LE50 or LE50(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the time instance lies with probability 50 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	55

Table D.58 — Time accuracy at 90 % significance level

Line	Component	Description
1	Name	time accuracy at 90 % significance level
2	Alias	–
3	Element name	accuracy of a time measurement
4	Basic measure	LE90 or LE90(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the time instance lies with probability 90 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	56

Table D.59 — Time accuracy at 95 % significance level

Line	Component	Description
1	Name	time accuracy at 95 % significance level
2	Alias	–
3	Element name	accuracy of a time measurement
4	Basic measure	LE95 or LE95(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the time instance lies with probability 95 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	57

Table D.60 — Time accuracy at 99 % significance level

Line	Component	Description
1	Name	time accuracy at 99 % significance level
2	Alias	–
3	Element name	accuracy of a time measurement
4	Basic measure	LE99 or LE99(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the time instance lies with probability 99 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	58

Table D.61 — Time accuracy at 99,8 % significance level

Line	Component	Description
1	Name	time accuracy at 99,8 % significance level
2	Alias	–
3	Element name	accuracy of a time measurement
4	Basic measure	LE99.8 or LE99.8(r), depending on the evaluation procedur
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the time instance lies with probability 99,8 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	59

D.5.2 Temporal consistency

One data quality measure for the data quality element temporal consistency is provided in Table D.62.

Table D.62 — Chronological error

Line	Component	Description
1	Name	chronological error
2	Alias	–
3	Element name	temporal consistency
4	Basic measure	error indicator
5	Definition	indication that an event is incorrectly ordered against the other events
6	Description	–
7	Parameter	–
8	Value type	Boolean (true indicates that the event is incorrectly ordered)
9	Value structure	–
10	Source reference	–
11	Example	True (5 historical events are present in the dataset but are not ordered correctly).
12	Identifier	159

D.5.3 Temporal validity

The temporal validity may be treated with the same data quality measures as for other domain specific attribute values (see data quality measures in Tables D.14 to D.18 of the data quality element domain consistency).

D.6 Thematic accuracy

D.6.1 Classification correctness

The assignment of an item to a certain class can either be correct or incorrect. Depending on the item that is classified, several data quality measures are given in Tables D.63 to D.67.

Table D.63 — Number of incorrectly classified features

Line	Component	Description
1	Name	number of incorrectly classified features
2	Alias	–
3	Element name	classification correctness
4	Basic measure	error count
5	Definition	number of incorrectly classified features
6	Description	–
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	60

Table D.64 — Misclassification rate

Line	Component	Description
1	Name	misclassification rate
2	Alias	–
3	Element name	classification correctness
4	Basic measure	error rate
5	Definition	number of incorrectly classified features relative to the number of features that should be there
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	61

Table D.65 — Misclassification matrix

Line	Component	Description
1	Name	misclassification matrix
2	Alias	confusion matrix
3	Element name	classification correctness
4	Basic measure	–
5	Definition	matrix that indicates the number of items of class (i) classified as class (j)
6	Description	The misclassification matrix (MCM) is a quadratic matrix with n columns and n rows. n denotes the number of classes under consideration. MCM (i,j) = [# items of class (i) classified as class (j)] The diagonal elements of the misclassification matrix contain the correctly classified items, and the off diagonal elements contain the number of misclassification errors.
7	Parameter	Name: n Definition: number of classes under consideration Value Type: Integer
8	Value type	Integer
9	Value structure	Matrix ($n \times n$)
10	Source reference	–
11	Example	
12	Identifier	62

Example (Line 11):

		Dataset class			
		A	B	C	Count
True class	A	7	2	1	10
	B	1	2	2	5
	C	1	1	3	5
	Count	9	5	6	20

Table D.66 — Relative misclassification matrix

Line	Component	Description
1	Name	relative misclassification matrix
2	Alias	–
3	Element name	classification correctness
4	Basic measure	–
5	Definition	matrix that indicates the number of items of class (i) classified as class (j) divided by the number of items of class (i)
6	Description	The relative misclassification matrix (RMCM) is a quadratic matrix with n columns and n rows. n denotes the number of classes under consideration. RMCM (i,j) = [# items of class (i) classified as class (j)] / (# items of class (i)] $\times 100$ %
7	Parameter	Name: n Definition: number of classes under consideration Value Type: Integer
8	Value type	Real
9	Value structure	Matrix ($n \times n$)
10	Source reference	–
11	Example	–
12	Identifier	63

Table D.67 — Kappa coefficient

Line	Component	Description
1	Name	kappa coefficient
2	Alias	–
3	Element name	classification correctness
4	Basic measure	–
5	Definition	coefficient to quantify the proportion of agreement of assignments to classes by removing misclassifications
6	Description	With the elements of the misclassification matrix MCM(*i,j*) given as data quality measure in Table D.65 the kappa coefficient (κ) can be calculated by $$\kappa = \frac{N \cdot \sum_{i=1}^{r} \text{MCM}(i,i) - \sum_{i=1}^{r}\left(\sum_{j=1}^{r}\text{MCM}(i,j) \cdot \sum_{j=1}^{r}\text{MCM}(j,i)\right)}{N^2 - \sum_{i=1}^{r}\left(\sum_{j=1}^{r}\text{MCM}(i,j) \cdot \sum_{j=1}^{r}\text{MCM}(j,i)\right)}$$ N is the number of classified items
7	Parameter	Name: r Definition: number of classes under consideration Value Type: Integer
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	64

D.6.2 Non-quantitative attribute correctness

The data quality measures for the data quality element non-quantitative attribute correctness are provided in Tables D.68 to D.70.

Table D.68 — Number of incorrect attribute values

Line	Component	Description
1	Name	number of incorrect attribute values
2	Alias	–
3	Element name	non-quantitative attribute correctness
4	Basic measure	error count
5	Definition	total number of erroneous attribute values within the relevant part of the dataset
6	Description	count of all attribute values where the value is incorrect
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	5 (5 geographical names are misspelled)
12	Identifier	65

Table D.69 — Rate of correct attribute values

Line	Component	Description
1	Name	rate of correct attribute values
2	Alias	–
3	Element name	non-quantitative attribute correctness
4	Basic measure	correct items rate
5	Definition	number of correct attribute values in relation to the total number of attribute values
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	66

Table D.70 — Rate of incorrect attribute values

Line	Component	Description
1	Name	rate of incorrect attribute values
2	Alias	–
3	Element name	non-quantitative attribute correctness
4	Basic measure	error rate
5	Definition	number of attribute values where incorrect values are assigned in relation to the total number of attribute values
6	Description	–
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	67

D.6.3 Quantitative attribute accuracy

The data quality measures for the data quality element quantitative attribute accuracy are provided in Tables D.71 to D.76.

Table D.71 — Attribute value uncertainty at 68,3 % significance level

Line	Component	Description
1	Name	attribute value uncertainty at 68,3 % significance level
2	Alias	–
3	Element name	quantitative attribute accuracy
4	Basic measure	LE68.3 or LE68.3(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 68,3 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	68

Table D.72 — Attribute value uncertainty at 50 % significance level

Line	Component	Description
1	Name	attribute value uncertainty at 50 % significance level
2	Alias	–
3	Element name	quantitative attribute accuracy
4	Basic measure	LE50 or LE50(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 50 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	69

Table D.73 — Attribute value uncertainty at 90 % significance level

Line	Component	Description
1	Name	attribute value uncertainty at 90 % significance level
2	Alias	–
3	Element name	quantitative attribute accuracy
4	Basic measure	LE90 or LE90(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 90 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	70

Table D.74 — Attribute value uncertainty at 95 % significance level

Line	Component	Description
1	Name	attribute value uncertainty at 95 % significance level
2	Alias	–
3	Element name	quantitative attribute accuracy
4	Basic measure	LE95 or LE95(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 95 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	71

Table D.75 — Attribute value uncertainty at 99 % significance level

Line	Component	Description
1	Name	attribute value uncertainty at 99 % significance level
2	Alias	–
3	Element name	quantitative attribute accuracy
4	Basic measure	LE99 or LE99(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 99 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	72

Table D.76 — Attribute value uncertainty at 99,8 % significance level

Line	Component	Description
1	Name	attribute value uncertainty at 99,8 % significance level
2	Alias	–
3	Element name	quantitative attribute accuracy
4	Basic measure	LE99.8 or LE99.8(r), depending on the evaluation procedure
5	Definition	half length of the interval defined by an upper and a lower limit, in which the true value for the quantitative attribute lies with probability 99,8 %
6	Description	see G.3.2
7	Parameter	–
8	Value type	Measure
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	73

D.7 Aggregation Measures

In a data product specification, several requirements are set up for a product to conform to the specification. The data quality measures for this element are provided in Tables D.77 to D.81.

Table D.77 — Data product specification passed

Line	Component	Description
1	Name	data product specification passed
2	Alias	–
3	Element name	usability element
4	Basic measure	correctness indicator
5	Definition	indication that all requirements in the referred data product specification are fulfilled
6	Description	
7	Parameter	–
8	Value type	Boolean (true if all the requirements in the referred data product specification are fulfilled)
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	101

　　　　　　　　　　　　　　　　　　　　　　　　　　　　105

Table D.78 — Data product specification fail count

Line	Component	Description
1	Name	data product specification fail count
2	Alias	–
3	Element name	usability element
4	Basic measure	error count
5	Definition	number of data product specification requirements that are not fulfilled by the current product/dataset
6	Description	
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	102

Table D.79 — Data product specification pass count

Line	Component	Description
1	Name	data product specification pass count
2	Alias	–
3	Element name	usability element
4	Basic measure	correct items count
5	Definition	number of the data product specification requirements that are fulfilled by the current product/dataset
6	Description	
7	Parameter	–
8	Value type	Integer
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	103

Table D.80 — Data product specification fail rate

Line	Component	Description
1	Name	data product specification fail rate
2	Alias	–
3	Element name	usability element
4	Basic measure	error rate
5	Definition	number of the data product specification requirements that are not fulfilled by the current product/dataset in relation to the total number of data product specification requirements
6	Description	
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	104

Table D.81 — Data product specification pass rate

Line	Component	Description
1	Name	data product specification pass rate
2	Alias	–
3	Element name	usability element
4	Basic measure	correct items rate
5	Definition	number of the data product specification requirements that are fulfilled by the current product/dataset in relation to the total number of data product specification requirements
6	Description	
7	Parameter	–
8	Value type	Real
9	Value structure	–
10	Source reference	–
11	Example	–
12	Identifier	105

Annex E
(informative)

Evaluating and reporting data quality

E.1 Introduction

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

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

E.2 Dataset description

E.2.1 Data product specification

E.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 in the dataset.

