ISO 8000-2:2022(E)

ISO TC 184/SC 4/WG 13

Date: 2022-07

Deleted: 02-14

Data quality — Part 2: Vocabulary

Qualité des données — Partie 2: Vocabulaire

ISO copyright office

CP 401 • Ch. de Blandonnet 8

CH-1214 Vernier, Geneva

Phone: +41 22 749 01 11

Email: copyright@iso.org

Website: www.iso.org

Published in Switzerland

Deleted: www.iso.org

Deleted: 2 © ISO 2022 – All rights reserved¶

Contents

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 of the voluntary nature of standards, 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 www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 184, *Automation systems and integration*, Subcommittee SC 4, *Industrial data*.

This fifth edition cancels and replaces the fourth edition (ISO 8000-2:2020), which has been technically revised. It also incorporates the Amendment ISO 8000-2:2020/Amd 1:2021.

The main changes are as follows:

Deleted: compared to the previous edition

— additional terminological entries to align the ISO 8000 series further with ISO 9000;

— updates where the updates originate from a new edition of ISO 10303-110;

— updates where the updates originate from converting ISO 8000-150 to an International Standard from a Technical Specification;

Deleted: 10303

— updates where the updates originate from a new edition of ISO 10303-59;

— other minor improvements to entries to improve consistency and readability of entries.

A list of all parts in the ISO 8000 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Deleted: 4 © ISO 2022 – All rights reserved¶

Introduction

Digital data deliver value by enhancing all aspects of organizational performance including:

— operational effectiveness and efficiency;

— safety;

— reputation with customers and the wider public;

— compliance with statutory regulations;

— innovation;

— consumer costs, revenues and stock prices.

In addition, many organizations are now addressing these considerations with reference to the United Nations Sustainable Development Goals[1].

The influence on performance originates from data being the formalized representation of information[2]. This information enables organizations to make reliable decisions. Such decision making can be performed by human beings directly and also by automated data processing including artificial intelligence systems.

Through widespread adoption of digital computing and associated communication technologies, organizations become dependent on digital data. This dependency amplifies the negative consequences of lack of quality in these data. These consequences are the decrease of organizational performance.

The biggest impact of digital data comes from two key factors:

— the data having a structure that reflects the nature of the subject matter;

EXAMPLE 1 A research scientist writes a report using a software application for word processing. This report includes a table that uses a clear, logical layout to show results from an experiment. These results indicate how material properties vary with temperature. The report is read by a designer, who uses the results to create a product that works in a range of different operating temperatures.

— the data being computer processable (machine readable) rather than just being for a person to read and understand.

EXAMPLE 2 A research scientist uses a database system to store the results of experiments on a material. This system controls the format of different values in the data set. The system generates an output file of digital data. This file is processed by a software application for engineering analysis. The application determines the optimum geometry when using the material to make a product.

ISO 9000 explains that quality is not an abstract concept of absolute perfection. Quality is actually the conformance of characteristics to requirements. This actuality means that any item of data can be of high quality for one purpose but not for a different purpose. The quality is different because the requirements are different between the two purposes.

EXAMPLE 3 Time data are processed by calendar applications and also by control systems for propulsion units on spacecraft. These data include start times for meetings in a calendar application and activation times in a control system. These start times require less precision than the activation times.

[1] https://sdgs.un.org/goals

[2] This document defines information as "knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning".

The nature of digital data is fundamental to establishing requirements that are relevant to the specific decisions made by an organization.

EXAMPLE 4 ISO 8000-8 identifies that data have syntactic (format), semantic (meaning) and pragmatic (usefulness) characteristics.

To support the delivery of high-quality data, the ISO 8000 series addresses:

— data governance, data quality management and maturity assessment;

EXAMPLE 5 ISO 8000-61 specifies a process reference model for data quality management.

— creating and applying requirements for data and information;

EXAMPLE 6 ISO 8000-110 specifies how to exchange characteristic data that are master data.

— monitoring and measuring information and data quality;

EXAMPLE 7 ISO 8000-8 specifies approaches to measuring information and data quality.

— improving data and, consequently, information quality;

EXAMPLE 8 ISO/TS 8000-81 specifies an approach to data profiling, which identifies opportunities to improve data quality.

— issues that are specific to the type of content in a data set.

