INTERNATIONAL STANDARD

ISO 8000-8

First edition
2015-11-15

Data quality —

Part 8: Information and data quality: Concepts and measuring

Qualité des données —

Partie 8: Informations et qualité des données: Concepts et mesurage

Reference number
ISO 8000-8:2015(E)

ISO copyright office
Ch. de Blandonnet 8 • CP 401
CH-1214 Vernier, Geneva, Switzerland
Tel. +41 22 749 01 11
Fax +41 22 749 09 47
copyright@iso.org
www.iso.org

Contents

Page

Foreword

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

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

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

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

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

The committee responsible for this document is Technical Committee ISO/TC 184, *Industrial automation systems and integration*, Subcommittee SC 4 *Industrial data*.

ISO 8000 is organized as a series of parts, each published separately. The structure of ISO 8000 is described in ISO/TS 8000–1.

A complete list of parts of ISO 8000 is available from the Internet:

http://www.iso.org/iso/home/store.htm

Introduction

The ability to create, collect, store, maintain, transfer, process and present information and to support business processes in a timely and cost effective manner requires both an understanding of the characteristics of the information and data that determine its quality, and an ability to measure, manage and report on information and data quality.

ISO 8000 defines characteristics of information and data that determine its quality, and provides methods to manage, measure, and improve the quality of information and data.

When assessing the quality of information and data, it is useful to perform the assessment in accordance with documented methods. It is also important to document the tailoring of standardized methods with respect to the expectation and requirements pertinent to the business.

ISO 8000 includes parts applicable to all types of data, and parts applicable to specific types of data. ISO 8000 can be used independently or in conjunction with quality management systems.

This part of ISO 8000 can be used on its own or in conjunction with other parts of ISO 8000.

This part of ISO 8000 is intended for use by those actors that have a vested interest in information or data quality, with a focus on one or more information systems both inter-organization and intra-organization views, and throughout all life cycle phases.

When assessing whether the quality of information and data is sufficient, it is necessary to establish the threshold, pertinent to the business, for each object to be measured. This part of ISO 8000 does not set these thresholds.

When talking of measured values, it is important to state the scale used. This part of ISO 8000 does not define the scales against which the quality of information and data are measured, but call for them to be stated.

When communicating the result of the quantification of the quality of information and data, it is useful for the receiver to be able to understand the confidence of the result. In particular, it is important to know if any rule was not applied, or if any information or data was not checked.

This part of ISO 8000 provides the following:

— a definition of information and data quality;

— a structured way to plan and perform information and data quality measurements;

— prerequisites for measuring information and data quality;

— requirements for reporting information and data quality measurements.

This part of ISO 8000 is applicable independent of status of organization, type of information or data, hardware storage medium, software, information security and information life cycle stage.

This part of ISO 8000 can be used in relation to activities that use or depend on information or data.

These activities include capturing, storing, archiving, retrieving, tracking, transferring, displaying, delivering, and disposal of data.

NOTE The planned ISO 8000–9[1)] is intended to provide guidance on how to apply this part of ISO 8000 in a quality management system and through the life cycle stages of systems and software.

1) Under preparation.

Data quality —

Part 8: **Information and data quality: Concepts and measuring**

1 Scope

This part of ISO 8000 describes fundamental concepts of information and data quality, and how these concepts apply to quality management processes and quality management systems.

It also specifies prerequisites for measuring information and data quality when executed within quality management processes and quality management systems.

2 Normative references

The following 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 8000–2, *Data quality — Part 2: Vocabulary*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 8000–2 and the following apply.

3.1
data
reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

[SOURCE: ISO/IEC 2382:2015, 2121272, modified — Notes to entry have been removed.]

3.2
entity
concrete or abstract thing in the domain under consideration

[SOURCE: ISO 19439:2006, 3.29, modified — The word “any” has been removed at the start of the definition.]

3.3
information
knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning

[SOURCE: ISO/IEC 2382:2015, 2121271, modified — Field of application and notes to entry have been removed.]

