

DIN EN ISO 19157

ICS 35.240.70

Entwurf

Einsprüche bis 2011-10-29
Vorgesehen als Ersatz für
DIN EN ISO 19113:2005-05,
DIN EN ISO 19114:2005-05 und
DIN EN ISO 19114
Berichtigung 1:2006-10

**Geoinformation –
Datenqualität (ISO/DIS 19157:2011);
Englische Fassung prEN ISO 19157:2011**

Geographic information –
Data quality (ISO/DIS 19157:2011);
English version prEN ISO 19157:2011

Information géographique –
Qualité des données (ISO/DIS 19157:2011);
Version anglaise prEN ISO 19157:2011

Anwendungswarnvermerk

Dieser Norm-Entwurf mit Erscheinungsdatum 2011-08-22 wird der Öffentlichkeit zur Prüfung und Stellungnahme vorgelegt.

Weil die beabsichtigte Norm von der vorliegenden Fassung abweichen kann, ist die Anwendung dieses Entwurfes besonders zu vereinbaren.

Stellungnahmen werden erbeten

- vorzugsweise als Datei per E-Mail an nabau@din.de in Form einer Tabelle. Die Vorlage dieser Tabelle kann im Internet unter www.din.de/stellungnahme oder für Stellungnahmen zu Norm-Entwürfen der DKE unter www.dke.de/stellungnahme abgerufen werden;
- oder online im Norm-Entwurfs-Portal des DIN unter www.entwuerfe.din.de, sofern dort wiedergegeben;
- oder in Papierform an den Normenausschuss Bauwesen (NABau) im DIN, 10772 Berlin (Hausanschrift: Burggrafenstr. 6, 10787 Berlin).

Die Empfänger dieses Norm-Entwurfs werden gebeten, mit ihren Kommentaren jegliche relevanten Patentrechte, die sie kennen, mitzuteilen und unterstützende Dokumentationen zur Verfügung zu stellen.

Gesamtumfang 173 Seiten

Normenausschuss Bauwesen (NABau) im DIN

Nationales Vorwort

Dieses Dokument (prEN ISO 19157:2011) wurde vom Technischen Komitee ISO/TC 211 „Geographic information/Geomatics“ in Zusammenarbeit mit dem Technischen Komitee CEN/TC 287 „Geoinformation“, dessen Sekretariat von BSI (Vereinigtes Königreich) gehalten wird, erarbeitet.

Der für die deutsche Mitarbeit zuständige Arbeitsausschuss im DIN Deutsches Institut für Normung e. V. ist der als Spiegelausschuss zum CEN/TC 287 und ISO/TC 211 eingesetzte Arbeitsausschuss NA 005-03-03 AA „Kartographie und Geoinformation“ des Normenausschusses Bauwesen (NABau).

Dieses Dokument ist derzeit zur Umfrage vorgelegt.

Das Ausgabedatum des Europäischen Norm-Entwurfs stand zum Zeitpunkt der Veröffentlichung dieses DIN EN ISO-Entwurfs noch nicht fest; der prEN ISO wird jedoch vom CCMC unter der angegebenen prEN-Nummer demnächst zur CEN-Umfrage verteilt. Um der deutschen Öffentlichkeit eine möglichst lange Einspruchsfrist einzuräumen, wurde dieser Norm-Entwurf bereits vorab veröffentlicht.

Das Präsidium des DIN hat mit seinem Beschluss 1/2004 festgelegt, dass von dem in den Regeln der europäischen Normungsarbeit von CEN/CENELEC verankerten Grundsatz, wonach Europäische Normen in den drei offiziellen Sprachen Deutsch, Englisch und Französisch veröffentlicht werden, in begründeten Ausnahmefällen abgewichen und auf die deutsche Sprachfassung verzichtet werden kann.

Die Genehmigung dafür hat die DIN-Geschäftsleitung entsprechend den festgelegten Kriterien für die vorliegende Norm auf Antrag des Normenausschusses Bauwesen (NABau), Fachbereich 03 „Vermessungswesen; Geoinformation“ erteilt.

Änderungen

Gegenüber DIN EN ISO 19113:2005-05, DIN EN ISO 19114:2005-05 und DIN EN ISO 19114 Berichtigung 1:2006-10 wurden folgende Änderungen vorgenommen:

- a) Inhalte der genannten Normen zusammengefasst.

Nationaler Anhang NA (informativ)

Begriffe

4.1 Begriffe

Für die Anwendung dieses Dokuments gelten die folgenden Begriffe.

4.1

Genauigkeit

Maß der Übereinstimmung zwischen einem Ermittlungsergebnis oder einem Messergebnis und dem wahren Wert

[ISO 6709:2008, Definition 4.1]

ANMERKUNG In dieser Internationalen Norm kann der wahre Wert ein Bezugswert sein, der als wahr angenommen wird.

4.2

Katalog

Sammlung von **Elementen** (4.16) oder ein elektronisches Dokument oder Papierdokument, das Informationen über die Sammlung von Elementen enthält

[ISO 10303-227:2005, Definition 3.3.10]

4.3

Konformität

Erfüllung festgelegter Anforderungen

[ISO 19105:2000, Definition 3.8]

4.4

Qualitätsgrenzlage

Schwellenwert oder Reihe von Schwellenwerten für die Ergebnisse zur **Qualität** (4.19) von Daten, mit denen bestimmt wird, wie gut ein **Datenbestand** (4.8) die in dessen **Datenprodukt-Spezifikation** (4.6) oder den Benutzeranforderungen festgelegten Kriterien erfüllt

4.5

Richtigkeit

Übereinstimmung mit dem **Gegenstandsbereich** (4.22)

4.6

Datenprodukt-Spezifikation

detaillierte Beschreibung eines **Datenbestands** (4.8) oder einer **Reihe von Datensätzen** (4.9) verbunden mit ergänzenden Informationen, die deren Erstellung, Lieferung an und Nutzung durch Dritte ermöglicht

[ISO 19131:2007, Definition 4.7]

4.7

Basismaß der Datenqualität

allgemeines Maß der **Datenqualität** (4.19), das als Grundlage für die Erstellung spezieller Maße der Datenqualität verwendet wird

ANMERKUNG Basismaße der Datenqualität sind abstrakte Datentypen. Sie können nicht unmittelbar zur Angabe der Datenqualität verwendet werden.

4.8

Datenbestand

identifizierbare Sammlung von Daten

[ISO 19115:2003, Definition 4.2]

ANMERKUNG Ein Datenbestand kann eine kleinere Gruppierung von Daten sein, die sich, trotz gewisser Einschränkungen wie beispielsweise der räumlichen Ausdehnung oder der **Featureklasse** (4.11), räumlich innerhalb eines größeren Datenbestands befindet. Theoretisch kann ein Datenbestand so klein sein wie ein einzelnes Feature oder **Featureattribut** (4.12), das in einem größeren Datenbestand enthalten ist.

4.9

Reihe von Datensätzen

Sammlung von **Datenbeständen** (4.8) mit derselben Produktspezifikation

[ISO 19115:2003, Definition 4.3]

4.10

direkte Ermittlungsmethode

Verfahren zur Bewertung der **Qualität** (4.19) eines **Datenbestands** (4.8), das auf der Überprüfung der **Elemente** (4.16) innerhalb des Datenbestands beruht

4.11

Feature

Abstraktion eines Phänomens der realen Welt

[ISO 19101:2002, Definition 4.11]

ANMERKUNG Ein Feature kann als Klasse oder Instanz vorkommen. Featureklasse oder Featureinstanz sollten verwendet werden, wenn nur eines von beiden gemeint ist.

4.12

Featureattribut

Merkmal eines **Features** (4.11)

[ISO 19101:2002, Definition 4.12]

ANMERKUNG Ein Featureattribut hat einen Namen, einen Datentyp und einen mit ihm verbundenen Wertebereich. Ein Featureattribut für eine Featureinstanz hat außerdem einen dem Wertebereich entnommenen Attributwert.

4.13

Featureoperation

Operation, die jede Instanz einer **Featureklasse** (4.11) ausführen kann

[ISO 19110:2005, Definition 4.5]

4.14

Geodaten

Daten mit einem impliziten oder expliziten Bezug zu einem Ort auf der Erde

[ISO 19110:2005, Definition 4.12]

4.15

indirekte Ermittlungsmethode

Verfahren zur Bewertung der **Qualität** (4.19) eines **Datenbestands** (4.8), das auf externem Wissen beruht

ANMERKUNG Beispiele für externes Wissen sind der Ursprung des Datenbestands, wie beispielsweise Produktionsverfahren oder Quelldaten.

4.16

Element

Einheit

das, was einzeln beschrieben und betrachtet werden kann

[ISO 2859-5:2005, Definition 3.4]

ANMERKUNG Ein Element kann jeglicher Teil eines **Datenbestands** (4.8) sein, zum Beispiel ein **Feature** (4.11), eine Featurebeziehung, ein **Featureattribut** (4.12) oder eine Kombination aus diesen Teilen.

