

---

---

## Geographic information — Data quality

*Information géographique — Qualité des données*

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO 19157:2013

<https://standards.iteh.ai/catalog/standards/sist/902e57e3-a427-47c2-89a5-e5a7e09375a8/iso-19157-2013>



**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO 19157:2013

<https://standards.iteh.ai/catalog/standards/sist/902e57e3-a427-47c2-89a5-e5a7e09375a8/iso-19157-2013>



**COPYRIGHT PROTECTED DOCUMENT**

© ISO 2013

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office  
Case postale 56 • CH-1211 Geneva 20  
Tel. + 41 22 749 01 11  
Fax + 41 22 749 09 47  
E-mail [copyright@iso.org](mailto:copyright@iso.org)  
Web [www.iso.org](http://www.iso.org)

Published in Switzerland

# Contents

	Page
<b>Foreword</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>v</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Conformance</b> .....	<b>1</b>
<b>3 Normative references</b> .....	<b>1</b>
<b>4 Terms and definitions</b> .....	<b>2</b>
<b>5 Abbreviated terms</b> .....	<b>4</b>
5.1 Abbreviations.....	4
5.2 Package abbreviations.....	5
<b>6 Overview of data quality</b> .....	<b>5</b>
<b>7 Components of data quality</b> .....	<b>6</b>
7.1 Overview of the components.....	6
7.2 Data quality unit.....	7
7.3 Data quality elements.....	8
7.4 Descriptors of data quality elements.....	11
7.5 Metaquality elements.....	14
7.6 Descriptors of a metaquality element.....	15
<b>8 Data quality measures</b> .....	<b>16</b>
8.1 General.....	16
8.2 Standardized data quality measures.....	16
8.3 User defined data quality measures.....	16
8.4 Catalogue of data quality measures.....	16
8.5 List of components.....	17
8.6 Component details.....	18
<b>9 Data quality evaluation</b> .....	<b>20</b>
9.1 The process for evaluating data quality.....	20
9.2 Data quality evaluation methods.....	21
9.3 Aggregation and derivation.....	23
<b>10 Data quality reporting</b> .....	<b>23</b>
10.1 General.....	23
10.2 Particular cases.....	24
<b>Annex A (normative) Abstract test suites</b> .....	<b>26</b>
<b>Annex B (informative) Data quality concepts and their use</b> .....	<b>28</b>
<b>Annex C (normative) Data dictionary for data quality</b> .....	<b>34</b>
<b>Annex D (normative) List of standardized data quality measures</b> .....	<b>50</b>
<b>Annex E (informative) Evaluating and reporting data quality</b> .....	<b>96</b>
<b>Annex F (informative) Sampling methods for evaluating</b> .....	<b>119</b>
<b>Annex G (normative) Data quality basic measures</b> .....	<b>127</b>
<b>Annex H (informative) Management of data quality measures</b> .....	<b>132</b>
<b>Annex I (informative) Guidelines for the use of Quality Elements</b> .....	<b>135</b>
<b>Annex J (informative) Aggregation of data quality results</b> .....	<b>144</b>
<b>Bibliography</b> .....	<b>146</b>

## Foreword

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

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

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

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

For an explanation on the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the WTO principles in the Technical Barriers to Trade (TBT) see the following URL: Foreword - Supplementary information

The committee responsible for this document is ISO/TC 211, Geographic information/Geomatics

This edition of ISO 19157:2013 cancels and replaces ISO/TS 19138:2006, ISO 19114:2003 and ISO 19113:2002, which have been technically revised.

## Introduction

Geographic data are increasingly being shared, interchanged and used for purposes other than their producers' intended ones. Information about the quality of available geographic data are vital to the process of selecting a data set in that the value of data are directly related to its quality. A user of geographic data may have multiple data sets from which to choose. Therefore, it is necessary to compare the quality of the data sets 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 data set best suited to application needs or requirements. Complete descriptions of the quality of a data set will encourage the sharing, interchange and use of appropriate data sets. Information on the quality of geographic data allows a data producer to evaluate how well a data set meets the criteria set forth in its product specification and assists data users in evaluating a product's ability to satisfy the requirements for their particular application. For the purpose of this evaluation, clearly defined procedures are used in a consistent manner.