NOTE   This is not a complete example of a data product specification (see ISO 19131:2007).

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

E.2.1.2   Feature Types

Each feature type, with zero or more attributes, is listed in Table E.1. Each attribute name is followed by a value type (string or integer) and by an optional value domain.

Table E.1 — Feature types

	Feature type	Attribute name	Value type	Value domain
Buildings	Industrial building			
	House	Family name	String	
		Number of occupants	Integer	
Transport Network	Path			
	Road	Condition	String	surfaced, unsurfaced
	Tree	Height	String	A : from 1 to 3 metre, B : from 3 to 5 metre, C : from 5 to 10 metre, D : more than 10 metre

E.2.1.3 Rules

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

— trees with a height of less than 1 metre 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").

E.2.1.4 Quality requirements

Overall data quality requirement: to be conformant with the data quality requirements, a dataset shall pass all the data quality requirements below.

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

Transport Network:

2) Max two items can be missing for each feature type

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

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

Buildings:

5) Max two items can be missing for each feature type

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

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

Trees:

8) Max 10% missing trees

9) Max 10% trees in excess

10) Max 20% of the trees can have wrong height

11) No feature instances can be misclassified as other feature types

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

The relationship between the three figures is as follows:

— Figure E.1 represents the "real world", which generally contains more features than will be contained in the dataset;

— Figure E.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;

— Figure E.3 represents the dataset as produced.

In all 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.
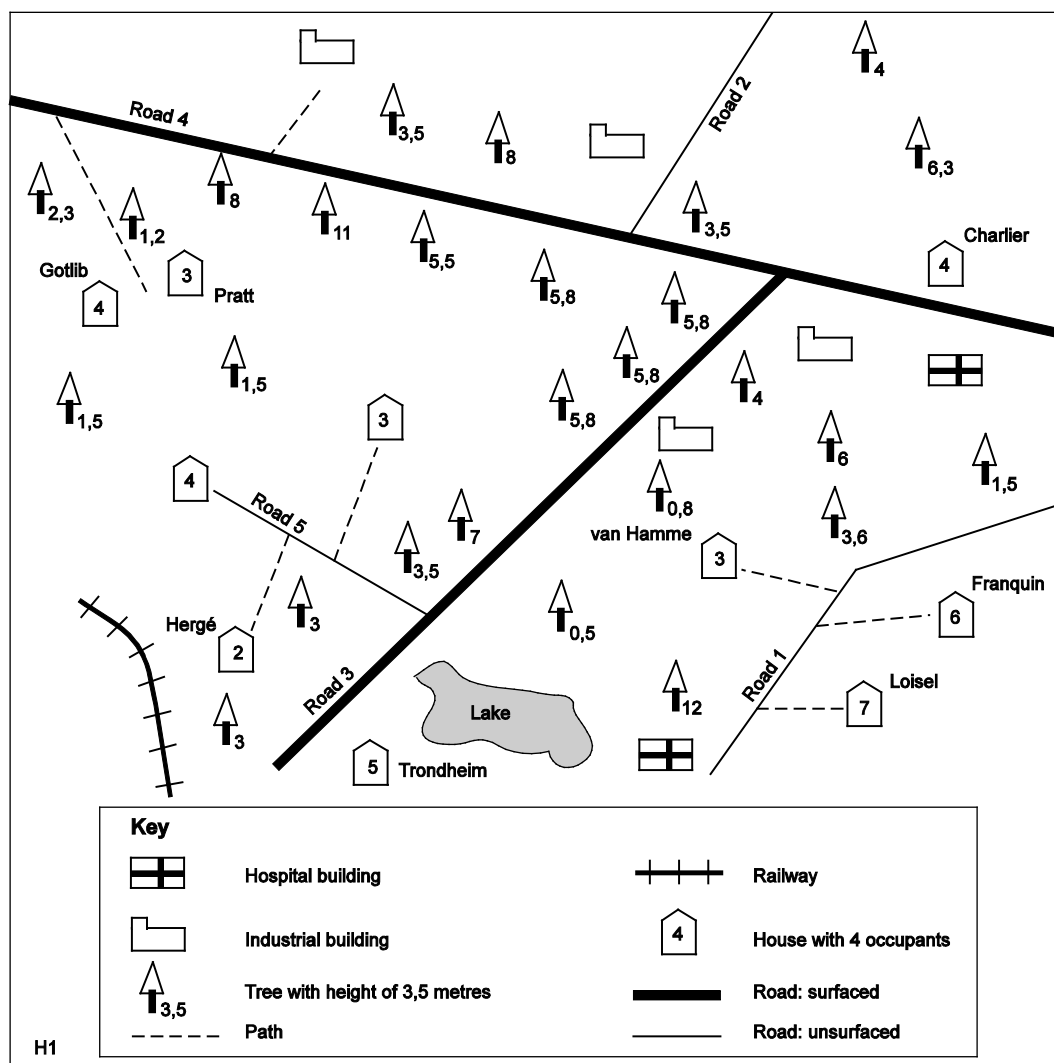
Figure E.1 — Graphical representation of the "real world"
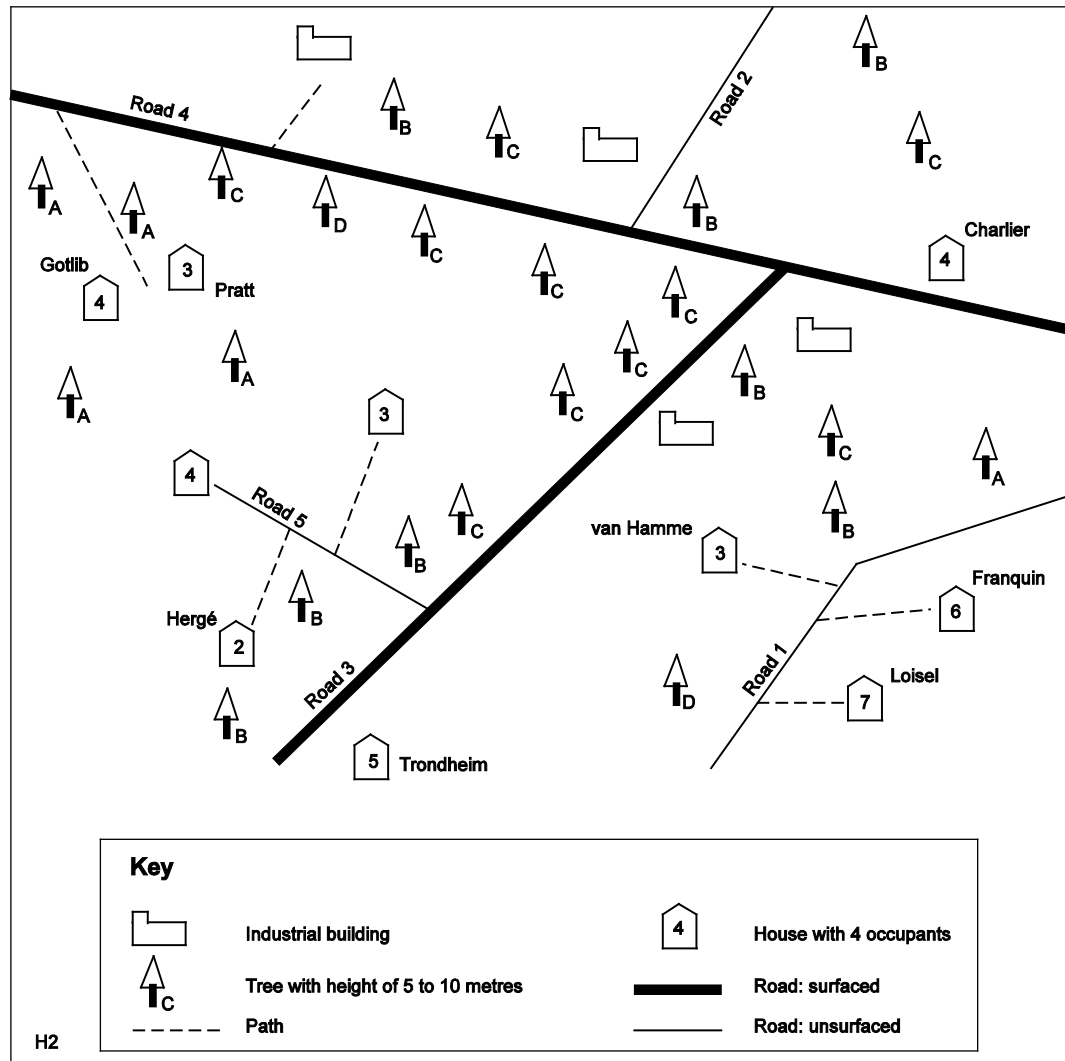
Figure E.2 — Graphical representation of the universe of discourse
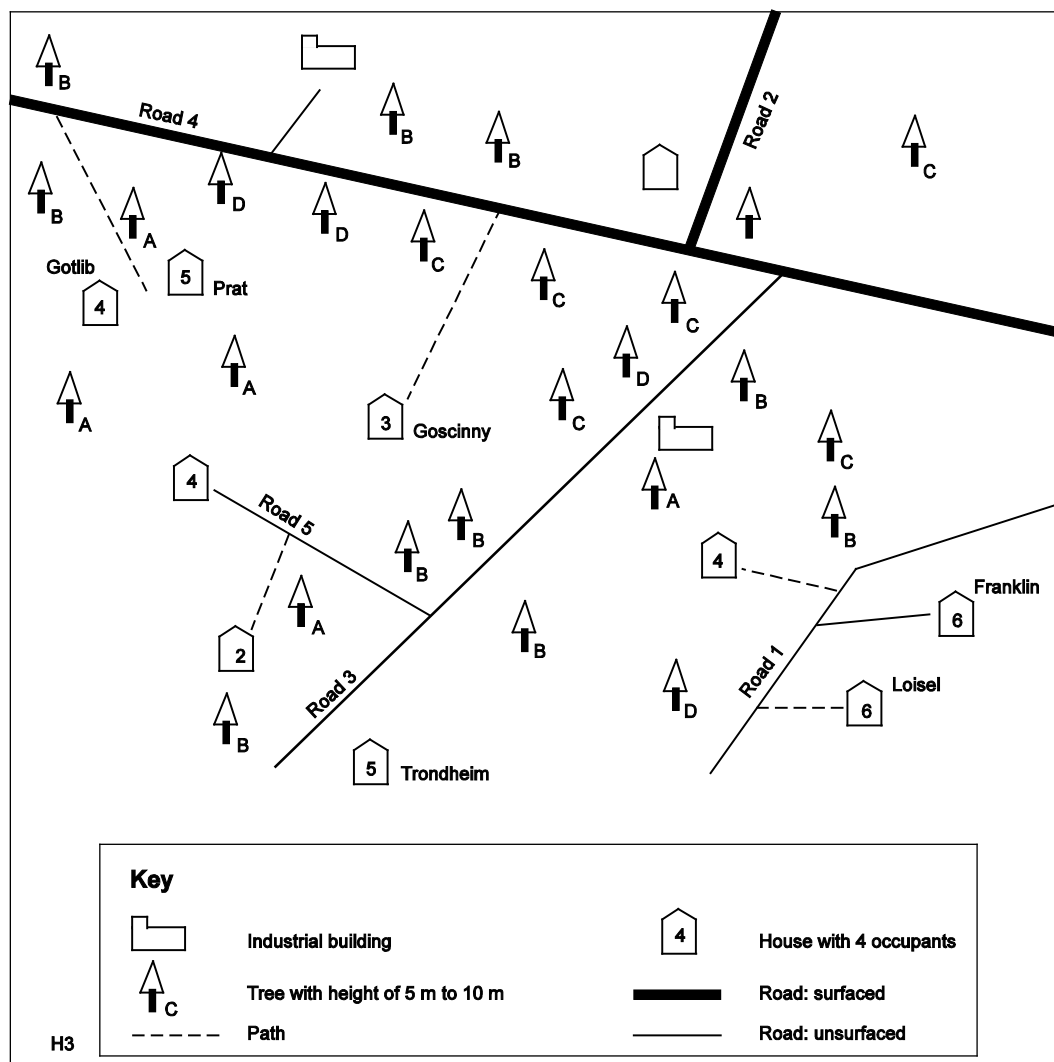
Figure E.3 — Graphical representation of the dataset