4.17

Metadaten

Daten über Daten

[ISO 19115:2003, Definition 4.5]

4.18

Metaqualität

Informationen, die die **Qualität** (4.19) der Datenqualität beschreiben

4.19

Qualität

Grad, in dem ein Satz inhärenter Merkmale Anforderungen erfüllt

[ISO 9000:2005, Definition 3.1.1]

4.20

Register

Satz von Dateien, die Identifikatoren enthalten, die **Elementen** (4.16) zugewiesen sind, mit Beschreibungen der zugehörigen Elemente

[ISO 19135:2005, Definition 4.1.9]

4.21

unabhängiger Qualitätsbericht

Freitext-Dokument, dass ausführliche Informationen über Bewertungen, Ergebnisse und Maßnahmen in Bezug auf die Daten**qualität** (4.19) zur Verfügung stellt

4.22

Gegenstandsbereich

Diskursuniversum

Darstellung der realen oder hypothetischen Welt, die alles enthält, was von Belang ist

[ISO 19101:2002, Definition 4.29]

— Leerseite —

CEN/TC 287

Date: 2011-08

prEN ISO 19157:2011

CEN/TC 287

Secretariat: BSI

Geographic information — Data quality (ISO 19157:2011)

Geoinformation — Datenqualität (ISO 19157:2011)

Information géographique — Qualité des données (ISO 19157:2011)

ICS:

Descriptors:

Document type: European Standard
Document subtype:
Document stage: CEN Enquiry
Document language: E

Contents

Page

Foreword.....	6
Introduction	7
1 Scope	8
2 Conformance.....	8
3 Normative references	8
4 Terms and definitions	9
5 Abbreviated terms	11
5.1 Abbreviations	11
5.2 Package abbreviations	11
6 Overview of data quality	12
7 Components of data quality	13
7.1 Overview of the components.....	13
7.2 Data quality unit.....	14
7.3 Data quality scope	15
7.4 Data quality elements	16
7.4.1 General.....	16
7.4.2 Completeness	17
7.4.3 Logical consistency	17
7.4.4 Positional accuracy	18
7.4.5 Thematic accuracy.....	18
7.4.6 Temporal quality	18
7.4.7 Usability element	18
7.5 Descriptors of data quality elements.....	19
7.5.1 General.....	19
7.5.2 Measure	19
7.5.3 Evaluation method.....	20
7.5.4 Result	20
7.6 Metaquality elements	22
7.7 Descriptors of a metaquality element.....	23
8 Data quality measures.....	24
8.1 General.....	24
8.2 Standardised data quality measures	24
8.3 User defined data quality measures	24
8.4 Catalogue of data quality measures	24
8.5 List of components.....	25
8.6 Component details.....	26
8.6.1 Measure identifier	26
8.6.2 Name	26
8.6.3 Alias	26
8.6.4 Element name.....	26
8.6.5 Basic measure.....	26
8.6.6 Definition	26
8.6.7 Description	26
8.6.8 Parameter	27
8.6.9 Value type	27
8.6.10 Value structure.....	27
8.6.11 Source reference.....	27
8.6.12 Example	27

9	Data quality evaluation	27
9.1	The process for evaluating data quality.....	27
9.1.1	Introduction.....	27
9.1.2	The process flow	27
9.1.3	Process steps	28
9.2	Data quality evaluation methods	29
9.2.1	Classification of data quality evaluation methods.....	29
9.2.2	Direct evaluation.....	30
9.2.3	Indirect evaluation	30
9.3	Aggregation and derivation.....	31
10	Data quality reporting	31
10.1	General	31
10.2	Particular cases	32
10.2.1	Reporting Aggregation (aggregated results).....	32
10.2.2	Reporting Derivation (derived results).....	32
10.2.3	Reference to the original data quality result	33
Annex A	(normative) Abstract test suites	34
A.1	Test case identifier: Quality evaluation process.....	34
A.2	Test case identifier: Data quality metadata	34
A.3	Test case identifier: Metadata conformity	34
A.4	Test case identifier: Standalone quality report	35
A.5	Test case identifier: Data quality measures	35
Annex B	(informative) Data quality concepts and their use	36
B.1	Framework of data quality concepts	36
B.2	The structure of datasets and components for quality description	37
B.3	When to use quality evaluation procedures	38
B.4	Reporting quality information	39
B.4.1	Why report data quality.....	39
B.4.2	When to report quality information	39
B.4.3	How to report quality information.....	40
Annex C	(normative) Data dictionary for data quality.....	42
C.1	Data dictionary overview	42
C.1.1	Introduction.....	42
C.1.2	Name/role name	42
C.1.3	Definition	42
C.1.4	Obligation/Condition	42
C.1.5	Maximum occurrence.....	43
C.1.6	Data type.....	43
C.1.7	Domain.....	43
C.2	Quality package data dictionaries	44
C.2.1	Data quality information	44
C.2.2	Measures information	52
C.3	CodeLists and enumerations	56
C.3.1	Introduction.....	56
C.3.2	DQ_EvaluationMethodTypeCode <<CodeList>>.....	56
C.3.3	DQM_ValueStructure <<CodeList>>.....	56
Annex D	(normative) List of standardised data quality measures	57
D.1	Introduction.....	57
D.2	Completeness	57
D.2.1	Commission	57
D.2.2	Omission	60
D.3	Logical consistency	61
D.3.1	Conceptual consistency	61
D.3.2	Domain consistency.....	66
D.3.3	Format consistency.....	68
D.3.4	Topological consistency.....	69

D.4	Positional accuracy	76
D.4.1	Absolute or external accuracy	76
D.4.2	Gridded data position accuracy	99
D.5	Temporal quality	100
D.5.1	Accuracy of a time measurement	100
D.5.2	Temporal consistency	103
D.5.3	Temporal validity	103
D.6	Thematic accuracy.....	103
D.6.1	Classification correctness	103
D.6.2	Non-quantitative attribute correctness	108
D.6.3	Quantitative attribute accuracy	109
D.7	Aggregation Measures	112
Annex E	(informative) Evaluating and reporting data quality.....	115
E.1	Introduction	115
E.2	Dataset description.....	115
E.2.1	Data product specification.....	115
E.2.2	Representation of the real world, the universe of discourse and the dataset	116
E.3	Quality evaluation process	119
E.3.1	Specify data quality unit(s)	119
E.3.2	Specify data quality measures	119
E.3.3	Specify data quality evaluation procedures	119
E.3.4	Determine the output of the data quality evaluation (Result)	120
E.4	Reporting data quality	126
E.4.1	Reporting as metadata	126
E.4.2	Reporting in a standalone quality report	134
E.5	Additional examples	134
E.5.1	General.....	134
E.5.2	Reporting descriptive results as metadata	135
E.5.3	Reporting metaquality as metadata	135
E.5.4	How to report sampling procedure	137
Annex F	(informative) Sampling methods for evaluating.....	139
F.1	Introduction	139
F.2	Lot and item.....	139
F.3	Sample size	139
F.4	Sampling strategies.....	140
F.4.1	Introduction	140
F.4.2	Probabilistic versus judgemental sampling	141
F.4.3	Feature-guided versus area-guided sampling	141
F.5	Probability-based sampling	143
F.5.1	General considerations	143
F.5.2	Existing standard for inspection by sampling.....	143
F.5.3	Sampling process	146
Annex G	(normative) Data quality basic measures	147
G.1	Purpose of data quality basic measures.....	147
G.2	Counting-related data quality basic measures.....	147
G.3	Uncertainty-related data quality basic measures.....	148
G.3.1	General.....	148
G.3.2	One-dimensional random variable, Z	148
G.3.3	Two-dimensional random variable X and Y	150
G.3.4	Three-dimensional random variable X , Y , Z	151
Annex H	(informative) Management of data quality measures.....	152
H.1	Introduction	152
H.2	Storage of data quality measures	152
H.2.1	General.....	152
H.2.2	Catalogue of data quality measures	153
H.2.3	Register of data quality measures	153

Annex I (informative) Guidelines for the use of Quality Elements	156
I.1 Overview	156
I.2 Data quality element categories	156
I.2.1 General	156
I.2.2 Ordering in data quality evaluation	156
I.3 The relationships between the data quality elements	158
I.3.1 General	158
I.3.2 Data quality elements related to missing attribute values	158
I.3.3 Relationships between the different aspects of accuracy	158
I.3.4 Dependency between completeness and accuracy	159
I.4 Data quality elements – example of use	160
I.4.1 Completeness	160
I.4.2 Logical consistency	160
I.4.3 Positional accuracy	162
I.4.4 Temporal quality	162
I.4.5 Thematic accuracy	163
I.5 Discussions on difficult cases	163
I.5.1 Relation between misclassification and completeness at feature type level	163
I.5.2 Quality elements related to unique identifiers	164
Annex J (informative) Aggregation of data quality results	165
J.1 Introduction	165
J.2 100% pass/fail	165
J.3 Weighted pass/fail	165
J.4 Maximum/minimum value	166
Bibliography	167

Foreword

This document (prEN ISO 19157:2011) has been prepared by Technical Committee ISO/TC 211 "Geographic information/Geomatics" in collaboration with Technical Committee CEN/TC 287 "Geographic Information" the secretariat of which is held by BSI.