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

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

The objective of this International Standard is to provide principles for describing the quality for geographic data and concepts for handling quality information for geographic data, and a consistent and standard manner to determine and report a data set's quality information. It aims also to provide guidelines for evaluation procedures of quantitative quality information for geographic data.

**iTeh STANDARD PREVIEW**  
**(standards.iteh.ai)**

ISO 19157:2013

<https://standards.iteh.ai/catalog/standards/sist/902e57e3-a427-47c2-89a5-e5a7e09375a8/iso-19157-2013>

# Geographic information — Data quality

## 1 Scope

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

- defines components for describing data quality;
- specifies components and content structure of a register for data quality measures;
- describes general procedures for evaluating the quality of geographic data;
- establishes principles for reporting data quality.

This International Standard also defines a set of data quality measures for use in evaluating and reporting data quality. It is applicable to data producers providing quality information to describe and assess how well a data set conforms to its product specification and to data users attempting to determine whether or not specific geographic data are of sufficient quality for their particular application.

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

## iTeh STANDARD PREVIEW (standards.iteh.ai)

## 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](#);  
<https://standards.iteh.ai/catalog/standards/sist/902e57e3-a427-47c2-89a5-c3a7c0933a0/iso-19157-2013>
- 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, in whole or in part, are normatively referenced in this document and are indispensable for its application. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

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

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

ISO 19115-1:2014, *Geographic information — Metadata — Part 1: Fundamentals*<sup>1)</sup>

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

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

1) Under preparation.

## 4 Terms and definitions

### 4.1

#### **accuracy**

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

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

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

### 4.2

#### **catalogue**

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

[SOURCE: ISO 10303-227:2005, 3.3.10, modified - Note has been deleted.]

### 4.3

#### **conformance**

fulfilment of specified requirements

[SOURCE: ISO 19105:2000, 3.8]

### 4.4

#### **conformance quality level**

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

### 4.5

#### **correctness**

correspondence with the *universe of discourse* (4.24)

### 4.6

#### **data product specification**

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

[SOURCE: ISO 19131:2007, 4.7, modified - Note has been deleted.]

### 4.7

#### **data quality basic measure**

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

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

### 4.8

#### **dataset**

identifiable collection of data

Note 1 to entry: A data set can be a smaller grouping of data which, though limited by some constraint such as spatial extent or *feature type* (4.15), is located physically within a larger data set. Theoretically, a data set can be as small as a single *feature* (4.11) or *feature attribute* (4.12) contained within a larger data set. A hardcopy map or chart can be considered a data set.

[SOURCE: ISO 19115-1:—, 4.3 ]<sup>2)</sup>

---

2) To be published.



**4.9****dataset series**

collection of *datasets* (4.8) sharing common characteristics  
 [SOURCE: ISO 19115-1:—, 4.10]<sup>3)</sup>

**4.10****direct evaluation method**

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

**4.11****feature**

abstraction of real world phenomena

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

[SOURCE: ISO 19101:2002, 4.11]

**4.12****feature attribute**

characteristic of a *feature* (4.11)

Note 1 to entry: A feature attribute has a name, a data type and a value domain associated with it. A feature attribute for a *feature instance* (4.13) also has an attribute value taken from the value domain.

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

**4.13****feature instance**

individual of a given *feature type* (4.15), having specified *feature attribute* (4.12) values  
 [SOURCE: ISO 19101-1:—, 4.1.14]<sup>4)</sup>

**4.14****feature operation**

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

[SOURCE: ISO 19110:2005, 4.5 - modified, Example and Note have been removed.]

**4.15****feature type**

class of *features* (4.11) having common characteristics

[SOURCE: ISO 19156:2011, 4.7]

**4.16****geographic data**

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

[SOURCE: ISO 19109:2005, 4.12, modified - Note has been deleted.]