E.3 Quality evaluation process

E.3.1 Specify data quality unit(s)

A data quality unit is composed of a scope and quality element(s). In this example the completeness and thematic accuracy 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 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).

— The last quality unit is composed by a usability element (overall conformance to the data product specification requirement) evaluated on the whole dataset.

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

E.3.2 Specify data quality measures

The measures used in this example come from the list of registered measures provided in Annex D.

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 accuracy the following measure is used:

— measure 62, "misclassification matrix".

For describing usability the following measure is used:

— measure 101, "data product specification passed".

E.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 E.5.4.

E.3.4 Determine the output of the data quality evaluation (Result)

E.3.4.1   Identification of errors

By comparing the dataset, represented by Figure E.3, with the universe of discourse, represented by Figure E.2, a list of errors in the example dataset can be produced, represented by Figure E.4.
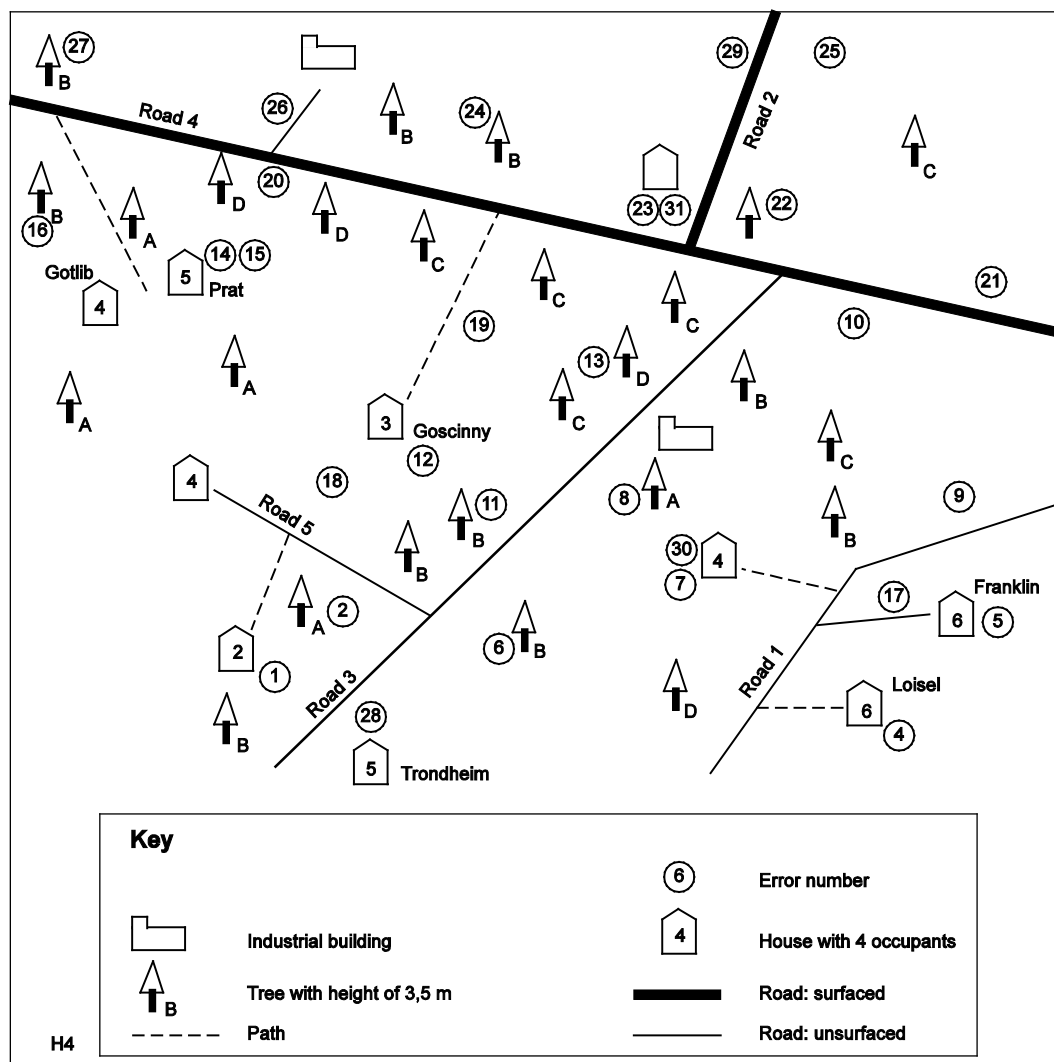
Figure E.4 — Graphical representation of dataset error locations

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).

— 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 accuracy 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.

E.3.4.2   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 E.2.

Table E.2 — Conformance result for logical consistency

Scope	Quality element	Data quality 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

E.3.4.3   Completeness

E.3.4.3.1   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 E.3 to E.5.

E.3.4.3.2   Quantitative result

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

Table E.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	3	43
Road	5	2	40	0	0
Tree	25	3	12	2	8
Industrial building	4	0	0	2	50
House	10	1	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

E.3.4.3.3   Derived conformance result

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

Table E.4 — Completeness conformance

Evaluation id	Quality element	Measure and measure id	Feature type	Requirement number	AQL	Error Count	Pop	Pass
1	Commission	Excess item (1)	Path	3	2	1	7	Yes
2	Omission	Missing item (5)	Path	2	2	3	7	No
3	Commission	Excess item (1)	Road	3	2	2	5	Yes
4	Omission	missing item (5)	Road	2	2	0	5	Yes
5	Commission	Excess item (1)	Tree	9	10%	3	25	No
6	Omission	Missing item (5)	Tree	8	10%	2	25	Yes
7	Commission	Excess item (1)	Industrial building	6	2	0	4	Yes
8	Omission	Missing item (5)	Industrial building	5	2	2	4	Yes
9	Commission	Excess item (1)	House	6	2	1	10	Yes
10	Omission	Missing item (5)	House	5	2	1	10	Yes

E.3.4.3.4   Aggregated conformance result

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

Table E.5 — Aggregated completeness conformance

Scope	Quality element	Data quality requirements	Number of evaluations and id (see Table E.4)	Counts yes/no	Pass
Transport Network	Omission	2) max two missing for each feature type	2 (evaluation No.2 and 4)	1/1	No
Transport Network	Commission	3) max two in excess for each feature type	2 (evaluation No.1 and 3)	2/0	Yes
Buildings	Omission	5) max two missing for each feature type	2 (evaluation No.8 and 10)	2/0	Yes
Buildings	Commission	6) max two in excess for each feature type	2 (evaluation No.7 and 9)	2/0	Yes

E.3.4.4   Thematic accuracy – classification correctness

E.3.4.4.1   General

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

E.3.4.4.2   Quantitative result

One way of depicting errors associated with thematic accuracy is by using the measure "misclassification matrix".

Table E.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 E.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 come from the missing and excess items.

E.3.4.4.3   Derived conformance result

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

Table E.7 — Thematic accuracy conformance

Evaluation id	Quality element	Measure	Feature type	Require-ment number	AQL	Mis-classification Count	Pass
11	Thematic classification correctness	number of incorrectly classified features	Path	4	2	2	Yes
12	Thematic classification correctness	number of incorrectly classified features	Road	4	2	0	Yes
13	Thematic classification correctness	number of incorrectly classified features	Industrial building	7	2	1	Yes
14	Thematic classification correctness	number of incorrectly classified features	House	7	2	0	Yes
15	Thematic classification correctness	number of incorrectly classified features	Tree	11	0	0	Yes

E.3.4.4.4   Aggregated conformance result

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

Table E.8 — Aggregated classification correctness conformance

Scope	Quality element	Data quality requirements	Number of evaluations and id (see Table E.7)	Counts yes/no	Pass
Transport Network	Thematic classification correctness	4) 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	7) max two feature instances misclassified as another of the Building feature types	2 (evaluation No. 13 and 14)	2/0	Yes

E.3.4.5   Thematic accuracy – quantitative attribute accuracy

E.3.4.5.1   General

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

E.3.4.5.2   Quantitative result

Attribute height of trees is shown in Table E.9.