This document is currently submitted to the parallel Enquiry.

Endorsement notice

The text of ISO/DIS 19157:2011 has been approved by CEN as an prEN ISO 19157:2011 without any modification.

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 standardises 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 may 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 may 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.

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 19107:2003, *Geographic information — Spatial schema*

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

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

ISO 19115:2003, *Geographic information — Metadata*

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

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

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

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

4 Terms and definitions

4.1

accuracy

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

[ISO 6709:2008, definition 4.1]

NOTE In this International standard, the true value may be a reference value that is accepted as true

4.2

catalogue

collection of **items** (4.16) or an electronic or paper document that contains information about the collection of items

[ISO 10303-227:2005, definition 3.3.10]

4.3

conformance

fulfilment of specified requirements

[ISO 19105:2000, definition 3.8]

4.4

conformance quality level

threshold value or set of threshold values for data **quality** (4.19) 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** (4.22)

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

[ISO 19131:2007, definition 4.7]

4.7

data quality basic measure

generic data **quality** (4.19) measure used as a basis for the creation of specific data quality measures

NOTE Data quality basic measures are abstract data types. They cannot be used directly when reporting data quality.

4.8

dataset

identifiable collection of data

[ISO 19115:2003, definition 4.2]

NOTE A dataset may be a smaller grouping of data which, though limited by some constraint such as spatial extent or **feature** (4.11) type, is located physically within a larger dataset. Theoretically, a dataset may be as small as a single feature or **feature attribute** (4.12) contained within a larger dataset.

4.9

dataset series

collection of **datasets** (4.8) sharing the same product specification

[ISO 19115:2003, definition 4.3]

4.10

direct evaluation method

method of evaluating the **quality** (4.19) of a **dataset** (4.8) based on inspection of the **items** (4.16) within the dataset

4.11

feature

abstraction of real world phenomena

[ISO 19101:2002, definition 4.11]

NOTE A feature may occur as a type or an instance. Feature type or feature instance should be used when only one is meant.

4.12

feature attribute

characteristic of a **feature** (4.11)

[ISO 19101:2002, definition 4.12]

NOTE A feature attribute has a name, a data type and a value domain associated with it. A feature attribute for a feature instance also has an attribute value taken from the value domain.

4.13

feature operation

operation that every instance of a **feature** (4.11) type may perform

[ISO 19110:2005, definition 4.5]

4.14

geographic data

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

[ISO 19109:2005, definition 4.12]

4.15

indirect evaluation method

method of evaluating the **quality** (4.19) of a **dataset** (4.8) based on external knowledge

NOTE Examples of external knowledge are dataset lineage, such as production method or source data.

4.16

item

anything that can be described and considered separately

[ISO 2859-5:2005, definition 3.4]

NOTE 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.

4.17

metadata

data about data

[ISO 19115:2003, definition 4.5]

4.18

metaquality

information describing the **quality** (4.19) of data quality

4.19

quality

degree to which a set of inherent characteristics fulfils requirements

[ISO 9000:2005, definition 3.1.1]

4.20

register

set of files containing identifiers assigned to **items** (4.16) with descriptions of the associated items

[ISO 19135:2005, definition 4.1.9]

4.21

standalone quality report

free text document providing fully detailed information about data **quality** (4.19) evaluations, results and measures used

4.22

universe of discourse

view of the real or hypothetical world that includes everything of interest

[ISO 19101:2002, definition 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 19115:2003]
CT	Catalogues [ISO/TS 19139:2007]

DQ	Data Quality [ISO 19157]
DQM	Data Quality Measure [ISO 19157]
EX	Extent [ISO 19115:2003]
GF	General Feature [ISO 19109:2005]
MD	Metadata [ISO 19115:2003]
QE	Quality Extended [ISO 19115-2:2009]
RE	Registration [ISO 19135:2005]

6 Overview of data quality

This International Standard can be used for

- aiding 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. Data product specifications should be established in conformance with 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.

NOTE 2 The process of evaluating data quality is described in Clause 9.

NOTE 3 How to report data quality is described in Clause 10.

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 are provided by purpose, usage and lineage. This information is reported as metadata in conformance with ISO 19115:2003.

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 4 The concept of metaquality is described in 7.6.

Figure 1 provides an overview of data quality information.

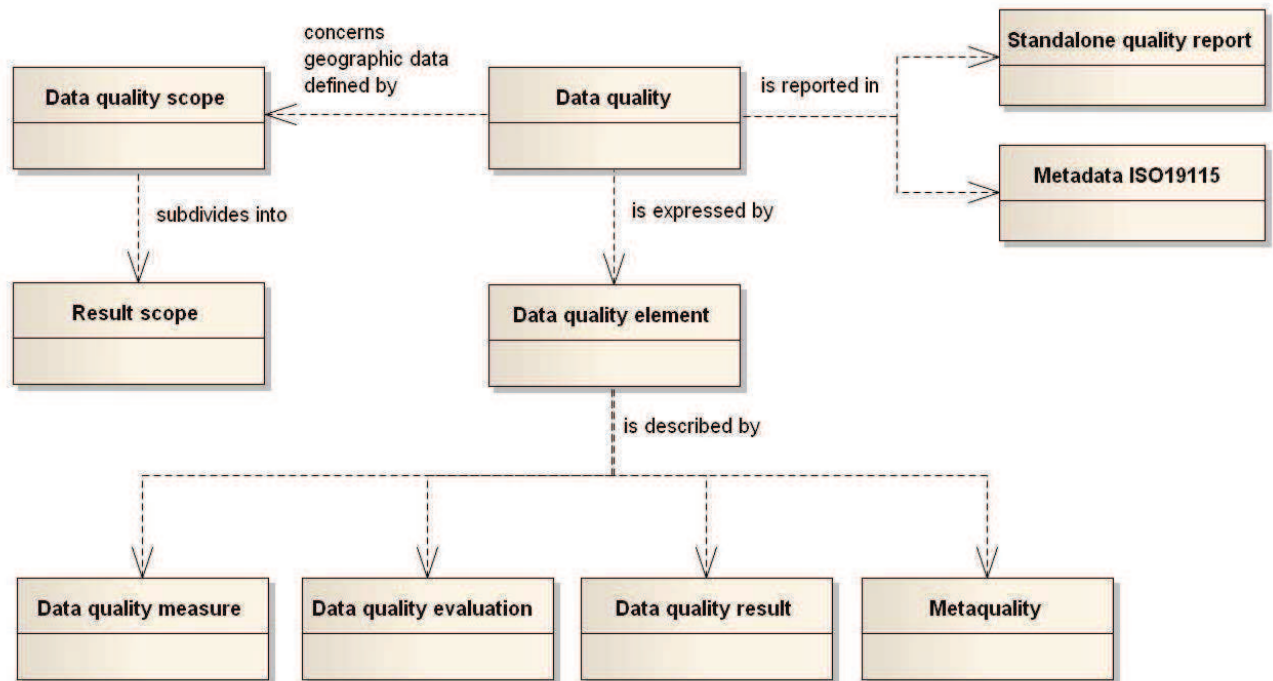


Figure 1 — Conceptual model on quality for geographic data

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. Each component is further described in the subsequent subclauses. See also the data dictionary defined in Annex C for more information about components and their attributes.