**4.17****indirect evaluation method**

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

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

3) To be published.

4) To be published.

**4.18**

**item**

anything that can be described and considered separately

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

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

**4.19**

**metadata**

information about a resource

[SOURCE: ISO 19115-1:—, 4.9]<sup>5)</sup>

**4.20**

**metaquality**

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

**4.21**

**quality**

degree to which a set of inherent characteristics fulfils requirements

[SOURCE: ISO 9000:2005, 3.1.1, modified - Original Notes have been removed.]

**4.22**

**register**

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

[SOURCE: ISO 19135:2005, 4.1.9]

**4.23**

**standalone quality report**

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

**4.24**

**universe of discourse**

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

[SOURCE: ISO 19101:2002, 4.29]

## 5 Abbreviated terms

### 5.1 Abbreviations

ADQR	aggregated data quality results
AQL	acceptance quality limit [ISO 3534-2:2006]
RMSE	root mean square error
UML	Unified Modeling Language
XML	Extensible Markup Language

---

5) To be published.

## 5.2 Package abbreviations

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

CI	Citation [ISO 19115-1:2014]
CT	Catalogues [ISO/TS 19139:2007]
DQ	Data Quality [ISO 19157]
DQM	Data Quality Measure [ISO 19157]
EX	Extent [ISO 19115-1:2014]
GF	General Feature [ISO 19109:2005]
MD	Metadata [ISO 19115-1:2014]
QE	Quality Extended [ISO 19115-2:2009]
RE	Registration [ISO 19135:2005]

## 6 Overview of data quality

Working with data quality includes:

- understanding of the concepts of data quality related to geographic data. [Annex B](#) is a description of data quality concepts used to establish the components for describing the quality of geographic data;
- defining data quality conformance levels in data product specifications or based on user requirements. Establishment of data product specifications is described in ISO 19131:2007;
- specifying quality aspects in application schemas;
- evaluating data quality;
- reporting data quality.

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

A data quality evaluation can be applied to data set series, a data set or a subset of data within a data set, 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 data set meets the criteria set forth in its data product specification or user requirements and provide quantitative quality information.

When data quality information describes data that have been created without a detailed data product specification or with a data product specification that lacks quantitative measures and descriptors, the data element may be evaluated in a non-quantitative subjective way as a descriptive result for each element.

Some quality related information is provided by purpose, usage and lineage. This information is reported as metadata in conformance with ISO 19115-1:2014.

NOTE 2 Purpose describes the rationale for creating a data set and contains information about its intended use, which may not be the same as the actual use of the data set. Usage describes the application(s) for which a data set has been used, either by the data producer or by other data users. Lineage describes the history of a data set and recounts the life cycle of a data set from collection and acquisition through compilation and derivation to its current form. This general, non-quantitative information is illustrative for users and can help assessing the quality of a data set, especially in cases where it is used for a particular application that differs from the intended application (see also 9.2.3).

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

NOTE 3 The concept of metaquality is described in 7.5.

Figure 1 provides an overview of data quality information.