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

Universe of discourse	Dataset				
	Class A 1 to 3 m	Class B 3 to 5 m	Class C 5 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.

E.3.4.5.3   Derived conformance result

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

Table E.10 — Thematic accuracy conformance

Quality element	Measure and measure id	Feature type / attribute	Requirement number	AQL	Misclassification Count	Pop	Pass
Quantitative attribute accuracy	Misclassification matrix (62)	Tree / height Class	10	20%	6	22	No

E.3.4.6   Usability – aggregated conformance to data product specification

In Table E.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", identifier 101 (see Table D.77).

Table E.11 — Usability – conformance to the data product specification

Scope	Quality element	Data quality requirements	Number of evaluations	Counts yes/no	Conformant
Dataset	Usability element	Overall data quality requirement: 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

E.4 Reporting data quality

E.4.1 Reporting as metadata

E.4.1.1   General

E.4.1.2 to E.4.1.4 present examples of how to report the quality results as metadata as described in this International Standard (Clause 10 and Annex C) and in ISO 19115-1:2013. Indeed, one instance of MD_Metadata aggregates one or more instances of DQ_DataQuality.

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

E.4.1.2   Reporting commission

Table E.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 E.12 — Reporting commission as metadata

XML element	Example	Comment
DQ_DataQuality		
scope : MD_Scope		
level: MD_ScopeCode	Dataset	Scope of this data quality unit
standaloneQualityReport: DQ_StandaloneQualityReportInformation		
reportReference: CI_Citation		Reference and abstract of the attached standalone quality report.
title: CharacterString	Reporting as standalone quality report, see E.4.2	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Creation	
abstract: CharacterString	The standalone quality report attached to this quality evaluation is providing more details on the derivation and aggregation method.	
report: DQ_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: DQ_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: DQ_FullInspection		
evaluationMethodType: DQ_EvaluationMethodTypeCode	directExternal	
evaluationMethodDescription: CharacterString	Compare count of items in the dataset against count of items in universe of discourse	

Table E.12 *(continued)*

XML element	Example	Comment
result: DQ_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	
value: Record	0	
valueUnit: UnitOfMeasure	None	
result: DQ_QuantitativeResult		
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Road	
value: Record	2	
valueUnit: UnitOfMeasure	None	
report: DQ_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: DQ_MeasureReference		
nameOfMeasure: CharacterString	excess item	
measureIdentification: MD_Identifier		
code: CharacterString	1	
measureDescription: CharacterString	Indication that an item is incorrectly present in the data	
evaluation: DQ_AggregationDerivation		
evaluationMethodType: DQ_EvaluationMethodTypeCode	indirect	
evaluationMethodDescription: CharacterString	Derivation from quantitative result	
derivedElement: DQ_Element	***quantitative_commission***	Reference to the original results

Table E.12 *(continued)*

XML element	Example	Comment
result: DQ_ConformanceResult		Derived conformance result for the path commission
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	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.
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Path	
specification: CI_Citation		
title: CharacterString	Data product specification (see E.2.1) requirement 2	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Creation	
pass: Boolean	True	
result: DQ_ConformanceResult		Derived conformance result for the road commission
resultScope: MD_Scope		
level: MD_ScopeCode	featureType	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.
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	Road	
specification: CI_Citation		
title: CharacterString	Data product specification (see E.2.1) requirement 2	
date: CI_Date		
date: Date	2010-07-05	
dateType: CI_DateTypeCode	Creation	
pass: Boolean	true	
DQ_DataQuality ***id = agg_commission1***		Aggregated conformance result for Transport Network
scope : MD_Scope		The scope is now the feature types for Transport Network => the data quality unit changed. That is why a new instance of DQ_DataQuality was created.
level: MD_ScopeCode	FeatureType	
levelDescription: MD_ScopeDescription		
features: GF_FeatureType	TransportNetwork (road and path)	
report: DQ_Commission		

Table E.12 *(continued)*

XML element			Example	Comment
evaluation: DQ_AggregationDerivation				Aggregation method
	evaluationMethodType: DQ_EvaluationMethodTypeCode		indirect	
	evaluationMethodDescription: CharacterString		100% pass fail aggregation of the conformance commission result for roads and paths	
	evaluationProcedure: CI_Citation			
		title: CharacterString	Annex J	
		Date: CI_Date		
		date: Date	2010-07-05	
		dateType: CI_DateTypeCode	Creation	
	derivedElement: DQ_Element		***conformance_commission***	Reference to the original results
	result: DQ_ConformanceResult			
	specification: CI_Citation			
		title: CharacterString	Data product specification (see E.2.1), requirement 2	
		date: CI_Date		
		date: Date	2010-07-05	
		dateType: CI_DateTypeCode	Creation	
	Pass: Boolean		true	

E.4.1.3   Reporting classification correctness

Table E.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 E.13 — Reporting classification correctness as metadata

XML element			Example	Comment
DQ_DataQuality				
	scope : MD_Scope			
		level: MD_ScopeCode	Dataset	Scope of this data quality unit
	standaloneQualityReport: DQ_StandaloneQualityReportInformation			

Table E.13 *(continued)*

XML element				Example	Comment
reportReference: CI_Citation					Reference and abstract of the attached standalone quality report.
	title: CharacterString			Reporting as standalone quality report see E.4.2	
	date: CI_Date				
		date: Date		2010-07-05	
		dateType: CI_DateTypeCode		Creation	
	abstract: CharacterString			The standalone quality 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: DQ_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: DQ_MeasureReference				
		nameOfMeasure: CharacterString		number of incorrectly classified features	
		measureIdentification: MD_Identifier			
			code: CharacterString	60	
	evaluation: DQ_AggregationDerivation				
		evaluationMethodType: DQ_EvaluationMethodTypeCode		indirect	
		evaluationMethodDescription: CharacterString		Derivation from quantitative results reported in the standalone quality report	
	standaloneQualityReportDetails: CharacterString			The original quantitative results are described in E.3.4.4.2 of the standalone quality report.	Reference to the original results
	result: DQ_ConformanceResult				Derived conformance result for the industrial buildings classification.
		resultScope: MD_Scope			
			level: MD_ScopeCode	featureType	The original quantitative result is intentionally not provided in metadata. It is described in the standalone quality report.
			levelDescription: MD_ScopeDescription		
			features: GF_FeatureType	Industrial Building	
		specification: CI_Citation			The attribute standaloneQualityReportDetails give the precise reference to the original result within the standalone quality report.
			title : CharacterString	Data product specification (see E.2.1), requirement 7	

Table E.13 *(continued)*

XML element					Example	Comment
			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 E.3.4.4.2 of the standalone quality report.	
		pass: Boolean			True	
	result: DQ_ConformanceResult					Derived conformance result for the industrial buildings classification. The original quantitative result is intentionally not provided in metadata. It is described in the standalone quality report. The attribute standaloneQualityReport Details give the precise reference to the original result within the standalone quality 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 E.2.1), requirement 7	
			date: CI_Date			
				date: Date	2010-07-05	
				dateType: CI_DateTypeCode	Creation	
		explanation: CharacterString			The original quantitative result is provided in standalone quality report.	
		pass: Boolean			True	
DQ_DataQuality *id = agg_classification2*						Aggregated classification correctness result for Buildings
	Scope : MD_Scope					The scope is now the Building feature types => the data quality unit changed. That is why a new instance of DQ_DataQuality was created.
		level: MD_ScopeCode			FeatureType	
		levelDescription: MD_ScopeDescription				
			features: GF_FeatureType		Buildings (industrial building and house)	
	report: DQ_ThematicClassificationCorrectness					
		evaluation: DQ_AggregationDerivation				Aggregation method
			evaluationMethodType: DQ_EvaluationMethodTypeCode		Indirect	
			evaluationMethodDescription: CharacterString		100% pass fail aggregation of the conformance classification correctness result for industrial buildings and houses	
			evaluationProcedure: CI_Citation			

125

Table E.13 *(continued)*

XML element			Example	Comment
		title: CharacterString	Annex J	
		date: CI_Date		
			date: Date	2010-07-05	
			dateType: CI_DateTypeCode	Creation	
	derivedElement: DQ_Element		***conformance_classification***	Reference to the original results
	result: DQ_ConformanceResult			
		specification: CI_Citation		
			title: CharacterString	Data product specification (see E.2.1), requirement 7	
			date: CI_Date		
				date: Date	2010-07-05	
				dateType: CI_DateTypeCode	Creation	
		pass: Boolean	True	

E.4.1.4   Reporting conformance to the data product specification using Usability

Table E.14 presents an example of how to express the conformance to the data product specification by aggregating the results for the different requirements. The quality element used for that is Usability.

Table E.14 — Reporting usability as metadata

XML element				Example	Comment
DQ_DataQuality					
	scope : MD_Scope				
		level: MD_ScopeCode		Dataset	
	standaloneQualityReport: DQ_StandaloneQualityReportInformation				Reference and abstract of the attached standalone quality report.
		reportReference: CI_Citation			
			title: CharacterString	Reporting as standalone quality report see E.4.2	
			date: CI_Date		
				date: Date	2010-07-05	
				dateType: CI_DateTypeCode	Creation	
		abstract: CharacterString		The standalone quality report attached to this quality evaluation is providing fully detailed information about the evaluation applied and results obtained.	

