DRAFT FOR DEVELOPMENT

DD ISO/TS 19138:2006

Geographic information — Data quality measures

ICS 35.240.70

NO COPYING WITHOUT BSI PERMISSION EXCEPT AS PERMITTED BY COPYRIGHT LAW

Document Sharing Hub
British Standards

Full document details: DD ISO/TS 19138:2006

National foreword

This Draft for Development was published by BSI. It is the UK implementation of ISO/TS 19138:2006.

This publication is not to be regarded as a British Standard.

It is being issued in the Draft for Development series of publications and is of a provisional nature. It should be applied on this provisional basis, so that information and experience of its practical application can be obtained.

Comments arising from the use of this Draft for Development are requested so that UK experience can be reported to the international organization responsible for its conversion to an international standard. A review of this publication will be initiated not later than 3 years after its publication by the international organization so that a decision can be taken on its status. Notification of the start of the review period will be made in an announcement in the appropriate issue of *Update Standards*.

According to the replies received by the end of the review period, the responsible BSI Committee will decide whether to support the conversion into an international Standard, to extend the life of the Technical Specification or to withdraw it. Comments should be sent to the Secretary of the responsible BSI Technical Committee at British Standards House, 389 Chiswick High Road, London W4 4AL.

The UK participation in its preparation was entrusted to Technical Committee IST/36, Geographic information.

A list of organizations represented on IST/36 can be obtained on request to its secretary.

This publication does not purport to include all the necessary provisions of a contract. Users are responsible for its correct application.

This Draft for Development was published under the authority of the Standards Policy and Strategy Committee on 29 December 2006

© BSI 2006

ISBN 0 580 49896 4

Amendments issued since publication

Amd. No.	Date	Comments

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

TECHNICAL SPECIFICATION

ISO/TS 19138

First edition
2006-12-01

Geographic information — Data quality measures

Information géographique — Mesures de la qualité des données

Reference number
ISO/TS 19138:2006(E)

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

Contents

Page

Foreword iv

Introduction v

1 Scope 1

2 Conformance 1

3 Normative references 1

4 Terms and definitions 1

5 Relationships to other standards 2

6 Register 3

7 Components of a data quality measure 3
7.1 List of components 3
7.2 Component details 4
7.3 Standardized data quality measures 6

Annex A (normative) **Abstract test suite** 7
A.1 Test case identifier: Component test 7
A.2 Test case identifier: Name test 7
A.3 Test case identifier: Data quality element and subelement test 7
A.4 Test case identifier: Data quality basic measure test 8
A.5 Test case identifier: Definition test 8
A.6 Test case identifier: Description test 8
A.7 Test case identifier: Parameter test 8
A.8 Test case identifier: Data quality value type test 8
A.9 Test case identifier: Source reference test 9
A.10 Test case identifier: Example test 9

Annex B (normative) **Structure of data quality measures** 10
B.1 Components defining a data quality measure 10
B.2 Mapping of the components to ISO 19115 and ISO 19135 11
B.3 UML-diagram for data quality measure 11

Annex C (normative) **Data quality basic measures** 14
C.1 Purpose of data quality basic measures 14
C.2 Counting-related data quality basic measures 14
C.3 Uncertainty-related data quality basic measures 15

Annex D (normative) **List of data quality measures** 19
D.1 Completeness 19
D.2 Logical consistency 23
D.3 Positional accuracy 37
D.4 Temporal accuracy 58
D.5 Thematic accuracy 61

Bibliography 68

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

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.

In other circumstances, particularly when there is an urgent market requirement for such documents, a technical committee may decide to publish other types of normative document:

— an ISO Publicly Available Specification (ISO/PAS) represents an agreement between technical experts in an ISO working group and is accepted for publication if it is approved by more than 50 % of the members of the parent committee casting a vote;

— an ISO Technical Specification (ISO/TS) represents an agreement between the members of a technical committee and is accepted for publication if it is approved by 2/3 of the members of the committee casting a vote.