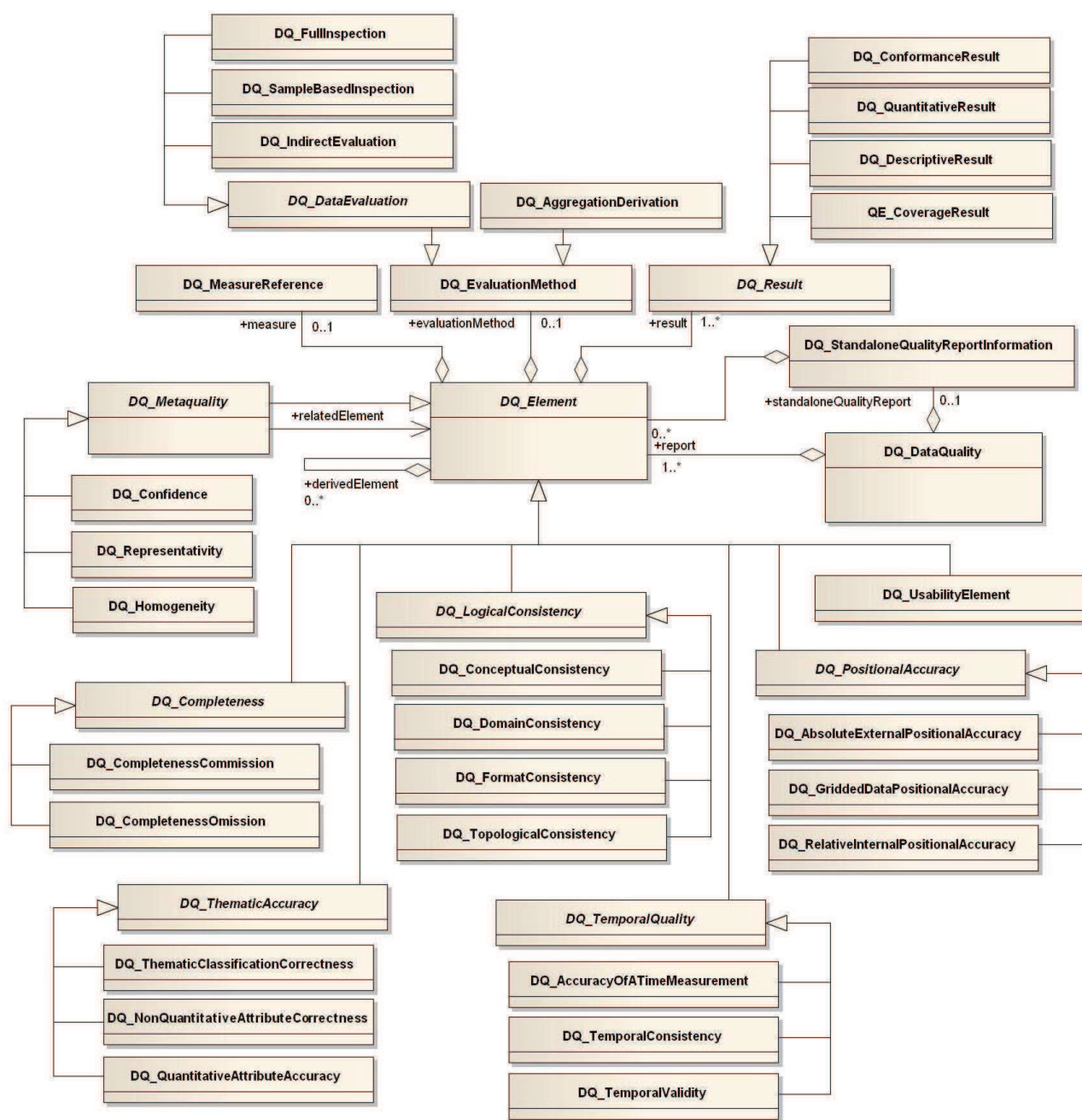


Figure 2 — Overview of the components of data quality

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.

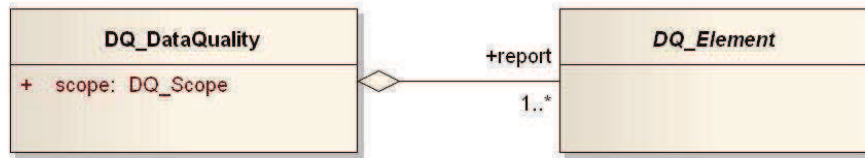


Figure 3 — Data quality unit

7.3 Data quality scope

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 Figure 4):

- 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).

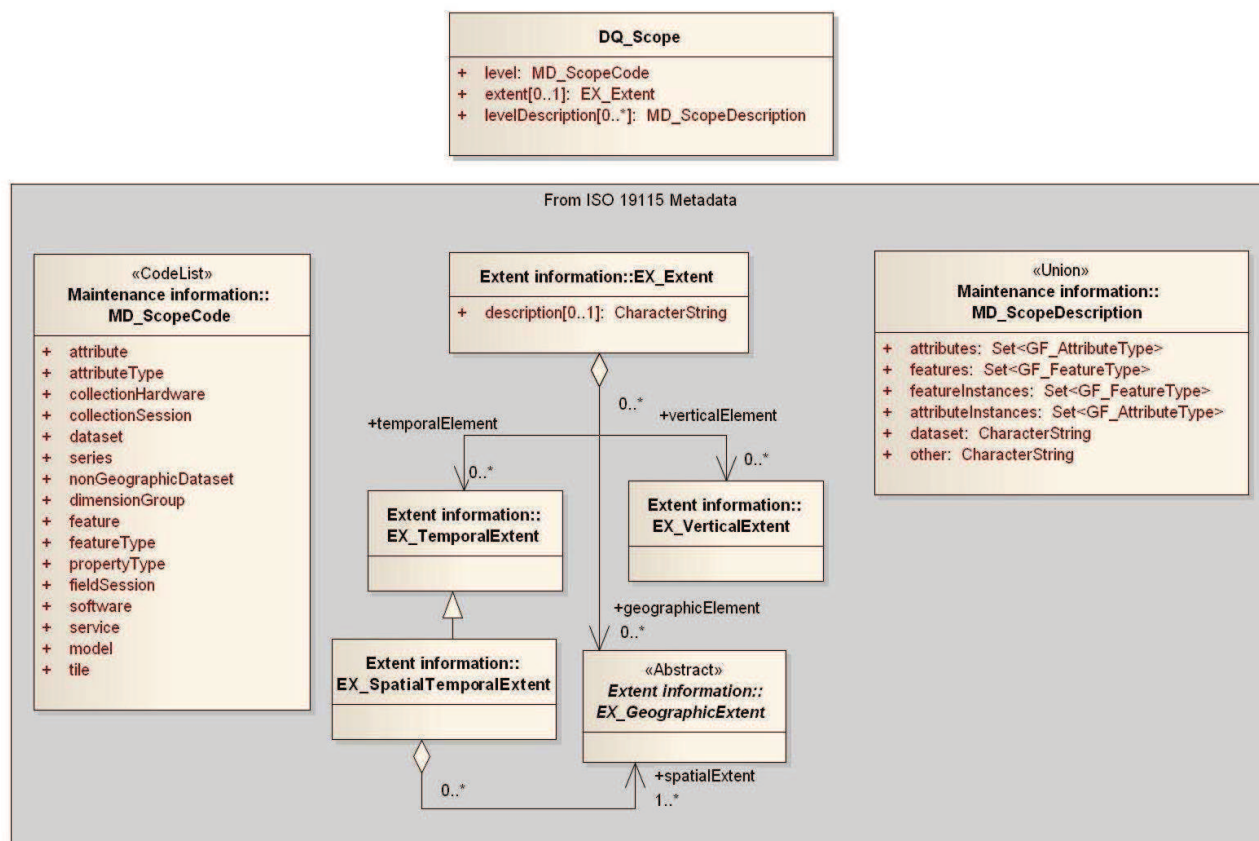


Figure 4 — Data quality scope

7.4 Data quality elements

7.4.1 General

A data quality element is a component describing a certain aspect of the quality of geographic data and these have been organised into different categories. These categories are shown in Figure 5.

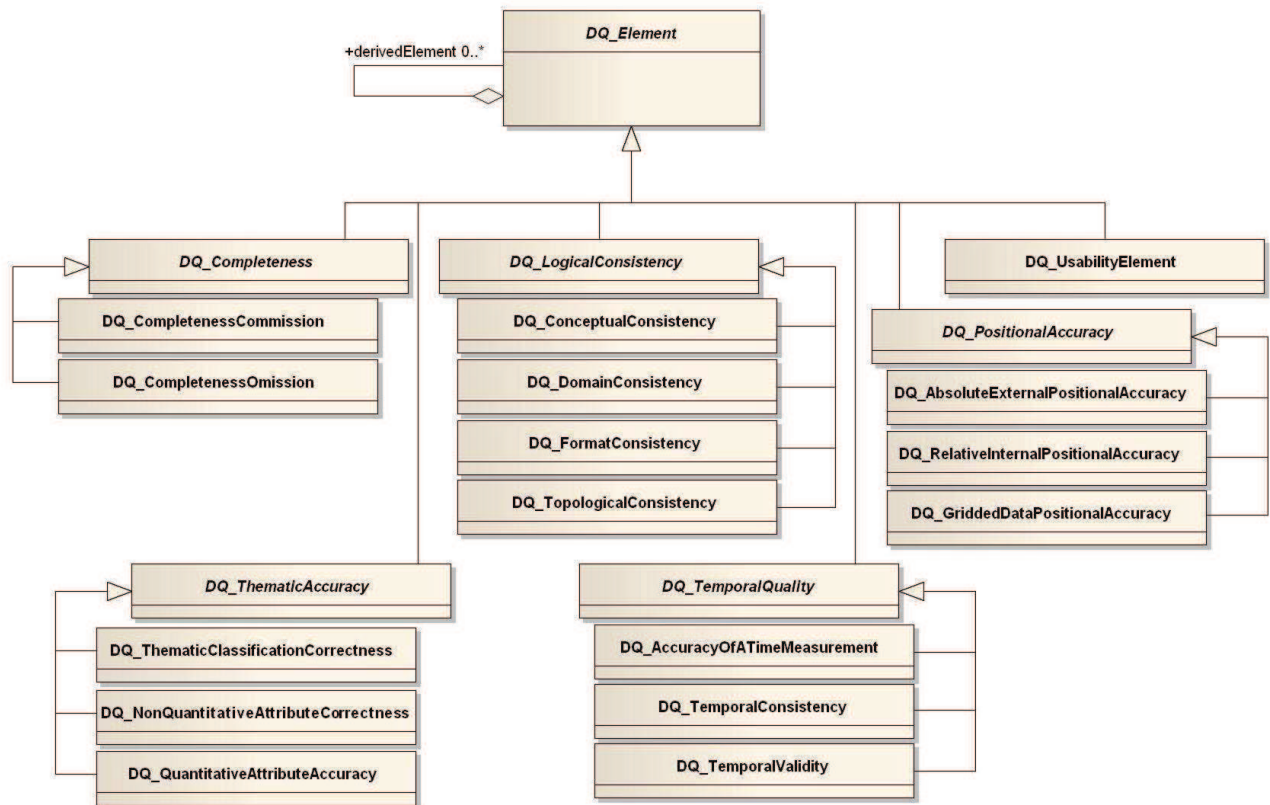


Figure 5 — Overview of the data quality elements

7.4.2 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.