Table E.14 *(continued)*

XML element				Example	Comment
report: DQ_UsabilityElement					This element is used to report the conformance of the dataset to the data product specification.
	measure: DQ_MeasureReference				
		nameOfMeasure: CharacterString		Data product specification passed	
		measureIdentification: MD_Identifier			
			code: CharacterString	101	
		measureDescription: CharacterString		Indication that all requirements in the referred data product specification are fulfilled.	
	evaluation: DQ_AggregationDerivation				
		evaluationMethodType: DQ_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 J	
			date: CI_Date		
				date: Date	2010-07-05
				dateType: CI_DateTypeCode	Creation
	standaloneQualityReportDetails: CharacterString			The original results are described in E.3.4.2, E.3.4.3.4, E.3.4.4.4 and E.3.4.5.3 of the standalone quality report.	Reference to the original results in the standalone quality report (conceptual consistency conformance result, quantitative attribute accuracy conformance result for tree heights…)
	derivedElement: DQ_Element			***agg_commission1***	Reference to the aggregated commission conformance result for transport network described previously in the metadata
	derivedElement: DQ_Element			*(id)*	Reference to the aggregated commission conformance result for buildings described previously in the metadata
	derivedElement: DQ_Element			*(id)*	Reference to the commission conformance result for trees described previously in the metadata

Table E.14 *(continued)*

XML element				Example	Comment
	derivedElement: DQ_Element			*(id)*	Reference to the aggregated omission conformance result for transport network described previously in the metadata
	derivedElement: DQ_Element			*(id)*	Reference to the aggregated omission conformance result for buildings described previously in the metadata
	derivedElement: DQ_Element			*(id)*	Reference to the omission conformance result for trees described previously in the metadata
	derivedElement: DQ_Element			*(id)*	Reference to the aggregated classification correctness conformance result for transport network described previously in the metadata
	derivedElement: DQ_Element			***agg_classification2***	Reference to the aggregated classification correctness conformance result for buildings described previously in the metadata
	derivedElement: DQ_Element			*(id)*	Reference to the classification correctness conformance result for trees described previously in the metadata
	result: DQ_ConformanceResult				
		specification: CI_Citation			
			title: CharacterString	Data product specification (see E.2.1)	
			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	

E.4.2 Reporting in a standalone quality report

The structure of the standalone quality report is free. E.2 and E.3 are examples of standalone quality reports.

E.5 Additional examples

E.5.1 General

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

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

E.5.2 Reporting descriptive results as metadata

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

Table E.15 — Reporting descriptive result as metadata

XML element			Example	Comment
DQ_DataQuality				
	scope : MD_Scope			
		level: MD_ScopeCode	Dataset	The dataset is describing archaeological objects
	report: DQ_RelativeInternalPositionalAccuracy			
		evaluation: DQ_IndirectEvaluation		
		evaluationMethodType: DQ_EvaluationMethodTypeCode	indirect	
		evaluationMethodDescription: CharacterString	Compare 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: DQ_DescriptiveResult		
		statement : CharacterString	Relative positional accuracy between archaeological objects and rivers is higher than the absolute positional accuracy of the archaeological objects (5 metres)	

E.5.3 Reporting metaquality as metadata

The absolute positional accuracy of the topological survey on an archaeological site is evaluated: the result is 5 metres 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 E.16 describes how to report metaquality as metadata.

Table E.16 — Reporting metaquality as metadata

XML element	Example	Comment
DQ_DataQuality		
scope : MD_Scope		
level: MD_ScopeCode	Dataset	
report: DQ_AbsolutExternalPositionalAccuracy ***id=positionalaccuracy1***		Absolute positional accuracy report.
measure: DQ_MeasureReference		An id is provided to the data quality element in order to be able to reference it in the following metaquality element.
nameOfMeasure: CharacterString	Root mean square error	
measureIdentification: MD_Identifier		
code: CharacterString	39	
measureDescription: CharacterString	Standard deviation where the true value is not estimated from the observations but known a priori	All optional attributes have not been filled here.
evaluation: DQ_FullInspection		
evaluationMethodType: DQ_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: DQ_QuantitativeResult		
value: Record	5	
valueUnit: UnitOfMeasure	Metre	
report: DQ_Confidence		Metaquality report (confidence) related to the previous accuracy report.
relatedElement : DQ_Element	***positionalaccuracy1***	
measure: DQ_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: DQ_FullInspection		

Table E.16 *(continued)*

XML element			Example	Comment
	evaluationMethodType: DQ_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: DQ_QuantitativeResult			
	value: Record		2.4	
	valueUnit: UnitOfMeasure			

E.5.4 How to report sampling procedure

This example is based upon a Topographic Database (TDB) 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 E.17.

Table E.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 x 1 km squares are selected so that the number of features in the sample is sufficient for an AQL = 4.

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

Table E.18 — Reporting sampling evaluation as metadata

XML element	Example
DQ_DataQuality	
scope : MD_Scope	
level: MD_ScopeCode	Feature Type
levelDescription: MD_ScopeDescription	
features: GF_FeatureType	Road
report: DQ_Commission	
measure: DQ_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: DQ_SampleBasedInspection	
evaluationMethodType: DQ_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 F
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 (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 x 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 x 1 km) is recommended as a practical lot size.

Annex F
(informative)

Sampling methods for evaluating

F.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 3951-1:2005 may be applied. These standards were originally developed for non-spatial use. This Annex describes how to apply the ISO 2859 series and ISO 3951-1:2005 and other spatial sampling techniques to geographic data.

F.2 Lot and item

Lot and item are important concepts in the sampling inspection method specified in the ISO 2859 series and ISO 3951-1:2005. 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.

F.3 Sample size

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

The difference between the perspectives is illustrated in Figure F.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.

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

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 F.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

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

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

F.4 Sampling strategies

F.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 F.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 F.2 — Sampling strategy relationships

F.4.2 Probabilistic versus judgemental sampling

F.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.

F.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.

F.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.

F.4.2.4 Semi-random sampling

Semi-random or systematic sampling applies random selection of the initial sample items (e.g. location, time, 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.

F.4.3 Feature-guided versus area-guided sampling

F.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 may 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.

F.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 x 1 km grid areas in order to evaluate the attributes of the objects contained in that area.

Figure F.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).
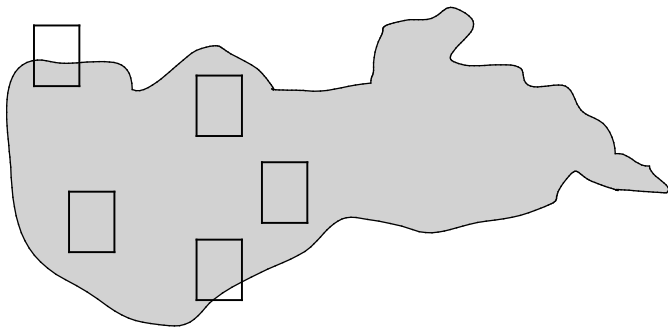
Figure F.3 — Example of area-guided random sampling

When coverage of the entire area is important, then the sample locations should be determined according to a regular or semi-regular pattern. Figure F.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.
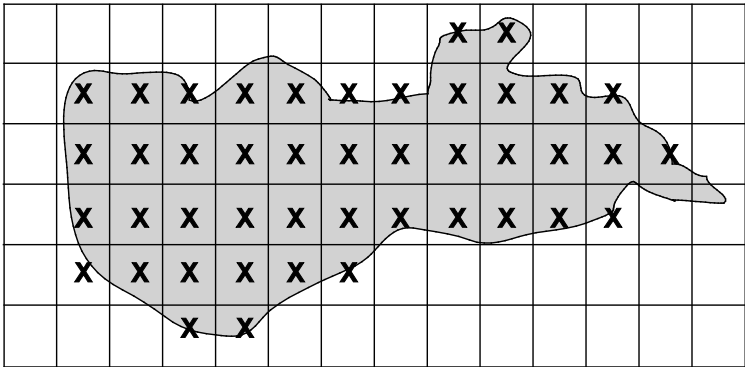
Figure F.4 — Example of area-guided regular and non-random sampling

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.

F.5 Probability-based sampling

F.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.

F.5.2 Existing standard for inspection by sampling

F.5.2.1 General

Based on the characteristics of production and in accordance with the data product specification, suitable International Standards 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 is applied for skip-lot sampling procedures. ISO 3951-1:2005 is 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 3534-2:2006. It was previously called acceptable quality level in ISO 2859-1, ISO 2859-3 and ISO 3951-1:2005 and LQ (limiting quality) in the case of ISO 2859-2 based on the data product specification.

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 3951-1:2005, quality statistics should be specified based on the data product specification.

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

F.5.2.2.1 General

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

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

Table F.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 (reference [20]). It is assumed that the deviations fit this distribution.

How to use the table:

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

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

3) Carry out the evaluation, and count number of "fail items";

4) 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 F.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	1200	80	3	3	5	6	7	8
1201	3200	125	3	4	6	8	10	11
3201	10000	200	4	6	8	11	14	16
10001	35000	315	5	7	12	16	20	23
35001	150000	500	6	10	16	23	28	34
150001	500000	800	9	14	24	33	42	51
> 500000		1250	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 values ranges than the one presented in Table F.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 2440 buildings.

Sample size (from Table F.2) is n = 125. Field check shows that 2 buildings are missing, giving an estimated missing rate of: $2/(125+2) \times 100\% = 1,6\%$.

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

1,6% is higher than 0,5%, but can the dataset 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 F.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.

F.5.2.2.3   Standard deviation

Table F.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 F.4 below is based on normal distribution, and assumes normal distribution of deviations.

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

Table F.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	$\dfrac{s}{\sqrt{F}}.s \times \sqrt{F}$
Standard deviation too high if:	$\sigma < \dfrac{s}{\sqrt{F}}$

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 F.4) is higher than the AQL.

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

Population size From	To	Sample size (n)	$\sqrt{F_{0.05,n-1,\infty}}$
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 = 21cm, Accepted Quality Level (AQL) = 19cm.

Lower limit for confidence interval = 21cm/1,23 (from Table F.4) = 17,1 cm. The AQL (19cm) 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.

F.5.3 Sampling process

F.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.

F.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.

F.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.

F.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.

F.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 International Standards.

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.

F.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 G
(normative)

Data quality basic measures

G.1 Purpose of data quality basic measures

The concept of data quality basic measure is introduced in this International Standard to avoid the repetitive definition of the same concept. There are data quality measures that have certain commonalities. For example, the counting-related data quality measures are dealing with the concept of counting errors. The number of errors may be used to construct different kind of data quality measures. The concept of constructing these data quality measures is defined for the generic data quality basic measures and are used for the creation of data quality measures that share these commonalities.