An ISO/PAS or ISO/TS is reviewed after three years in order to decide whether it will be confirmed for a further three years, revised to become an International Standard, or withdrawn. If the ISO/PAS or ISO/TS is confirmed, it is reviewed again after a further three years, at which time it must either be transformed into an International Standard or be withdrawn.

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/TS 19138 was prepared by Technical Committee ISO/TC 211, *Geographic information/Geomatics*.

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

Introduction

Knowledge of the quality of geographic data is often crucial for the application of the data, as different users and different applications often have different data quality requirements. 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. To facilitate such 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.

Data quality needs to be reported by the producer and evaluated by the user against his or her requirements for different criteria and data quality measures. It is essential that reported quality for a dataset contains the quality measurements that may be of interest to a potential user of the dataset, and that the metrics used to determine the quality are reported and available to the user.

ISO 19113 establishes the principles for the description of geographic data quality and specifies components for reporting quality information. Procedures for the evaluation of geographic data quality are described in ISO 19114.

The objective of this Technical Specification is to guide the producer in choosing the right data quality measures for data quality reporting, and the user in the evaluation of the usefulness of a dataset by standardizing the components and structures of data quality measures and by defining commonly used data quality measures.

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

blank

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

Geographic information — Data quality measures

1 Scope

This Technical Specification defines a set of data quality measures. These can be used when reporting data quality for the data quality subelements identified in ISO 19113. Multiple measures are defined for each data quality subelement, and the choice of which to use will depend on the type of data and its intended purpose.

The data quality measures are structured so that they can be maintained in a register established in conformance with ISO 19135.

This Technical Specification does not attempt to describe every possible data quality measure, only a set of commonly used ones.

2 Conformance

Any set of data quality measures claiming conformance with this Technical Specification shall pass all of the conditions specified in the abstract test suite (Annex A).

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 19113:2002, *Geographic information — Quality principles*

ISO 19115:2003, *Geographic information — Metadata*

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

4 Terms and definitions

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

4.1
correctness
correspondence with the universe of discourse

4.2
data quality basic measure
generic data quality 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.

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

4.3
data quality scope
extent or characteristic(s) of the data for which quality information is reported

[ISO 19113]

NOTE A data quality scope for a dataset can comprise a dataset series to which the dataset belongs, the dataset itself, or a smaller grouping of data located physically within the dataset sharing common characteristics. Common characteristics can be an identified feature type, feature attribute, or feature relationship; data collection criteria; original source; or a specified geographic or temporal extent.

4.4
error
discrepancy with the universe of discourse

4.5
measurand
particular quantity subject to measurement

[International Vocabulary of Basic and General Terms in Metrology (VIM)]

4.6
universe of discourse
view of the real or hypothetical world that includes everything of interest

[ISO 19101]

5 Relationships to other standards

ISO 19113 describes relevant data quality elements and their corresponding data quality subelements and it indicates how quality should be reported. ISO 19114 describes procedures for the evaluation of quantitative quality. ISO 19115 contains elements and classes for data quality reporting within the UML models and data dictionaries.

ISO 19113 specifies a set of descriptors for a data quality subelement, for use in recording data quality. One of these descriptors is the data quality measure. A data quality measure is described by the components listed in 7.1.

Table 1 provides a list of data quality elements and data quality subelements as defined in ISO 19113.

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

Table 1 — Data quality elements and data quality subelements with definitions (ISO 19113)

Data quality element	Data quality subelement	Definition
completeness	commission	excess data present in a dataset
	omission	data absent from a dataset
logical consistency	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
positional accuracy	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 position values to values accepted as or being true
temporal accuracy	accuracy of a time measurement	correctness of the temporal references of an item (reporting of error in time measurement)
	temporal consistency	correctness of ordered events or sequences, if reported
	temporal validity	validity of data with respect to time
thematic accuracy	classification correctness	comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference dataset)
	non-quantitative attribute correctness	correctness of non-quantitative attribute
	quantitative attribute accuracy	accuracy of quantitative attributes

6 Register

A register of data quality measures shall contain a set of data quality measures, described using the components listed in 7.1. The registration procedures shall be performed according to ISO 19135.