7.4.3 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.

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

7.4.5 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.

7.4.6 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 may be either a defined point in time or a period.

NOTE 2 March 33 is an example of invalid data.

7.4.7 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 7.5) 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 may be used to declare the conformance of the dataset to a particular specification.

7.5 Descriptors of data quality elements

7.5.1 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 6, and are described in 7.5.2, 7.5.3 and 7.5.4.

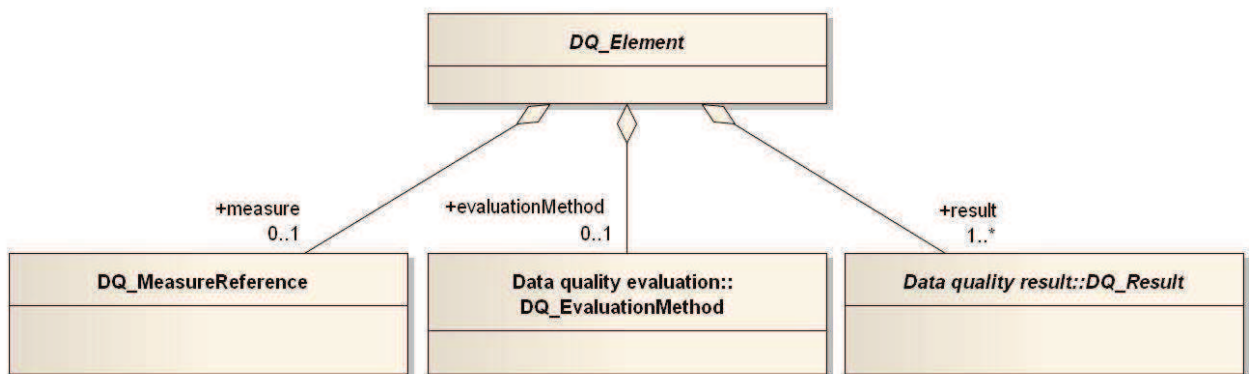


Figure 6 — Data quality element descriptors

7.5.2 Measure

A data quality element should refer to one measure only, by means of a measure reference (see Figure 7), 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.

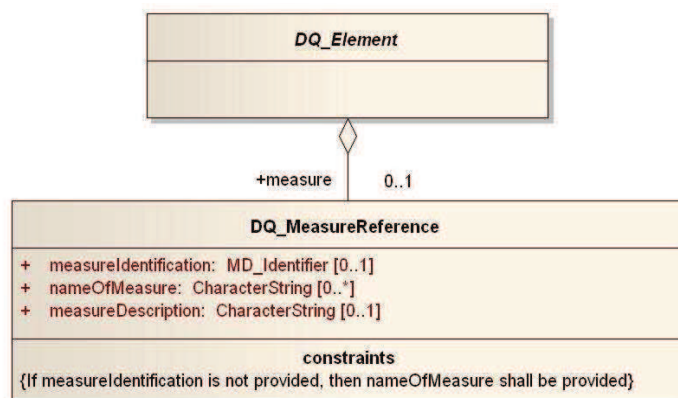


Figure 7 — Data quality measure reference

Data quality measures are further described in Clause 8 of this standard and Annex D contains a list of standardised 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 comport one instance of DQ_Element for each measure applied.

7.5.3 Evaluation method

Data quality evaluation method describes those procedures and methods which are applied to the geographic data to arrive at a data quality result, see Figure 8. 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 in conformance with ISO 19108:2002. If the evaluation was carried out on non-consecutive dates, each single date should be included.

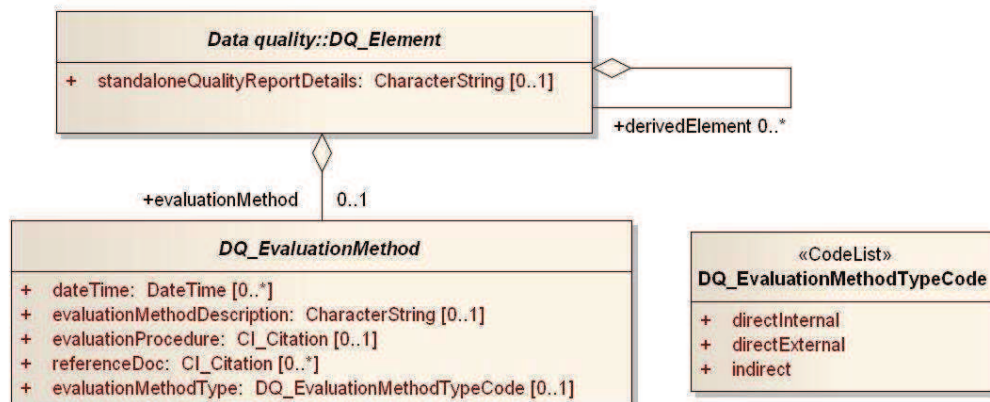


Figure 8 — Data quality evaluation method

7.5.4 Result

7.5.4.1 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 9.

NOTE 1 Different types of results may be provided for the same data quality element.

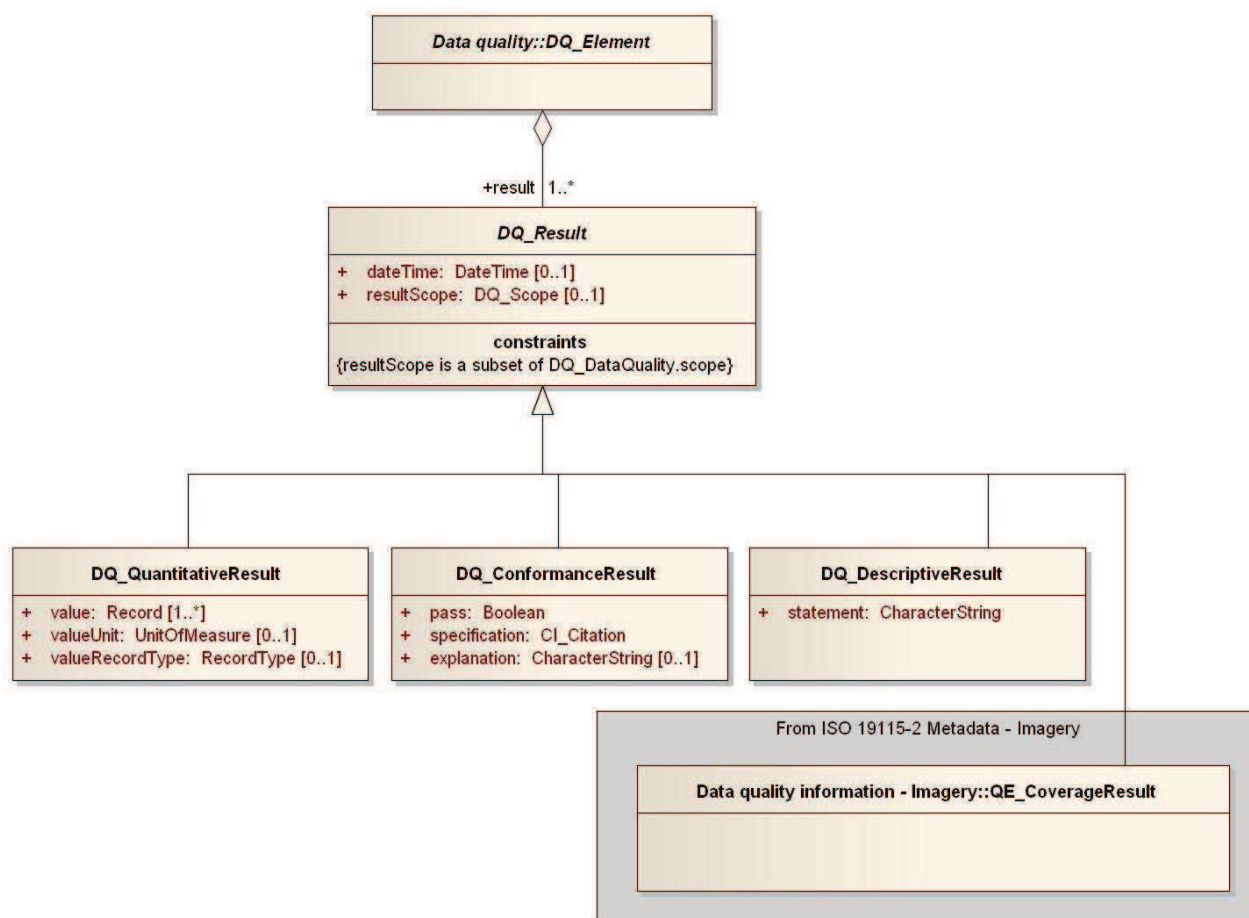


Figure 9 — Data quality result

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 7.3).