3.4
metadata
data that defines and describes other data

[SOURCE: ISO/IEC 11179-1:2004, 3.2.16]

3.5
requirement
need or expectation that is stated, generally implied or obligatory

[SOURCE: ISO 9000:2015, 3.6.4, modified — Notes to entry have been removed.]

3.6
verification
confirmation, through the provision of objective evidence, that specified requirements have been fulfilled

[SOURCE: ISO 9000:2015, 3.8.12, modified — Notes to entry have been removed.]

3.7
validation
confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled

[SOURCE: ISO 9000:2015, 3.8.13, modified — Notes to entry have been removed.]

4 Fundamental concepts

An information system is a designed system that collects, stores, processes, and distributes information about the state of a domain. The information is represented and formalized in the system as data, and presented through user interfaces.

NOTE 1 For an explanation of states of domains, see Reference [13].

The main purpose of this part of ISO 8000 is to provide a foundation for measuring information and data quality. Information and data quality is defined and measured according to the following categories:

— syntactic quality, which is the degree to which data conforms to its specified syntax, i.e. requirements stated by the metadata;

— semantic quality, which is the degree to which data corresponds to what it represents;

— pragmatic quality, which is the degree to which data is found suitable and worthwhile for a particular purpose.

Measuring syntactic and semantic quality is performed through a verification process, while measuring pragmatic quality is performed through a validation process.

When measuring a physical object, the measurements can be described as the dimensions length, width and height. If the object is a cylinder, radius is a candidate characteristic. Establishing a useful set of characteristics pertinent to the case at hand is important in order to cater for communication of the measurements. This part of ISO 8000 offers a set of dimensions for quantifying the quality of information and data.

An activity model supporting information and data quality measurement is provided in Annex D.

See ISO/TS 8000–1 for the overall description and architecture of ISO 8000.

NOTE 2 The planned ISO/TS 8000–3[2)] is intended to further explain the architecture of ISO 8000.

2) Under preparation.

5 Information and data quality

5.1 Syntactic quality

Syntactic quality is the degree to which data conforms to its specified syntax, i.e. requirements stated by the metadata.

EXAMPLE 1 A specific set of data conforms to its implementation schema.

EXAMPLE 2 A specific sentence conforms to the grammatical rules of the language and use words from the language dictionary.

For a syntactic verification to comply with this part of ISO 8000, i.e. that syntactic quality can be measured, the following shall be available:

— a complete set of syntactic quality rules (see Annex B);

— a formal specification of how the information is expressed;

— for each syntactic rule:

  — the definition of how compliance is to be measured;

  — a description of the issues that will arise in the data set if the data does not comply to the rule;

  — a description of what it means not to comply to a rule;

  EXAMPLE 3 Duplicates could occur when the entity integrity rule is violated.

— if there are any rules defined that are not subject to checking then these shall be listed;

— the definition of how deviations are registered and presented;

— the number of checks performed for each rule;

— the number of occurrences that comply with each rule.

5.2 Semantic quality

Semantic quality is the unique and unambiguous correspondence of identifiable data units to the entities represented. Consequently, the requirement to fulfil is the correspondence of the data to the entities represented as these are viewed through a conceptual model. Semantic quality is the degree to which the data corresponds to what it represents.

NOTE 1 For an explanation of conceptual model and conceptual modelling, see Reference [13].

EXAMPLE 1 A specific sentence is required to be a true statement.

Correspondence between the data and what it represents is established by verifying that these criteria are fulfilled:

— mapped completely: every entity in the domain of interest shall be represented;

  EXAMPLE 2 Violation of criterion: in an employee register, an employee is not registered.

— mapped consistently: each entity in the domain of interest shall be represented by one of the following:

  — at most, one identifiable data unit; or

  — multiple but consistent identifiable units; or

  — multiple identifiable units whose inconsistencies are resolved within an acceptable time frame;

EXAMPLE 3 Violation of criterion: in an employee register, one employee has been given two employee numbers.