Annex D of this Technical Specification contains the list of standardized data quality measures. A register shall contain these data quality measures and may also contain additional data quality measures submitted through the procedures defined within ISO 19135. The registration process also allows retiring data quality measures.

7 Components of a data quality measure

7.1 List of components

Each data quality measure shall be described using the following technical components:

— name (7.2.1)

— alias (7.2.2)

AVA Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

— data quality element (7.2.3)

— data quality subelement (7.2.4)

— data quality basic measure (7.2.5)

— definition (7.2.6)

— description (7.2.7)

— parameter (7.2.8)

— data quality value type (7.2.9)

— data quality value structure (7.2.10)

— source reference (7.2.11)

— example (7.2.12)

— identifier (7.2.13)

7.2 Component details

7.2.1 Name

Name refers to the name of the data quality measure.

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

NOTE The component name is specified in the base standard for registers, ISO 19135.

7.2.2 Alias

Alias refers to other 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.

7.2.3 Data quality element

Data quality element refers to the name of the data quality element to which this data quality measure applies.

NOTE A list of data quality elements is provided in Table 1.

7.2.4 Data quality subelement

Data quality subelement refers to the name of the data quality subelement to which this data quality measure applies.

NOTE A list of data quality subelements is provided in Table 1.

7.2.5 Data quality basic measure

Each data quality basic measure is described by its name, definition and value type. Data quality basic measures are identified by their names.

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

A variety of data quality measures are based on counting of erroneous items. There are also several data quality measures dealing with the uncertainty of numerical values. In order to avoid repetition, all possible methods of constructing counting-related data quality measures as well as general statistical measures for one- and two-dimensional random variables shall be defined in terms of data quality basic measures.

The data quality basic measures are defined in Annex C.

If a data quality measure is based on one of the set of data quality basic measures, the name of the data quality basic measure shall be provided in the field data quality basic measure. If the data quality measure is not based on a data quality basic measure, it shall be indicated in this field that a data quality basic measure is not applicable. The data quality basic measures shall also be used as appropriate for creating new data quality measures, for instance for reporting unclosed surface patches or other application-dependent data quality measures.

7.2.6 Definition

Definition states the fundamental concept of the data quality measure.

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

NOTE The component definition is specified in the base standard for registers, ISO 19135.

7.2.7 Description

Description refers to the description of the data quality measure including methods of calculation, with all formulae and/or illustrations needed to establish the result of applying the measure.

If the data quality measure uses the concept of errors, it shall be stated how an item shall be classified as incorrect.

NOTE The component description is specified in the base standard for registers, ISO 19135.

7.2.8 Parameter

Parameter refers to an auxiliary variable used by the data quality measure. It shall include name, definition and description.

More than one parameter may be provided.

7.2.9 Data quality value type

Data quality value type refers to the value type for reporting a data quality result.

A data quality value type shall be provided for a data quality result. The data types defined in ISO/TS 19103 shall be used when appropriate.

Table 2 — Examples of data quality value types

Boolean
Real
Integer
Ratio (numerator of type integer : denominator of type integer)
Percentage
Measure(s) [value(s) + unit(s)]

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

7.2.10 Data quality value structure

Data quality value structure gives the structure for reporting a complex data quality result.

A data quality result may consist of multiple values. In this case, the data quality result shall be structured using the data quality value structures as given in Table 3. The structure may consist of homogeneous or heterogeneous data quality value types. The possible data quality value types are given in 7.2.9.

Table 3 — Data quality value structures

Bag
Set
Sequence
Table
Matrix
Coverage

NOTE The values within a structure can be multiple. For example, the covariance matrix as given in Table D.32 is reported as matrix of measure, where the matrix elements may have different units of measure. A list may consist of different data quality value types.

7.2.11 Source reference

Source reference gives the citation of the source of the data quality measure.