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).

7.5.4.2 Quantitative result

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

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

NOTE The attribute valueRecordType is of type RecordType, which is a generic data type defined in ISO/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 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 "%".

7.5.4.3 Conformance result

A conformance result is the outcome of comparing the value or set of values obtained from applying a measure to the data specified by a data quality scope with a specified 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 used conformance quality level 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.

7.5.4.4 Descriptive result

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

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

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

7.5.4.5 Coverage result

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

7.6 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 10:

— Confidence – trustworthiness of a data quality result.

NOTE 1 Quantitative figures for confidence may 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 may 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 input by different operators, depending on the acquisition zone or the acquisition date.

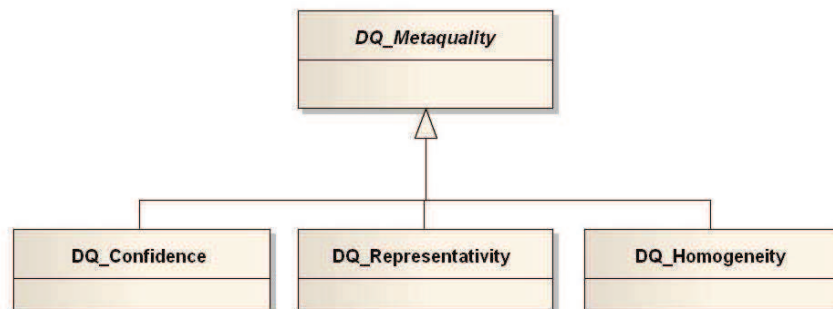


Figure 10 — Metaquality elements

7.7 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 7.5 and Figure 11). 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.

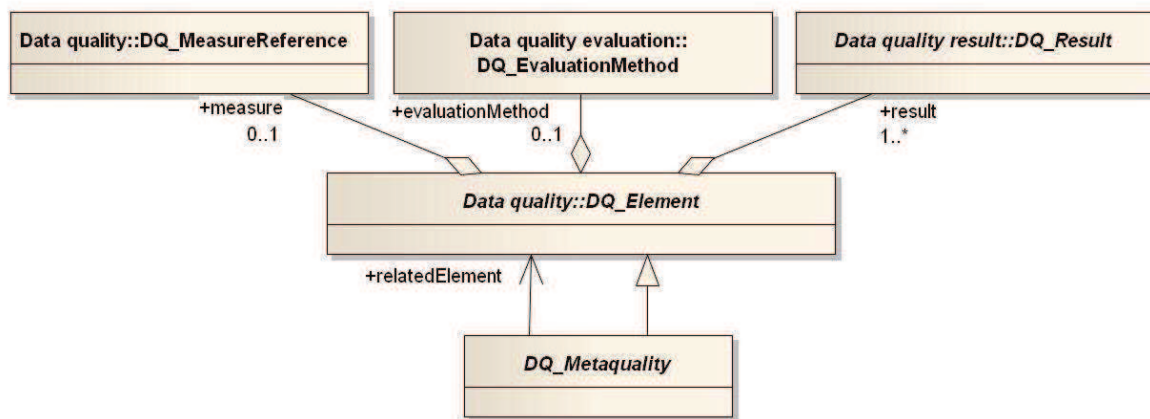


Figure 11 — Metaquality descriptors

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, standardised data quality measures should be used.

8.2 Standardised data quality measures

A list of standardised 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 standardised 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 standardised data quality measures cannot be complete. There may be cases where the user of this International Standard has to devise other data quality measures. These measures should 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 12 represents the components of data quality measures.

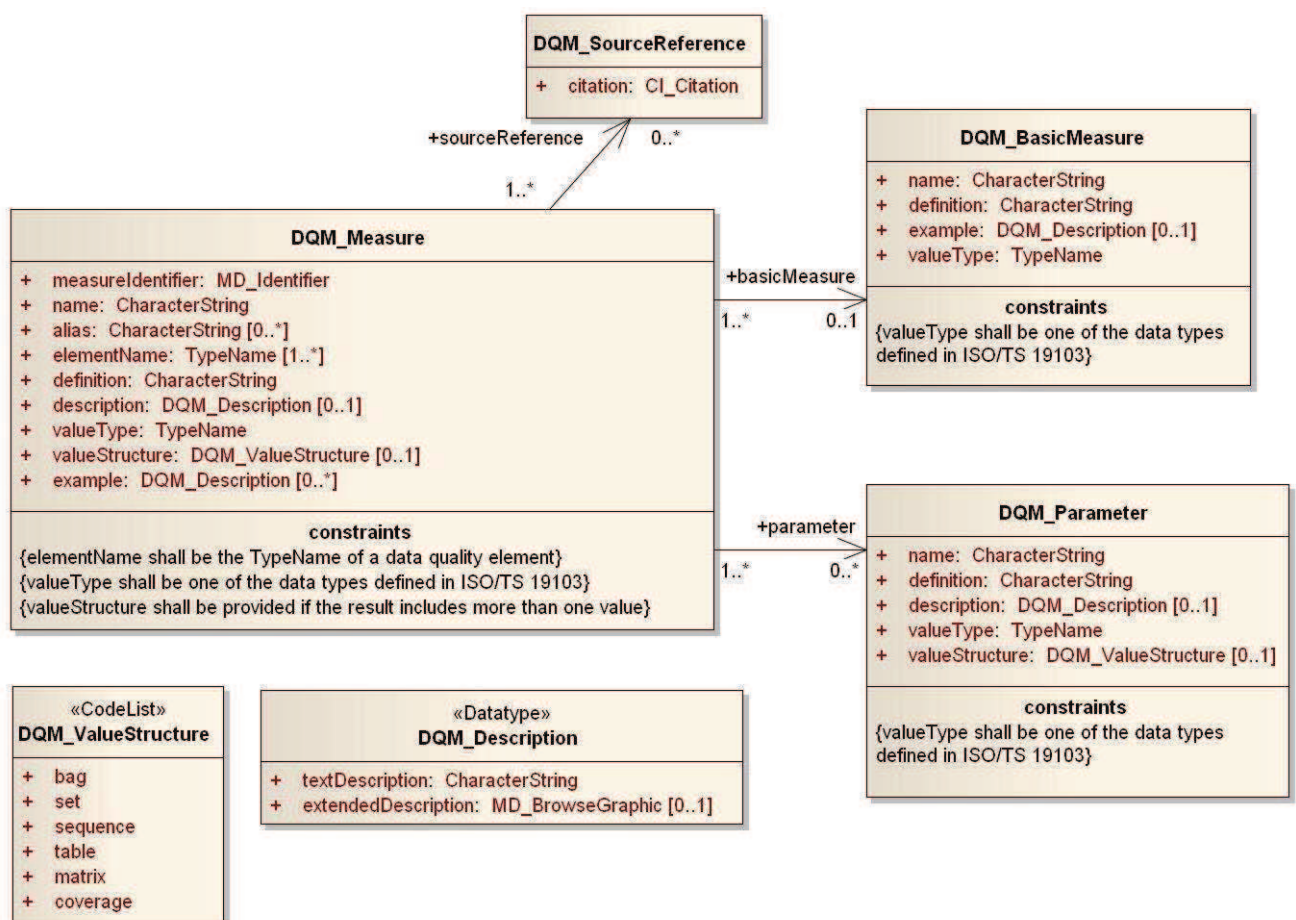


Figure 12 — Data quality measures

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 7.5.2)

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 7.4 and 7.6) 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 include 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 13 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.