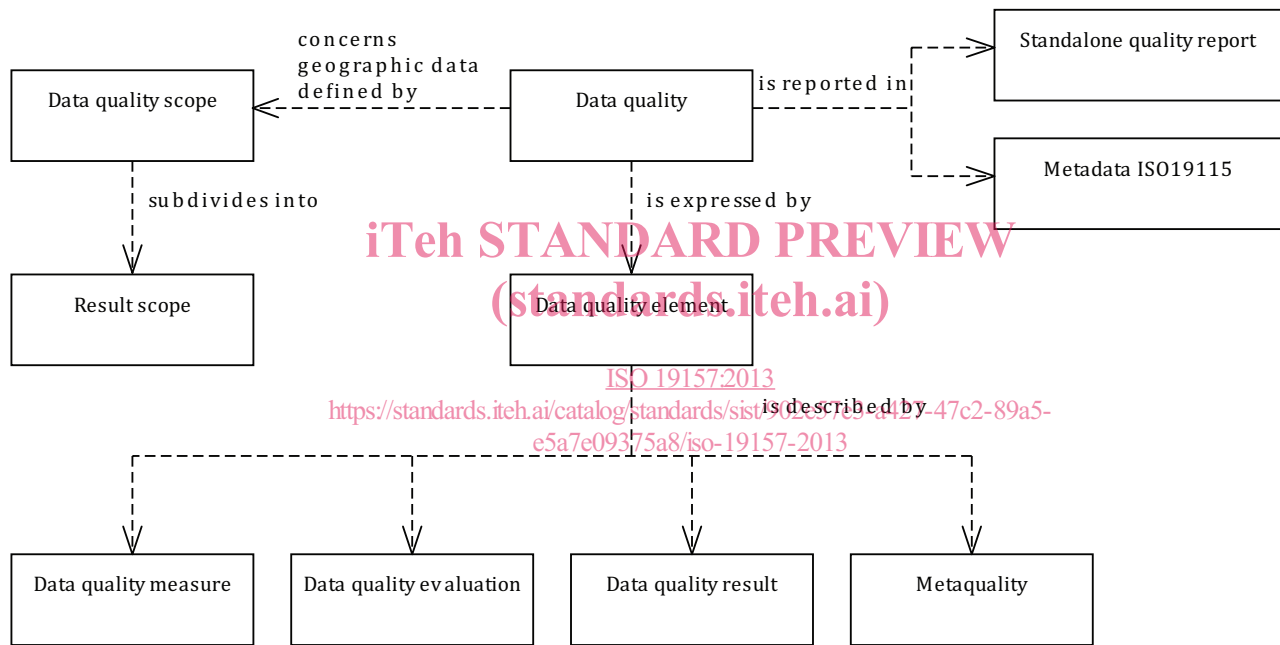


Figure 1 — Conceptual model of quality for geographic data

## 7 Components of data quality

### 7.1 Overview of the components

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

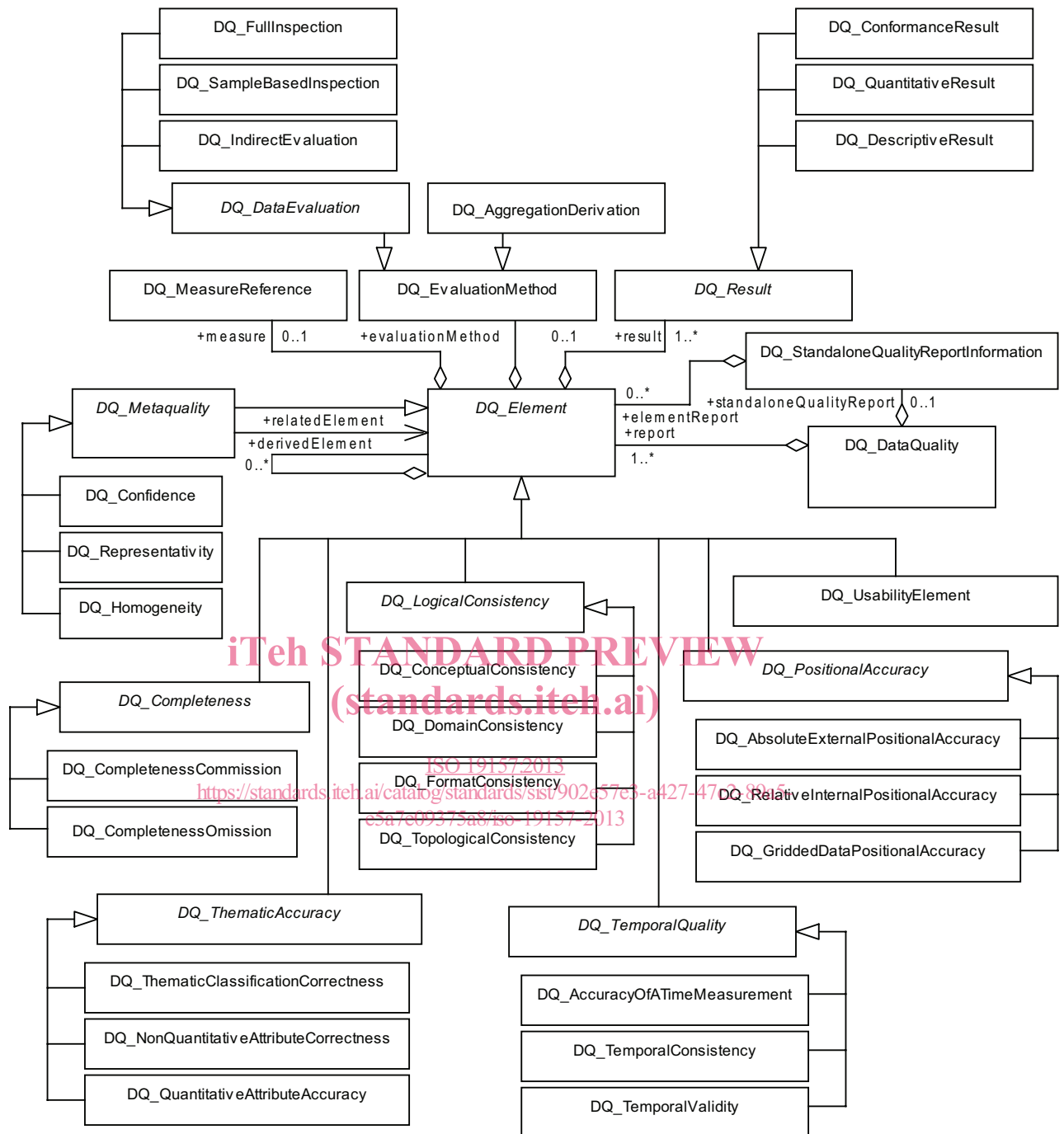


Figure 2 — Overview of the components of data quality

## 7.2 Data quality unit

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

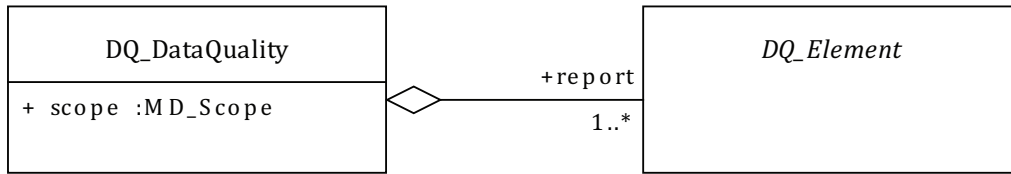


Figure 3 — Data quality unit

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

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

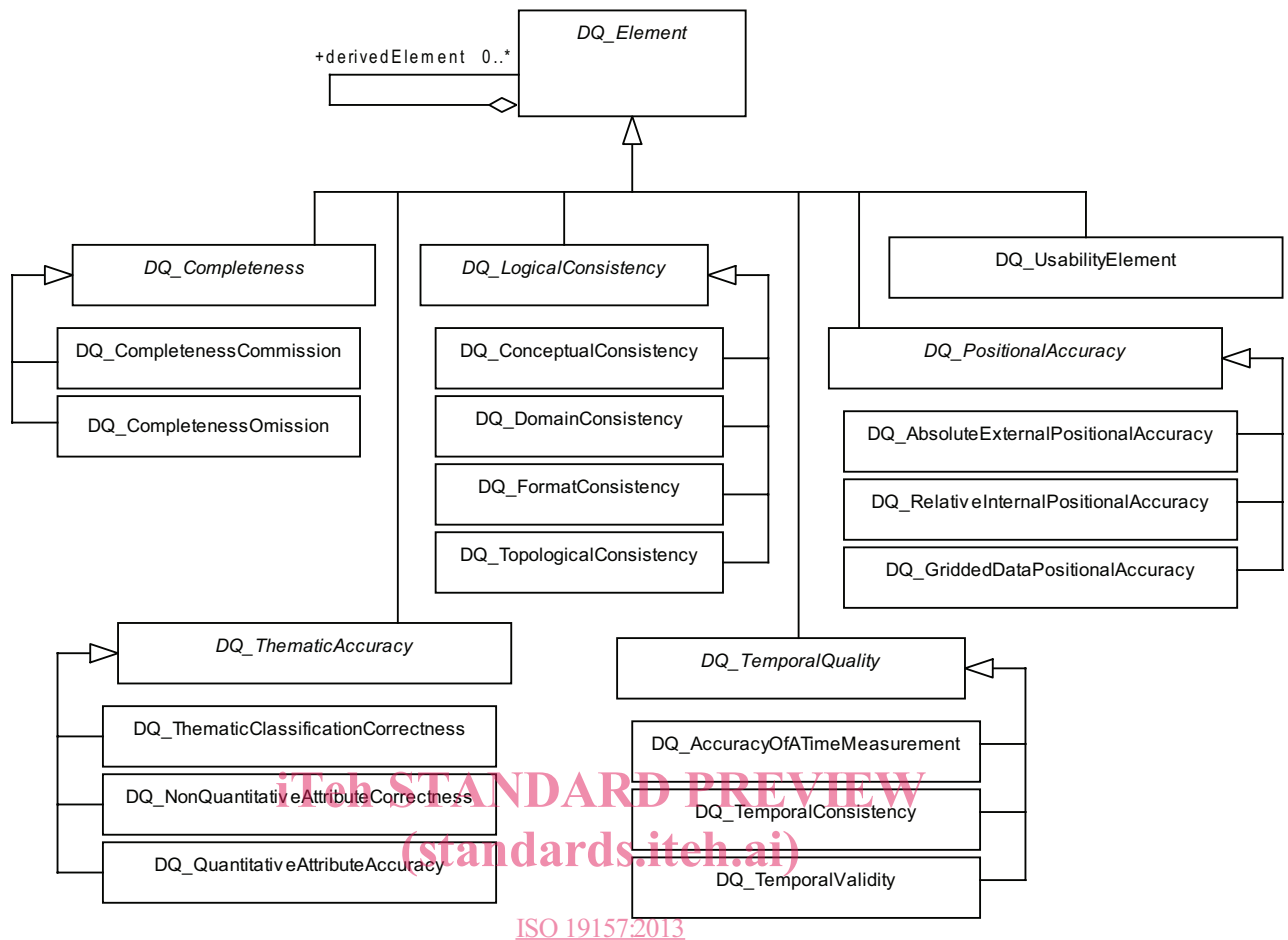
The following are examples of what defines a data quality scope (see also MD\_Scope in ISO 19115-1):

- a) a data set series;
- b) a data set;
- c) a subset of data defined by one or more of the following characteristics:
  - 1) types of items (sets of feature types, feature attributes, feature operations or feature relationships);
  - 2) specific items (sets of feature instances, attribute values or instances of feature relationships);
  - 3) geographic extent;
  - 4) temporal extent (the time frame of reference and accuracy of the time frame).

### 7.3 Data quality elements

#### 7.3.1 General

A data quality element is a component describing a certain aspect of the quality of geographic data and these have been organized into different categories. These categories are shown in [Figure 4](#).



ISO 19157:2013  
<https://standards.itech.ai/catalog/standards/sist/902e57e3-a427-47c2-89a5-c5b76375a710-19157-2013>  
**Figure 4 — Overview of the data quality elements**

### 7.3.2 Completeness

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

- commission: excess data present in a data set;
- omission: data absent from a data set.

### 7.3.3 Logical consistency

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

- conceptual consistency: adherence to rules of the conceptual schema;
- domain consistency: adherence of values to the value domains;
- format consistency: degree to which data are stored in accordance with the physical structure of the data set;
- topological consistency: correctness of the explicitly encoded topological characteristics of a data set.