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

NOTE The component source reference is specified in the base standard for registers, ISO 19135.

7.2.12 Example

Example may provide examples of applying the data quality measure or the result obtained for the data quality measure.

More than one example may be provided.

7.2.13 Identifier

Identifier consists of an integer number that uniquely identifies a data quality measure.

If data quality measures are administered in a register, then identifiers may only be assigned by the register manager.

NOTE The component identifier is specified in the base standard for registers, ISO 19135.

7.2.14 Obligation of the above-listed components

Some of the components are mandatory, others are conditional or optional. Table B.1 provides further information on the obligation of each technical component.

7.3 Standardized data quality measures

In order to make data quality related metadata and data quality reports comparable, standardized data quality measures shall be used in evaluating and reporting data quality, where appropriate. Annex D gives a list of commonly used data quality measures with all required components for data quality measures as specified in this Technical Specification.

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

Annex A
(normative)

Abstract test suite

A.1 Test case identifier: Component test

a) Test purpose: to determine conformance by ensuring that all necessary components of a data quality measure are provided.

b) Test method: examine the entry for the data quality measure and verify that the components have been provided as required by Table B.1.

c) Reference: 7.2 and Annex B.

d) Test type: Capability.

A.2 Test case identifier: Name test

a) Test purpose: to determine if a distinct name for the data quality measure is used.

b) Test method: determine if the name for the data quality measure is distinct from other measures with different concepts, and if the name is not in conflict with other data quality basic measures, their definitions and descriptions.

c) Reference: 7.2.1.

d) Test type: Capability.

A.3 Test case identifier: Data quality element and subelement test

a) Test purpose: to determine

— if data quality element and subelement are assigned;

— if they are taken from the list of data quality elements and subelements in ISO 19113 or if they are an additional data quality element and subelement created in conformance with the rules of ISO 19113;

— if the data quality measure is relevant for the given data quality element and subelement.

b) Test method: check if proper values are assigned to the data quality element and subelement components and if the data quality measure has bearing on these.

c) Reference: 7.2.3 and 7.2.4.

d) Test type: Capability.

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

A.4 Test case identifier: Data quality basic measure test

a) Test purpose: to determine if a data quality measure is properly derived from a data quality basic measure.

b) Test method: check if an appropriate data quality basic measure for the data quality measure exists and, if it does, that the data quality measure is utilizing this data quality basic measure in conformance with this Technical Specification.

c) Reference: 7.2.5.

d) Test type: Capability.

A.5 Test case identifier: Definition test

a) Test purpose: to determine if a fitting, correct and complete definition is provided.

b) Test method: check that the given definition contains no ambiguities and that it is in conformance with characteristics of a definition as stated in ISO 19135:2005, 7.3.1.

c) Reference: 7.2.6 and ISO 19135:2005, 7.3.1.

d) Test type: Capability.

A.6 Test case identifier: Description test

a) Test purpose: to determine if an exhaustive description is provided.

b) Test method: check if the description contains a comprehensive explanation with all required formulae to facilitate the application of the data quality measure.

c) Reference: 7.2.7.

d) Test type: Capability.

A.7 Test case identifier: Parameter test

a) Test purpose: to determine if required parameters are provided.

b) Test method: check if all parameters occurring in the description are provided in the parameter component.

c) Reference: 7.2.8.

d) Test type: Capability.

A.8 Test case identifier: Data quality value type test

a) Test purpose: to determine if a proper data quality value type is provided.

b) Test method: check if the provided data quality value type is included in the list in Table 3.

c) Reference: 7.2.9.

d) Test type: Capability.

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

A.9 Test case identifier: Source reference test

a) Test purpose: to determine if a proper source reference is provided.

b) Test method: check if the cited reference source exists and if it reflects the concept of the provided data quality measure.

c) Reference: 7.2.11.

d) Test type: Capability.