EXAMPLE 9 ISO/TS 8000-311 specifies how to address quality considerations for product shape data.

Data quality management covers all aspects of data processing, including creating, collecting, storing, maintaining, transferring, exploiting and presenting data to deliver information.

Effective data quality management is systemic and systematic, requiring an understanding of the root causes of data quality issues. This understanding is the basis for not just correcting existing nonconformities but also implementing solutions that prevent future reoccurrence of those nonconformities.

EXAMPLE 10 If a data set includes dates in multiple formats including "yyyy-mm-dd", "mm-dd-yy" and "dd-mm-yy", then data cleansing can correct the consistency of the values. Such cleansing requires additional information, however, to resolve ambiguous entries (such as, "04-05-20"). The cleansing also cannot address any process issues and people issues, including training, that have caused the inconsistency.

As a contribution to this overall capability of the ISO 8000 series, this document specifies the single, common vocabulary for the ISO 8000 series. This vocabulary is ideal reading material by which to understand the overall subject matter of data quality. This document presents the vocabulary structured by a series of topic areas (for example, terms relating to quality and terms relating to data and information).

Organizations can use this document on its own or in conjunction with other parts of the ISO 8000 series.

This document supports activities that affect:

— one or more information systems;

— data flows within the organization and with external organizations;

— any phase of the data life cycle.

Deleted: 6 © ISO 2022 – All rights reserved¶

By implementing parts of the ISO 8000 series to improve organizational performance, an organization achieves the following benefits:

— objective validation of the foundations for digital transformation of the organization;

— a sustainable basis for data in digital form becoming a fundamental asset class the organization relies on to deliver value;

— securing evidence-based trust from other parties (including supply chain partners and regulators) about the repeatability and reliability of data and information processing in the organization;

— portability of data with resulting protection against loss of intellectual property and reusability across the organization and applications;

— effective and efficient interoperability between all parties in a supply chain to achieve traceability of data back to original sources;

— readiness to acquire or supply services where the other party expects to work with common understanding of explicit data requirements.

ISO 8000-1 provides a detailed explanation of the structure and scope of the whole ISO 8000 series.

ISO has identified this document, ISO 8000-1 and ISO 8000-8 as horizontal deliverables[3].

Annex A contains an identifier that conforms to ISO/IEC 8824-1. The identifier unambiguously identifies this document in an open information system.

[3] A horizontal deliverable is a deliverable dealing with a subject relevant to a number of committees or sectors or of crucial importance to ensure coherence across standardization deliverables.

Deleted: © ISO 2022 – All rights reserved 7¶

Data quality — Part 2: Vocabulary

1 Scope

This document defines terms relating to data quality. These terms are used by the ISO 8000 series of standards.

Deleted: parts

2 Normative references

There are no normative references in this document.

3 Terms and definitions

ISO and IEC maintain terminology databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

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

Deleted: terminological

Field Code Changed

Deleted: http

3.1 Terms relating to quality

3.1.1
process
set of interrelated or interacting activities that use inputs to deliver an intended result

Deleted: , noun

[SOURCE: ISO 9000:2015, 3.4.1, modified — Notes to entry have been removed.]

3.1.2
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.]

3.1.3
quality
degree to which a set of inherent characteristics of an object fulfils *requirements* (3.1.2)

Note 1 to entry: The term "quality" can be used with adjectives such as poor, good or excellent.

Note 2 to entry: "Inherent", as opposed to "assigned", means existing in the object.

[SOURCE: ISO 9000:2015, 3.6.2]

3.1.4
quality management system
part of a management system with regard to *quality* (3.1.3)

[SOURCE: ISO 9000:2015, 3.5.4]

3.1.5
nonconformity
non-fulfilment of a *requirement* (3.1.2)

[SOURCE: ISO 9000:2015, 3.6.9, modified — Note 1 to entry has been removed.]

3.1.6
defect
nonconformity (3.1.5) related to an intended or specified use

Note 1 to entry: The distinction between the concepts defect and nonconformity is important as it has legal connotations, particularly those associated with *product* (3.5.2) and service liability issues.

Note 2 to entry: The intended use as intended by the customer can be affected by the nature of the *information* (3.2.1), such as operating or maintenance instructions, provided by the provider.