Counting- and uncertainty-related data quality measures can be identified. Therefore two principle categories of data quality basic measures are listed in this Annex. The counting-related data quality basic measures are based on the concept of counting errors or correct items. The uncertainty-related data quality basic measures are based on the concept of modelling the uncertainty of measurements with statistical methods. The measured quantity can be embedded in different dimensions. Depending on the dimension of the measured quantity, different types of data quality basic measures are used to construct data quality measures.

G.2 Counting-related data quality basic measures

The data quality basic measures based on different methods of counting errors or counting the number of correct values is listed in Table G.1.

Table G.1 — Data quality basic measures for counting-related data quality measures

Data quality basic measure name	Data quality basic measure definition	Example	Data quality value type
Error indicator	Indicator that an item is in error	False	Boolean (if the value is true the item is not correct)
Correctness indicator	Indicator that an item is correct	True	Boolean (if the value is true the item is correct)
Error count	Total number of items that are subject to an error of a specified type	11	Integer
Correct items count	Total number of items that are free of errors of a specified type	571	Integer
Error rate	Number of the erroneous items with respect to the total number of items that should have been present	0,0189	Real
Correct items rate	Number of the correct items with respect to the total number of items that should have been present	0,9811	Real
NOTE 1 Error rate can either be presented as percentage or as a ratio. The value unit in the quantitative result (see 7.5.4.2) can be used to specify that the result is presented in percentage or as a ratio. NOTE 2 Correct items rate can either be presented as percentage or as a ratio. The value unit in the quantitative result (see 7.5.4.2) can be used to specify that the result is presented in percentage or as a ratio.			

NOTE   Number of items is defined using number of items in the universe of discourse for the dataset specified by data quality scope.

EXAMPLE   Use number of items found in the real world or reference dataset.

G.3 Uncertainty-related data quality basic measures

G.3.1 General

Numerical values that are obtained by measurement can only be observed to a certain accuracy. By treating the measured quantity as a random variable, this uncertainty can be quantified. The different ways of describing uncertainty with statistical methods are used for the definition of uncertainty-related data quality basic measures.

The statistical methods used for the definition of uncertainty-related data quality measures are based on certain assumptions:

— uncertainties are homogeneous for all observed values;

— the observed values are not correlated;

— the observed values have a normal distribution.

G.3.2 One-dimensional random variable, Z

For a measured quantity that takes real values, it is impossible to give the probability of a single value to be the true value. But it is possible to give the probability for the true value to be within a certain interval. This interval is called the confidence interval. It is given by the probability P of the true value being between the lower and the upper limit. This probability P is also called the significance level.

P(lower limit $\leq$ true value $\leq$ upper limit)=P

If the standard deviation σ is known, the limits are given by the quantiles u of the normal (Gaussian) distribution $P(z_t - u \times \sigma \leq \text{true value} \leq z_t + u \times \sigma) = P$.

See also Table G.2

Table G.2 — Relation between the quantiles of the normal distribution and the significance level

Probability P	Quantile	Data quality basic measure	Name	Data quality value type
P = 50 %	$u_{50\%}$ 0,6745	$u_{50\%} \cdot \sigma_Z$	LE50	Measure
P = 68,3 %	$u_{68,3\%} = 1$	$u_{68,3\%} \cdot \sigma_Z$	LE68.3	Measure
P = 90 %	$u_{90\%} = 1,645$	$u_{90\%} \cdot \sigma_Z$	LE90	Measure
P = 95 %	$u_{95\%} = 1,960$	$u_{95\%} \cdot \sigma_Z$	LE95	Measure
P = 99 %	$u_{99\%} = 2,576$	$u_{99\%} \cdot \sigma_Z$	LE99	Measure
P = 99,8 %	$u_{99,8\%} = 3$	$u_{99,8\%} \cdot \sigma_Z$	LE99.8	Measure

If the standard deviation σ is unknown, but the one-dimensional random variable Z is measured redundantly by N independent observations, it is possible to estimate the standard deviation from the observations (see Table G.3).

z_{mi} represents the ith measurement for the value. If the true value z_t for Z is known, the standard deviation can be estimated by

$$s_Z = \sqrt{\frac{1}{r}\sum_{i=1}^{N}(z_{mi} - z_t)^2}$$

with redundancy r being the number of observations $r = N$. If the true value is unknown, it may be estimated as the arithmetic mean of the observations $z_t = \sum_{i=1}^{N} z_{mi}$.

The standard deviation may then be estimated using the same formula, with $r = N - 1$.

If the standard deviation is estimated by redundant measurements, the confidence interval can be derived from the Student's t-distribution with parameter r:

$$P\left(-t \times s_z \le Z - z_t \le t \times s_z\right) = P \text{ with } \frac{(Z - z_t)}{s_z} \sim t(r)$$

Table G.3 — Relation between the quantiles of the Student's t-distribution and the significance level for different redundancies r

Probability P	Quantile for r = 10	Quantile for r = 5	Quantile for r = 4	Quantile for r = 3	Quantile for r = 2	Quantile for r = 1
P = 50 %	t = 1,221	t = 1,301	t = 1,344	t = 1,423	t = 1,604	t = 2,414
P = 68,3 %	t = 1,524	t = 1,657	t = 1,731	t = 1,868	t = 2,203	t = 3,933
P = 90 %	t = 2,228	t = 2,571	t = 2,776	t = 3,182	t = 4,303	t = 12,706
P = 95 %	t = 2,634	t = 3,163	t = 3,495	t = 4,177	t = 6,205	t = 25,452
P = 99 %	t = 3,581	t = 4,773	t = 5,598	t = 7,453	t = 14,089	t = 127,321
P = 99,8 %	t = 4,587	t = 6,869	t = 8,610	t = 12,924	t = 31,599	t = 636,619

Table G.4 — Data quality basic measures for different probabilities P of a one-dimensional quantity, where the standard deviation is estimated from redundant measurements

Probability P	Data quality basic measure	Name	Data quality value type
P = 50,0 %	$t_{50\%}(r) \cdot s_Z$	LE50(r)	Measure
P = 68,3 %	$t_{68,3\%}(r) \cdot s_Z$	LE68.3(r)	Measure
P = 90,0 %	$t_{90\%}(r) \cdot s_Z$	LE90(r)	Measure
P = 95,0 %	$t_{95\%}(r) \cdot s_Z$	LE95(r)	Measure
P = 99,0 %	$t_{99\%}(r) \cdot s_Z$	LE99(r)	Measure
P = 99,8 %	$t_{99,8\%}(r) \cdot s_Z$	LE99.8(r)	Measure
NOTE The values of t for a number of redundancies r can be obtained from Table G.3			

The data quality basic measures for the uncertainty of one-dimensional quantities are given in Table G.2 and Table G.4. They both aim to measure the uncertainty by giving the upper and lower limit of a confidence interval. The difference is in how the standard deviation is obtained. If it is known *a priori*, then Table G.2 is relevant. If the standard deviation is estimated from redundant measurements, then Table G.4 in conjunction with Table G.3 is relevant.

G.3.3 Two-dimensional random variable X and Y

The case of the one-dimensional random variable Z can be expanded to two dimensions where the measured quantity is always observed by two values. The result is given by the tuple X, Y. This has the same assumptions as in the case of the one-dimensional random variable.

The observations are x_{mi} and y_{mi}. The equivalence of the confidence interval in one dimension is the confidence area, which is usually described as a circle around the best estimation for the true value. The probability for the true value to lie in this area is calculated by area integration over the two-dimensional density function of the normal distribution. A circular area is characterized by its radius. This radius, R, is used as measure for the accuracy of two-dimensional random variables (see also Table G.5):

$$P(R, \sigma_X, \sigma_Y) = \frac{1}{2\pi \sigma_X \sigma_Y} \iint\limits_{(x-x_t)^2 + (y-y_t)^2 = R^2} e^{-\frac{1}{2}\left(\frac{(x-x_t)^2}{\sigma_X^2} + \frac{(y-y_t)^2}{\sigma_Y^2} \right)} \mathrm{d}x \mathrm{d}y$$

For some particular probabilities, the radius can be calculated depending on the standard deviations σ_x and σ_y.

Table G.5 — Relationship between the probability P and the corresponding radius of the circular area

Probability P	Data quality basic measure	Name	Data quality value type
P = 39,4 %	$\dfrac{1}{\sqrt{2}}\sqrt{\sigma_x^2 + \sigma_y^2}$	CE39.4	Measure
P = 50 %	$\dfrac{1,1774}{\sqrt{2}}\sqrt{\sigma_x^2 + \sigma_y^2}$	CE50	Measure
P = 90 %	$\dfrac{2,146}{\sqrt{2}}\sqrt{\sigma_x^2 + \sigma_y^2}$	CE90	Measure
P = 95 %	$\dfrac{2,4477}{\sqrt{2}}\sqrt{\sigma_x^2 + \sigma_y^2}$	CE95	Measure
P = 99,8 %	$\dfrac{3,5}{\sqrt{2}}\sqrt{\sigma_x^2 + \sigma_y^2}$	CE99.8	Measure

G.3.4 Three-dimensional random variable X, Y, Z

The case of the one-dimensional random variable Z can be expanded to three dimensions where the result is always observed by three values. The result is given by the tuple X, Y, Z. They underlay the same assumptions as in the case of the one-dimensional random variable.

The observations are x_{mi}, y_{mi} and z_{mi}. The equivalence of the confidence interval in one dimension is the confidence volume, which is usually described as a sphere around the best estimation for the true value. The probability for the true value to lie in this volume is calculated by volume integration over the three-dimensional density function of the normal distribution. A spherical volume is characterized by its radius. This radius is used as measure for the accuracy of three-dimensional random variables (see Table G.6).