A.10 Test case identifier: Example test

a) Test purpose: to determine if the example, if provided, is a valid example for the data quality measure.

b) Test method: check if the example is free of errors and if it is representative of the usage of the data quality measure.

c) Reference: 7.2.12.

d) Test type: Capability.

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

Annex B
(normative)

Structure of data quality measures

B.1 Components defining a data quality measure

Table B.1 shall be used for the technical specification of every data quality measure. The descriptor for obligation/condition may have the following values: M (mandatory), C (conditional), or O (optional).

Table B.1 — Components defining a data quality measure

Line	Component	Description	Obligation/condition
1	Name	Name of the data quality measure applied to the data	M
2	Alias [a]	Another recognized name, an abbreviation or a short name for the same data quality measure	O
3	Data quality element	Name of the data quality element for which quality is reported	M
4	Data quality subelement	Name of the data quality subelement for which quality is reported	M
5	Data quality basic measure	Name of the data quality basic measure from which the data quality measure is derived	C/if derived from basic measure
6	Definition	Definition of the fundamental concept for the data quality measure	M
7	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
8	Parameter [a]	Auxiliary variable used by the data quality measure, including its name, definition and optionally its description	C/if required
9	Data quality value type [a]	Value type for reporting a data quality result	M
10	Data quality value structure	Structure for reporting a complex data quality result	O
11	Source reference [a]	Reference to the source of an item that has been adopted from an external source	C/if an external source exists
12	Example [a]	Illustration of the use of a data quality measure	O
13	Identifier	Integer number, uniquely identifying a data quality measure	C/if data quality measures are administered in a register

[a] Multiple entries are allowed. When values for the optional or conditional elements are not present, this should be indicated by assigning the character "—" to the appropriate component.

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

B.2 Mapping of the components to ISO 19115 and ISO 19135

Table B.2 — Mapping of the components to ISO 19115 and ISO 19135

Line	Component	ISO 19115 element name	ISO 19135 element name
1	Name	nameOfMeasure	name
2	Alias	–	alternativeExpressions
3	Data quality element	DQ_Element	–
4	Data quality subelement	lines 108-127 [B.2.4.3 Data quality element information]	–
5	Data quality basic measure	–	–
6	Definition	–	definition
7	Description	measureDescription	description
8	Parameter	–	–
9	Data quality value type	–	–
10	Data quality value structure	–	–
11	Source reference	–	source
12	Example	–	–
13	Identifier	measureIdentification	itemIdentifier

B.3 UML-diagram for data quality measure

Figure B.1 defines the components for data quality measures and Figure B.2 defines the relationship of data quality measures and registered items from ISO 19135. Both figures are in UML notation.

The UML models describe the content model, if a register for data quality measures is implemented.

The class RE_RegisteredItem is defined in ISO 19135.