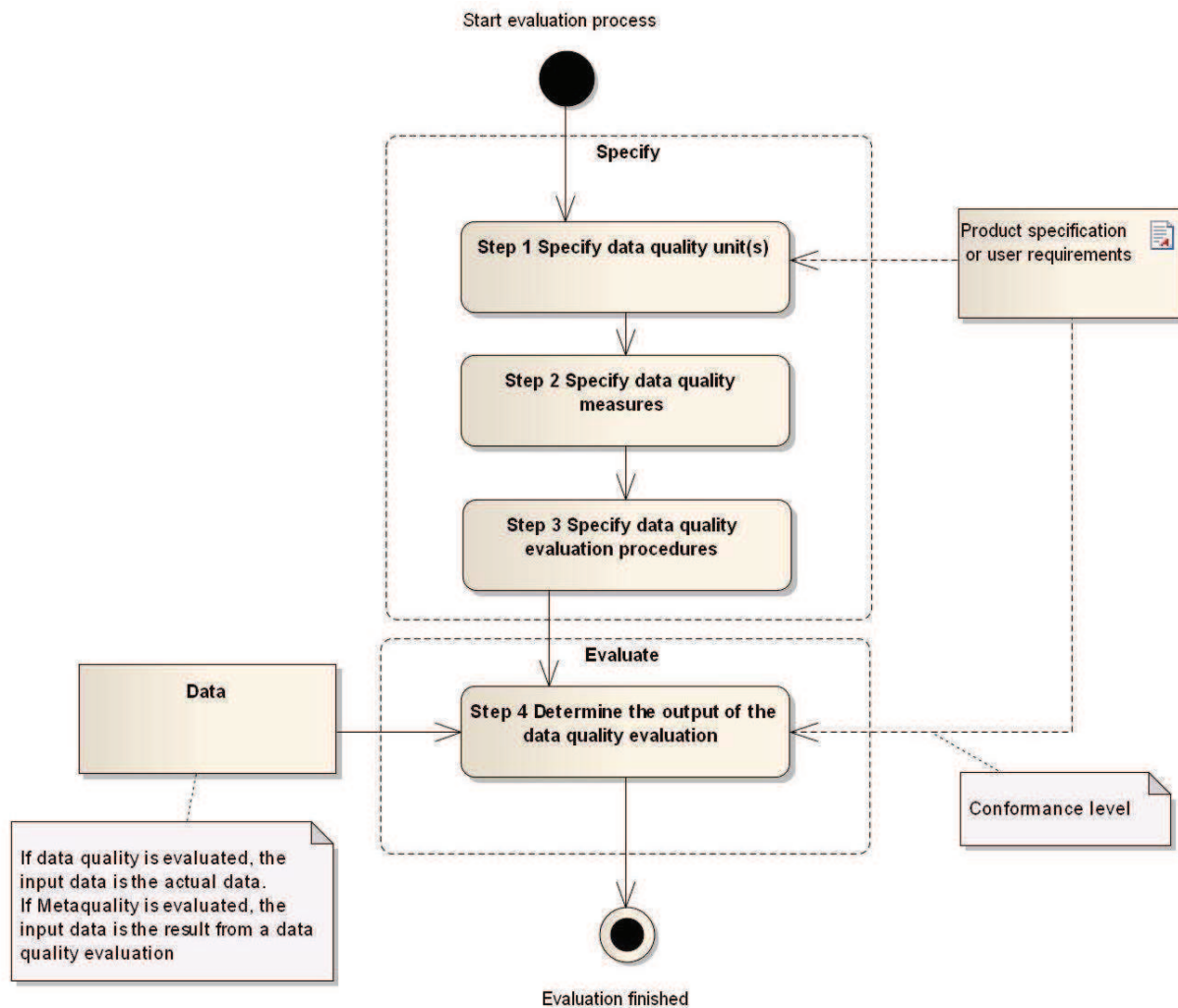


Figure 13 — Evaluating data quality

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 by a scope and quality element(s), see 7.2 and 7.3. 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 7.4, 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 14 shows the classes used describing the evaluation methods.

NOTE lineage is described in ISO 19115:2003

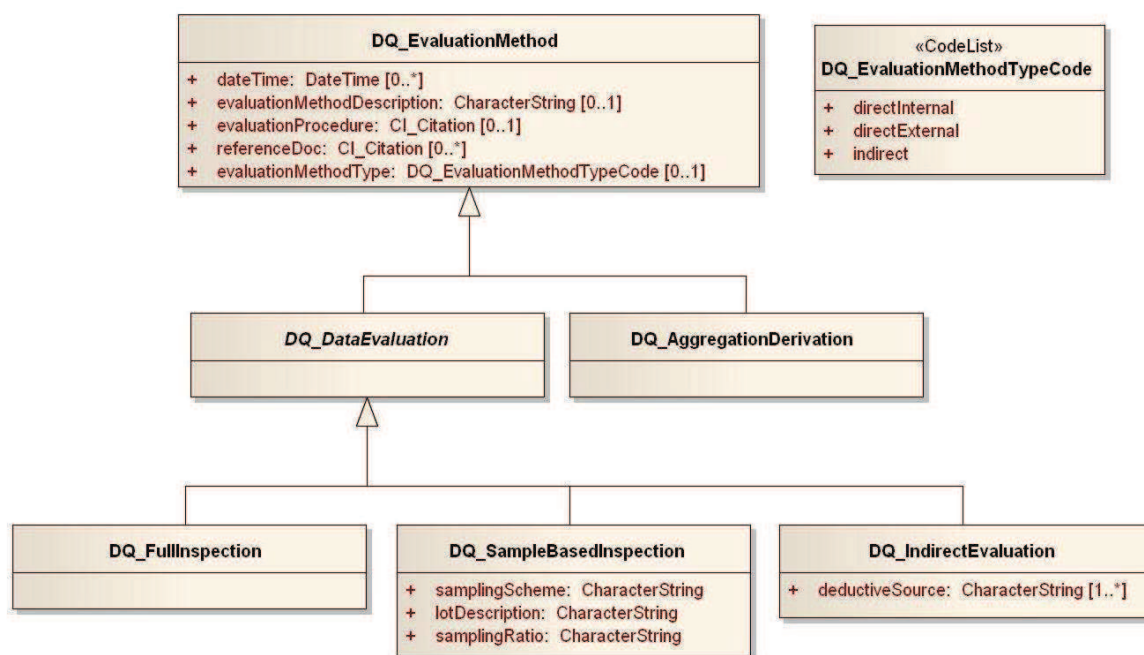


Figure 14 — Data quality evaluation methods

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 19115:2003) or other data quality reports on the dataset or data used to produce the dataset. It 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 7.5.4.4.

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, ISO 19115:2003 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 15).

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.

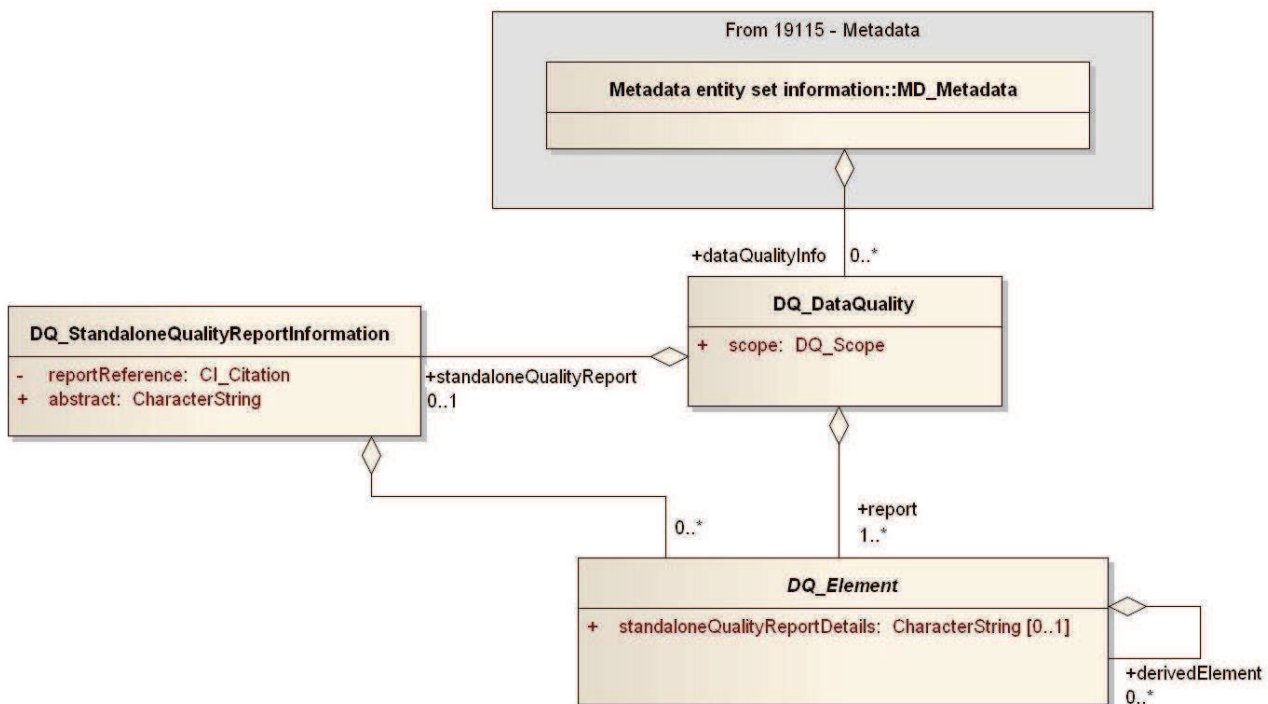


Figure 15 — Reporting data quality

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 have 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 19115:2003 and ISO 19115-2:2009.
- b) Test method: Check abstract test suites provided in ISO 19115:2003, D.2, D.2.1, D.2.2, D.2.4, D.2.5, D.2.6
- c) Reference: ISO 19115:2003, D.2, D.2.1, D.2.2, D.2.4, D.2.5, D.2.6
- 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 19115:2003). 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 provide 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 data product specification or user requirements used, if these are changed then quality evaluation should be repeated againsts changed specification or requirements. Care should be taken when comparing quality results where universe of discourse is different. Typical example of this is related to model transformation in Spatial Data Infrastructures or generalization. For example if geometry of a feature type is changed then positional accuracy results are changed as well.