Table G.6 — Relationship between the probability P and the corresponding radius of the spherical volume

Probability P	Data quality basic measure	Name	Data quality value type
P = 50 %	$0,51 \cdot \left(\sigma_x + \sigma_y + \sigma_z \right)$	spherical error probable (SEP)	Measure
P = 61 %	$\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$	mean radial spherical error (MRSE)	Measure
P = 90 %	$0,833 \cdot \left(\sigma_x + \sigma_y + \sigma_z \right)$	90 % spherical accuracy standard	Measure
P = 99 %	$1,122 \cdot \left(\sigma_x + \sigma_y + \sigma_z \right)$	99 % spherical accuracy standard	Measure

Annex H
(informative)

Management of data quality measures

H.1 Introduction

This Annex provides the description of how to store data quality measures, basic measures and parameters in a register or a catalogue.

H.2 Storage of data quality measures

H.2.1 General

Full description of data quality measures, data quality basic measures and parameters may be stored either in a register, or in a catalogue. These two types of organization are compatible and complement each other. The register is used for global use case (e.g. register for all the measures used in an organization) and the catalogue present a set of information specific to one particular use case (e.g. catalogue for the set of measures used for the data quality evaluation of one particular dataset).
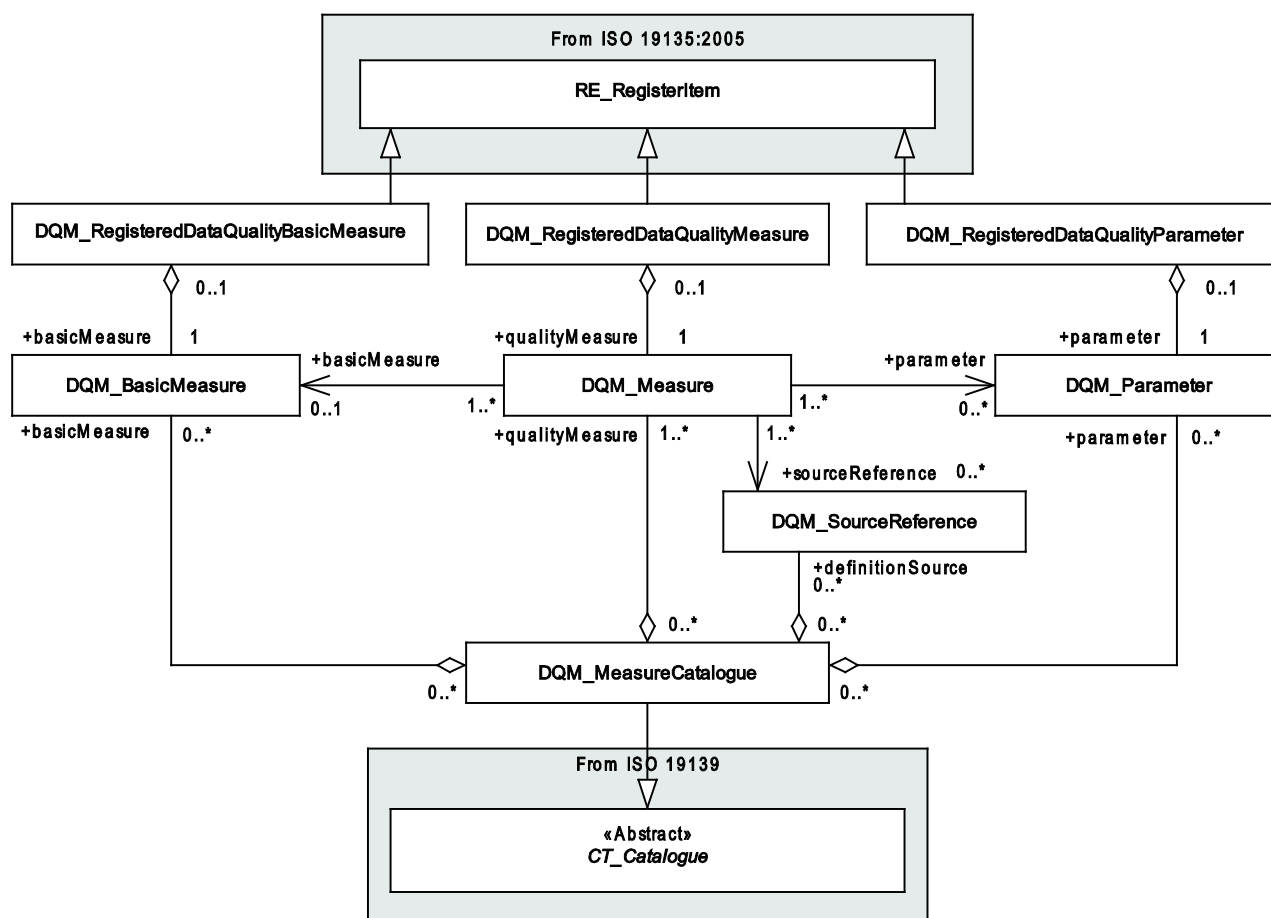
Figure H.1 — Registered items, catalogue and data quality measures

H.2.2 Catalogue of data quality measures

Measures, basic measures, source references and parameters may be provided within a measure catalogue: DQM_MeasureCatalogue, derived from the class CT_Catalogue defined in ISO/TS 19139:2007.

DQM_MeasureCatalogue should aggregate all wanted instances of DQM_Measure, DQM_BasicMeasure, DQM_SourceReference and DQM_Parameter as shown in Figure H.1.

H.2.3 Register of data quality measures

In order to manage data quality measures, a register of data quality measure may be created. In this case, the register of data quality measures should follow the register specification provided in ISO 19135:2005, which describes the structure and attributes of registered items.

Figure H.2 presents the structure of the class RE_RegisteredItem compared to the classes DQM_Measure, DQM_BasicMeasure and DQM_Parameter.
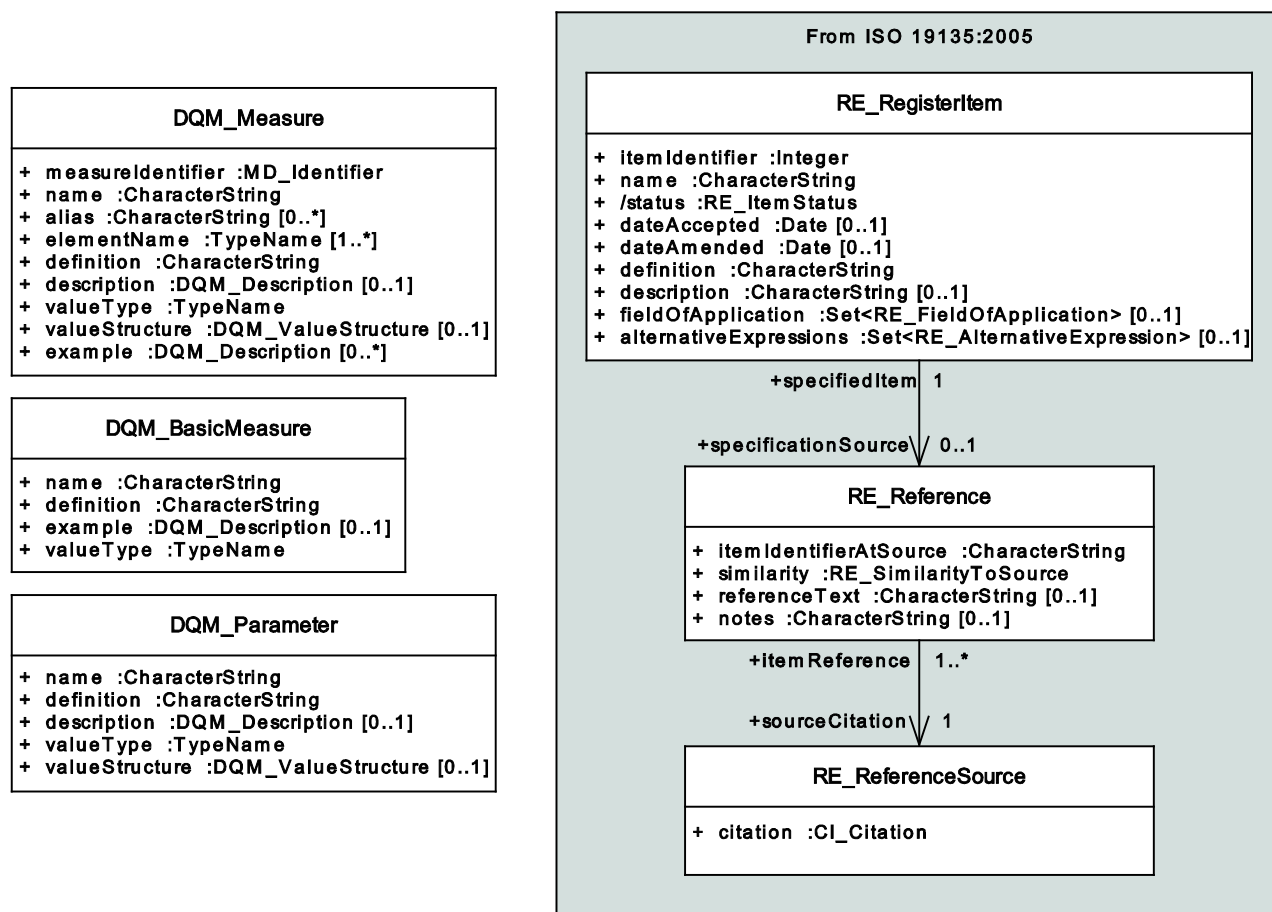
Figure H.2 — Structural similarities between registered items and data quality measures

Some descriptors of the data quality measures, basic measures and parameters (as defined in Clause 8) may be reused as the attributes of registered measures, basic measures and parameters (see Figure H.1 and Table H.1) derived from RE_RegisteredItem defined in ISO 19135:2005. The other descriptors of registered items should be provided in compliance with ISO 19135:2005.