AVA Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

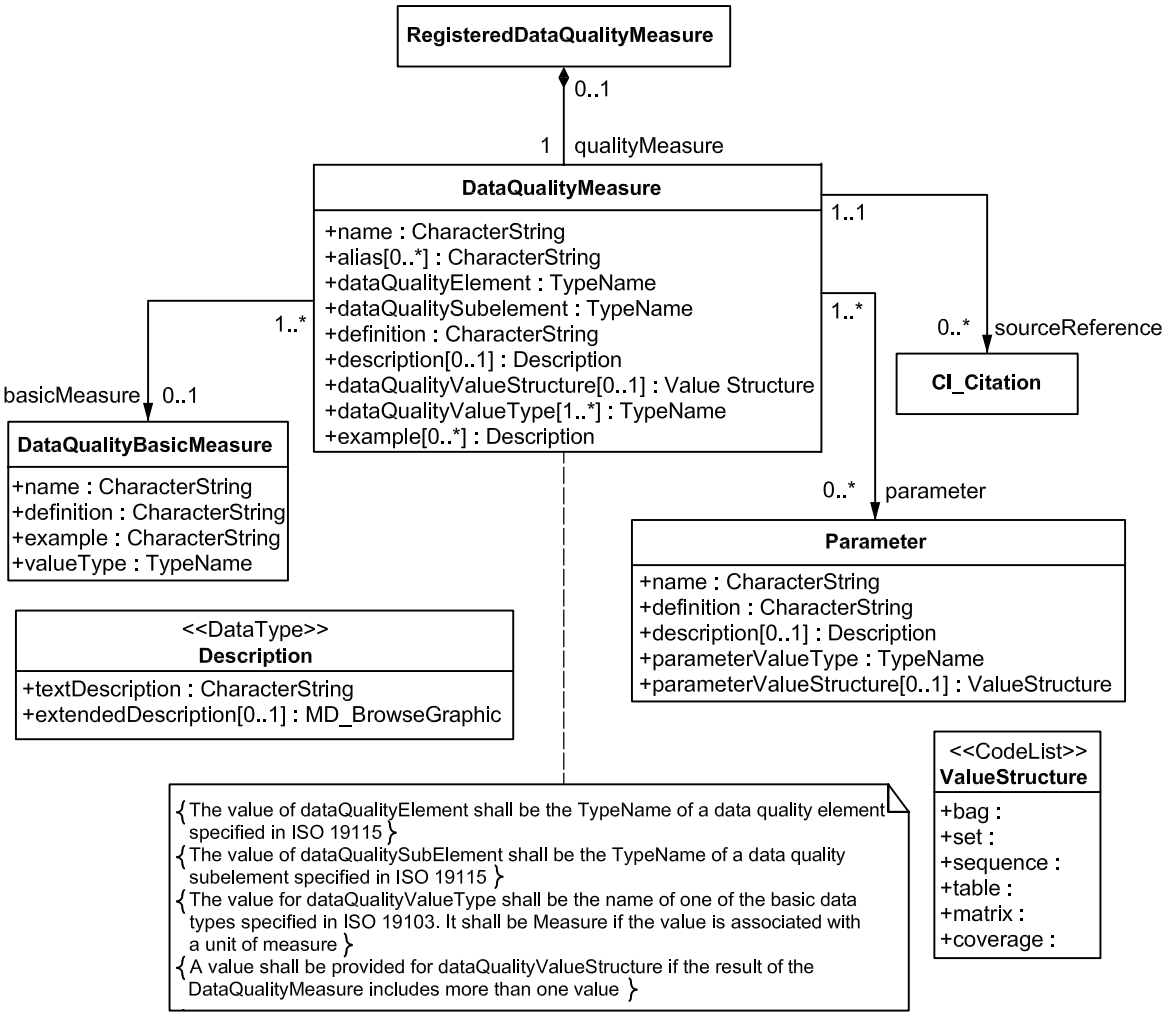

Figure B.1 — Data quality measure

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

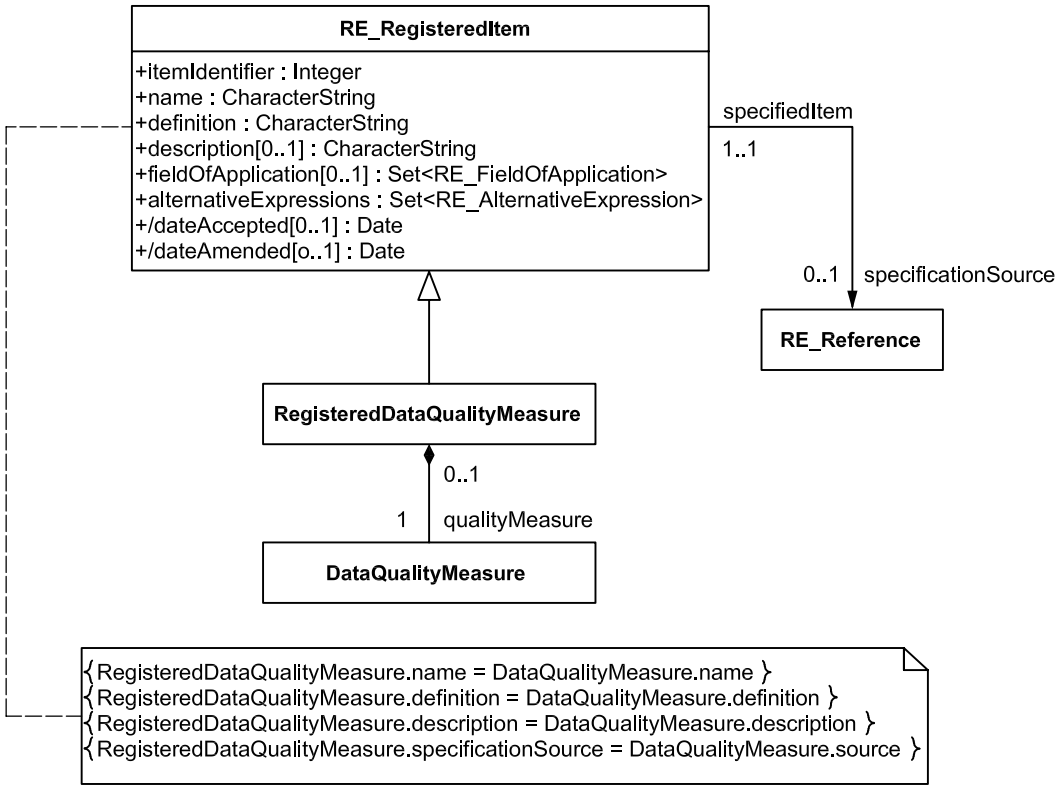

Figure B.2 — Relationship between registered item of ISO 19135 and data quality measure

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

Annex C
(normative)

Data quality basic measures

C.1 Purpose of data quality basic measures

The concept of data quality basic measures is introduced in this Technical Specification to avoid the repetitive definition of the same concept. There are data quality measures that have certain communalities. 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 shall be used for the creation of data quality measures that share these communalities.

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 shall be used to construct data quality measures.

Annex D uses the data quality basic measures of Annex C where appropriate. When appropriate, the construction of new data quality measures shall be derived from one of the following data quality basic measures.

C.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 are listed in Table C.1.

Table C.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 not in error	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 1,89% 11,582	Error rate can either be presented as real, percentage or as ratio
Correct items rate	Number of the correct items with respect to the total number of items that should have been present	0,9811 98,11% 571:582	Correct items rate can either be presented as real, percentage or as ratio

NOTE 1 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.

NOTE 2 A list of data quality value types is provided in Table 2 (see 7.2.9).