— mapped meaningfully: each identifiable data unit shall represent at least one specific entity in the domain of interest;

EXAMPLE 4 Violation of criterion: in an employee register, one employee number does not correspond to any employee.

— mapped unambiguously: each identifiable data unit shall represent at most one specific entity in the domain of interest;

EXAMPLE 5 Violation of criterion: in an employee register, two employees have been given the same employee number.

— entities in the domain of interest mapped correctly: each identifiable data unit shall map to the correct entity in the domain of interest;

EXAMPLE 6 Violation of criterion: in an employee register, there is a mix-up in employee number regarding two employees.

— properties mapped correctly: all attribute values in an identifiable data unit shall match the property values for the represented entity in the domain of interest.

EXAMPLE 7 Violation of criterion: in an employee register an employee has been registered with an incorrect address.

For a semantic verification to comply with this part of ISO 8000, i.e. that semantic quality can be measured, the following shall be available:

— a documented conceptual model of the universe of discourse including relationships;

— a presentation of the data in a form that enables them to be seen through the conceptual model, and thus mapped to what they represent;

— a prescription of the methods used for verifying the data according to the conceptual model;

— a description of how the rules derived from the criteria are realized as measurements;

— for each semantic rule:

  — the definition of how compliance is to be measured;

  — a description of what it means to not comply to a rule;

  EXAMPLE 8 Registrations may be missing from a data set when the mapped completely rule is violated.

— if there are external entities or types represented in the conceptual model that are not subject to checking, these shall be listed;

— a definition of how deviations are registered and presented;

— the number of checks performed for each rule;

— the number of occurrences that comply with each rule.

NOTE 2 A complete check of correspondence between data units and the external entities they represent is in many cases close to an impossible task. Sampling techniques and statistical methods can then be applied to give a probabilistic view of semantic quality.

NOTE 3 Existing “Trusted Surrogates” is recommended to be considered used instead of checking for correspondence with entities in the domain of interest. For example, an electronic representation of entities in the domain of interest that has already been verified.

5.3 Pragmatic quality

Pragmatic quality is conformance to usage-based requirements.

EXAMPLE 1 The data has to be understandable to the user.

EXAMPLE 2 Date of obsolescence is not to be exceeded.

Pragmatic quality shall be expressed as dimensions. These dimensions are specific perspectives validating defined requirements not directly covered by the syntactic or semantic quality categories.

For a pragmatic validation to comply with this part of ISO 8000, i.e. that pragmatic quality can be measured, the following shall be available:

— definitions of the quality dimensions used;

— a description of known dimension interdependencies;

— definitions of the methods applied for collecting the data to be used for measuring the quality dimension;

NOTE 1 Validation of pragmatic quality dimensions can be performed by utilizing formal requirements, questionnaires or interviews, or a combination of one or more of these.

— a description of metrics to measure the quality dimensions, including the following:

  — the definition of each metric;

  — the purpose of each metric;

NOTE 2 It is recommended to provide example of usage.

  — a description of how the measurements are presented;

  — results for each validated dimension.

NOTE 3 Different user groups can utilize the same data for different purposes. This is handled by declaring the purpose and viewpoint for each type of user.

For a coherent set of pragmatic quality dimensions, see Annex C.

6 Conformance

This part of ISO 8000 describes criteria for measuring data quality on three levels: syntactic, semantic and pragmatic. This part of ISO 8000 also provides criteria which constitute example sets to describe requirements for verification or validation of data quality. Verification and validation of the information and data quality conforms to this part of ISO 8000 when the corresponding conformance statement includes the following:

— reference to the subject data set;

— the applicable specified criteria for one or more of the following:

  — syntactic verification;

  — semantic verification;

  — pragmatic validation;

— the results of the measurements;

— when and by whom the verifications and validations are performed.

NOTE Further details are given in 5.1, 5.2 and 5.3.