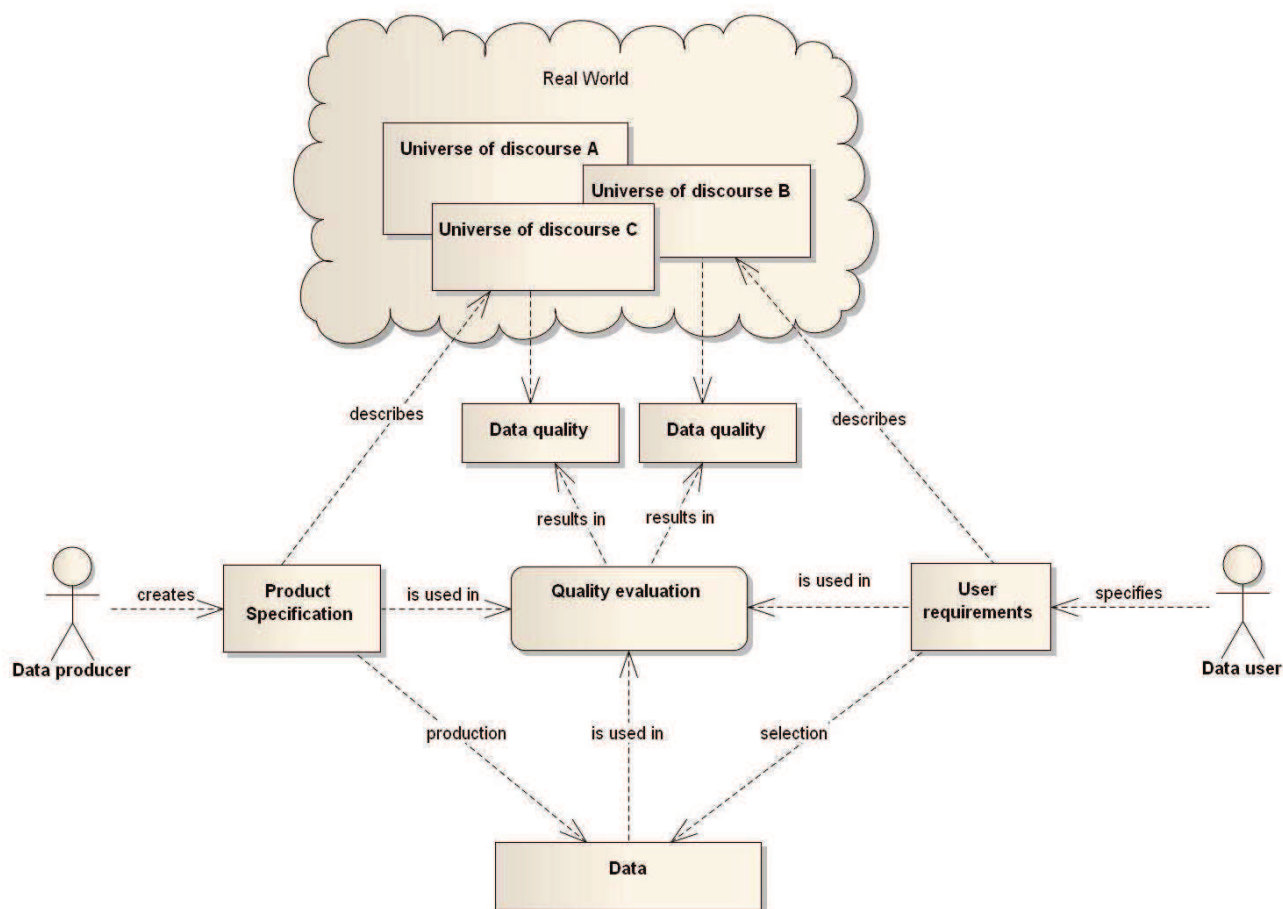


Figure B.1 — Framework of data quality concepts

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 may 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 recognises 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 recognises 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 ↓ Lower level	Series	
	Dataset	
	Subset	
	Feature type	Attribute type
	Feature instance	Attribute instance

NOTE Quality for an instance of feature, feature attribute or associations between features may 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 19115:2003, aggregates zero, one or several data quality units (instances of the class DQ_DataQuality, as specified in this International Standard), see Figure B.2.

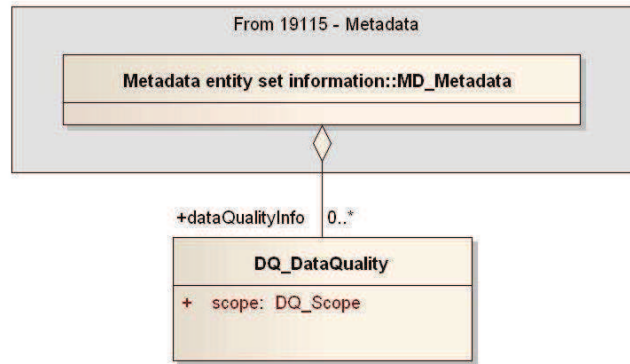


Figure B.2 — Data quality information

B.4.3.2.3 Reporting quality information within a standalone quality report

The standardisation 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 standardised so that each particular organisation 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 organisation 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 defined in Clauses 7, 8, 9 and 10. The dictionary is specified in a hierarchy to establish relationships and an organization for the information. The clause titles of several of the tables have been expanded to reflect class specification within the respective diagram. Each UML model class equates to a data dictionary entity. Each UML model class attribute equates to a data dictionary element. The shaded rows define entities. The entities and elements within the data dictionary are defined by six attributes described in C.1.2 to C.1.7.

NOTE The attributes are based on those specified in ISO/IEC 11179-3 for the description of data element concepts, (i.e. data elements without representation). The term “dataset” when used as part of a definition is synonymous with all types of geographic data resources (aggregations of datasets, individual features and the various classes that compose a feature).

C.1.2 Name/role name

A label assigned to a metadata entity or to a metadata element. Metadata entity names start with an upper case letter. Spaces do not appear in a metadata entity name. Instead, multiple words are concatenated, with each new subword starting with a capital letter (example: XnnnYmmm). Metadata entity names are unique within the entire data dictionary of this International Standard. Metadata element names are unique within a metadata entity, not the entire data dictionary of this International Standard. Metadata element names are made unique, within an application, by the combination of the metadata entity and metadata element names. Role names are used to identify metadata abstract model associations and are preceded by “Role name:” to distinguish them from other metadata elements. Names and role names may be in a language other than that used in this International Standard.

C.1.3 Definition

This is the metadata entity/element description.

C.1.4 Obligation/Condition

C.1.4.1 General

This is a descriptor indicating whether a metadata entity or metadata element shall always be documented in the metadata 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 metadata entity or metadata element shall be documented.

C.1.4.3 Conditional (C):

Specifies an electronically manageable condition under which at least one metadata entity or a metadata element 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 metadata entity or a metadata element if another element has been documented.
- Documenting a metadata element if a specific value for another metadata element 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 electronic user interface.

If the answer to the condition is positive, then the metadata entity or the metadata element shall be mandatory.

C.1.4.4 Optional (O):

The metadata entity or the metadata element may or need not be documented. Optional metadata entities and optional metadata elements 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 entity is not used, the elements contained within that entity (including mandatory elements) will also not be used. Optional entities may have mandatory elements; those elements only become mandatory if the optional entity is used.

C.1.5 Maximum occurrence

Specifies the maximum number of instances the metadata entity or the metadata element 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 metadata elements; for example, integer, real, string, DateTime, and Boolean. The data type attribute is also used to define metadata entities, stereotypes, and metadata associations.

NOTE Data types are defined in ISO/TS 19103:2005, 6.5.2.

C.1.7 Domain

For an entity, the domain indicates the line numbers covered by that entity.

For a metadata element, 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 Quality package data dictionaries

C.2.1 Data quality information

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	DQ_Scope <<DataType>> (C.2.1.6)
3.	<i>Role name:</i> report	quantitative quality information for the data specified by the scope	M	N	Association	DQ_Element <<Abstract>> (C.2.1.2)
4.	<i>Role name:</i> standaloneQualityReport	reference to external standalone quality report	O	1	Association	DQ_StandaloneQualityReportInformation (C.2.1.7)