Table H.1 — Measures, basic measures and parameters attributes corresponding to registered items attribute
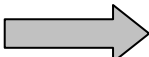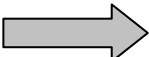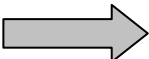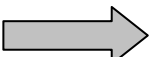
19157 measure element		19135 element
Registered data quality measure		
DQM_Measure.name		DQM_RegisteredDataQualityMeasure.name
DQM_Measure.definition		DQM_RegisteredDataQualityMeasure.definition
DQM_Measure.description.textDescription		DQM_RegisteredDataQualityMeasure.description
DQM_Measure.alias		DQM_RegisteredDataQualityMeasure.alternativeExpressions
DQM_Measure.measureIdentifier.code		DQM_RegisteredDataQualityMeasure.specifiedItem.itemIdAtSource
DQM_Measure.measureIdentifier.authority		DQM_RegisteredDataQualityMeasure.specifiedItem.sourceCitation
Registered data quality basic measures		
DQM_BasicMeasure.name		DQM_RegisteredDataQualityBasicMeasure.name
DQM_BasicMeasure.definition		DQM_RegisteredDataQualityBasicMeasure.definition
Registered data quality parameters		
DQM_Parameter.name		DQM_RegisteredDataQualityParameter.name

Table H.2 presents an example of the registered Measure 11 (see Table D.11).

Table H.2 — Example of registered item element - Measure 11

Registered Item element	Example value
DQM_RegisteredDataQualityMeasure.itemIdentifier	Identifier of the item within the register. Example: "1"
DQM_RegisteredDataQualityMeasure.status	Status of the item within the register
DQM_RegisteredDataQualityMeasure.name	"Number of invalid overlaps of surface"
DQM_RegisteredDataQualityMeasure.definition	"total number of erroneous overlaps within the data"
DQM_RegisteredDataQualityMeasure.description	"Which surfaces may overlap and which shall not is application dependent. Not all overlapping surfaces are necessarily erroneous. When reporting this data quality measure, the types of feature classes corresponding to the illegal overlapping surfaces shall be reported as well. "
DQM_RegisteredDataQualityMeasure.alternativeExpressions	"overlapping surfaces"
DQM_RegisteredDataQualityMeasure.specifiedItem.itemIdAtSource	"11"
DQM_RegisteredDataQualityMeasure.specifiedItem.sourceCitation	CI_Citation for ISO 19157

Annex I
(informative)

Guidelines for the use of Quality Elements

I.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 7.3.

I.2 Data quality element categories

I.2.1   General

Six different quality element categories are defined in 7.3:

— Completeness (7.3.2);

— Logical consistency (7.3.3);

— Positional accuracy (7.3.4);

— Thematic accuracy (7.3.5);

— Temporal quality (7.3.6);

— Usability element (7.3.7).

The usability element is used for a quality evaluation based on user requirements which can not be covered by the five others data quality categories. It may also be used to provide an aggregation result where results from several data quality categories are aggregated (for example, overall conformity to one specification). It is not further handled in this Annex.

Of the remaining five, 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 accuracy are used to describe how the dataset relates to the universe of discourse.

The last category (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).

I.2.2   Other candidates

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. Two possible (but not accepted) candidates "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 represent todays real world), the result will only be valid 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.

The two accepted data quality elements completeness and accuracy both describe how well the dataset represents real world at a given point of time. These two, possibly combined with metadata giving update policy for the dataset, will cover the need for up-to-dateness and timeliness.

I.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 I.1.
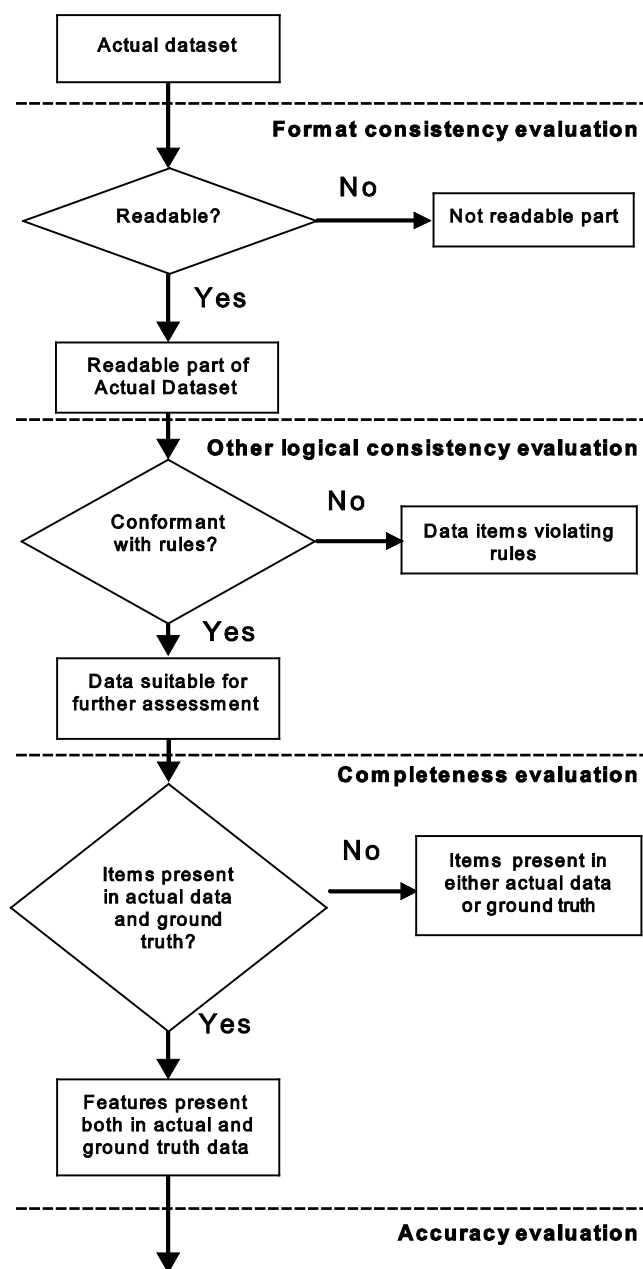
Figure I.1 — Ordering in data quality evaluation

When evaluating data quality, the usual ordering is:

1) 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.

2) 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.

3) 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.

4) 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.

I.3 The relationships between the data quality elements

I.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.

I.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 have 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 +-1m (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 metres, then the attribute completeness would become better (100%), but the attribute accuracy will probably become poorer.

I.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 vs. 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 vs. 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 vs. time: "The price ($/m^2) for the specific parcel is wrong by $20", or "this was the correct price 10 years ago"

I.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:

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

2) 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 I.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).
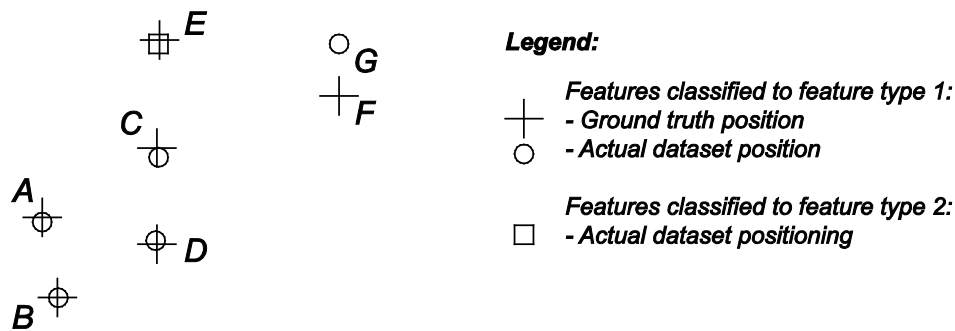
Figure I.2 — Accuracy versus completeness

I.4 Data quality elements – example of use

I.4.1 Completeness

I.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.

I.4.1.2 commission – excess data present in a dataset

This may be applied at the feature instance level. This means that data is 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

I.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.

I.4.2   Logical consistency

I.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.

I.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.

I.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"

I.4.2.4   format consistency – degree to which data is 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 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 may 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 shall 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 UPRN feature instance.

I.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.) should be reported as topological consistency errors.

I.4.3   Positional accuracy

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

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, and 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,

155

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

I.4.4 Temporal quality

I.4.4.1 General

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

I.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.

I.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 lifecycle 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".

I.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".

I.4.5 Thematic accuracy

I.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.

I.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);

EXAMPLE   Within a certain organization, this definition is used strictly. Classifications which are not defined within the dataset specification are not considered as classification correctness (these are considered to be domain consistency).

I.5 Discussions on difficult cases

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

At feature type level, completeness and thematic accuracy/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.

An example of this is provided in Annex E.

I.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 I.1.

Table I.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 J
(informative)

Aggregation of data quality results

J.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 telling 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 J.2 to J.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 10.2.1.

J.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 is determined by the equation,

$ADQR = v_1 * v_2 * v_3 * . . . * v_n$, 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.

J.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 is determined by the equation,

$ADQR = v_1 * w_1 + v_2 * w_2 + v_3 * w_3 + . . . + v_n * w_n$, 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 J.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 J.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 doesn't mean much to end-users and can be misleading depending on which weights the data producer has used.

J.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 is determined by either of the two equations,

$ADQR = MAX(v_i , in = 1...n)$ or $ADQR = MIN(v_i , in = 1...n)$ 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.

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 3951-1:2005, *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 6709:2008, *Standard representation of geographic point location by coordinates*

[5]   ISO 8601:2004, *Data elements and interchange formats — Information interchange — Representation of dates and times*

[6]   ISO 10303-227:2005, *Industrial automation systems and integration — Product data representation and exchange — Part 227: Application protocol: Plant spatial configuration*

[7]   ISO 19101:2002, *Geographic information — Reference model*

[8]   ISO 19105:2000, *Geographic information — Conformance and testing*

[9]   ISO 19107:2003, *Geographic information — Spatial schema*

[10]   ISO 19109:2005, *Geographic information — Rules for application schemas*

[11]   ISO 19123:2005, *Geographic information — Schema for coverage geometry and functions*

[12]   ISO 19131:2007, *Geographic information — Data product specifications*

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

[14]   ISO/IEC 11179-3:2013, *Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes*

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

[16]   ISO/TS 19139:2007 *Geographic information — Metadata — XML schema implementation*

[17]   *CRC Handbook of Tables for Probability and Statistics*, 2nd edition, 1982

[18]   Department of Defense (US). *Standard Practice: Mapping, Charting and Geodesy Accuracy.* MIL STD 600001, 1990

[19]   NATO STANAG 2215 IGEO, *Evaluation of land maps, aeronautical charts and digital topographic data,* 6th edition

[20]   Wikipedia: *http://en.wikipedia.org/wiki/Hypergeometric_distribution*