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006

C.3 Uncertainty-related data quality basic measures

C.3.1 General

Numerical values that are obtained by some kind of measuring procedure can only be observed to a certain accuracy. By treating the measured quantity (measurand) as 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 normal distribution.

C.3.2 One-dimensional random variable, Z

For a continuous measurand (i.e. the value domain of the measured quantities is the real numbers), 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.

$$P(\text{lower limit} \leqslant \text{true value} \leqslant \text{upper limit}) = \text{P}$$

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

Table C.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 = 68,3 %	$u_{68,3\%}$ = 1	$u_{68,3\%} \cdot \sigma_Z$	LE68.3	measure
P = 50 %	$u_{50\%}$ = 0,6745	$u_{50\%} \cdot \sigma_Z$	LE50	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.

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

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

Documents
Document Sharing Hub

Full document details: DD ISO/TS 19138:2006

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:

$$P\left(-t \cdot s_z \leqslant Z - z_t \leqslant t \cdot s_z\right) = \mathrm{P} \text{ with } (Z - z_t)/s_z \sim t(r)$$

Table C.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 C.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 C.3.			

The data quality basic measures for the uncertainty of one-dimensional quantities are given in Tables C.2 and C.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 C.2 is relevant. If the standard deviation is estimated from redundant measurements, then Table C.4 in conjunction with Table C.3 is relevant.

Documents
Document Sharing Hub
Full document details: DD ISO/TS 19138:2006