[SOURCE: ISO 9000:2015, 3.6.10]

3.1.7
quality management
management with regard to *quality* (3.1.3)

Note 1 to entry: Quality management can include establishing quality policies and quality objectives, and *processes* (3.1.1) to achieve these quality objectives through *quality planning* (3.1.8), *quality control* (3.1.9), *quality assurance* (3.1.10) and *quality improvement* (3.1.11).

[SOURCE: ISO 9000:2015, 3.3.4]

3.1.8
quality planning
part of *quality management* (3.1.7) focused on setting *quality* (3.1.3) objectives and specifying necessary operational *processes* (3.1.1) and related resources to achieve the quality objectives

Note 1 to entry: Establishing quality plans can be part of quality planning.

[SOURCE: ISO 9000:2015, 3.3.5]

3.1.9
quality control
part of *quality management* (3.1.7) focused on fulfilling *quality* (3.1.3) *requirements* (3.1.2)

[SOURCE: ISO 9000:2015, 3.3.7]

3.1.10
quality assurance
part of *quality management* (3.1.7) focused on providing confidence that *quality* (3.1.3) *requirements* (3.1.2) will be fulfilled

[SOURCE: ISO 9000:2015, 3.3.6]

3.1.11
quality improvement
part of *quality management* (3.1.7) focused on increasing the ability to fulfil *quality* (3.1.3) *requirements* (3.1.2)

Note 1 to entry: The quality requirements can be related to any aspect such as effectiveness, efficiency or traceability.

[SOURCE: ISO 9000:2015, 3.3.8]

3.1.12
inspection
determination of conformity to specified *requirements* (3.1.2)

[SOURCE: ISO 9000:2015, 3.11.7, modified — Notes to entry have been removed.]

3.2 Terms relating to data and information

3.2.1
information
knowledge concerning objects, such as facts, events, things, *processes* (3.1.1), 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.]

3.2.2
data
reinterpretable representation of *information* (3.2.1) in a formalized manner suitable for communication, interpretation, or processing

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

3.2.3
data exchange
storing, accessing, transferring, and archiving of *data* (3.2.2)

[SOURCE: ISO 10303-1:2021, 3.1.31]

3.2.4
data set
logically meaningful grouping of *data* (3.2.2)

EXAMPLE 1 Computer-aided design (CAD) files.

EXAMPLE 2 Electronic data interchange (EDI) transactions.

3.2.5
metadata
data (3.2.2) defining and describing other data

[SOURCE: ISO/IEC 11179-1:2015, 3.2.16, modified — The words “that defines and describes” have been replaced with “defining and describing”.]

3.2.6
objective evidence
data (3.2.2) supporting the existence or verity of something

Note 1 to entry: Objective evidence can be obtained through observing, *measuring* (3.4.1), testing or other means.

[SOURCE: ISO 9000:2015, 3.8.3, modified — Note 1 to entry has been modified and Note 2 to entry has been removed.]

3.2.7
data element
unit of *data* (3.2.2) that is considered in context to be indivisible

Note 1 to entry: The definition states that a data element is "indivisible" in some context. This means it is possible that a data element considered indivisible in one context (e.g. telephone number) can be divisible in another context (e.g. country code, area code, local number).

[SOURCE: ISO/IEC 11179-1:2015, 3.3.8, modified — The abbreviated term "DE" has been removed and the word "may" has been replaced by "can" in Note 1 to entry.]

3.2.8
value domain
set of permissible values

Note 1 to entry: The permissible values in a value domain can either be enumerated or expressed via a description.

[SOURCE: ISO/IEC 11179-1:2015, 3.3.31, modified — The abbreviated term "VD" has been removed and the word "may" has been replaced by "can" in Note 1 to entry.]

3.2.9
data element concept
concept that is an association of a property with an object class

Note 1 to entry: A data element concept is implicitly associated with both the property and the object class whose combination it expresses.

Note 2 to entry: A data element concept can also be associated with zero or more conceptual domains, each of which expresses its value meanings.

Note 3 to entry: A data element concept can also be associated with zero or more *data elements* (3.2.7) each, of which provides representation for the data element concept via its associated *value domain* (3.2.8).

[SOURCE: ISO/IEC 11179-1:2015, 3.3.9, modified — The abbreviated term "DEC" has been removed and the word "may" has been replaced by "can" in Notes 2 and 3 to entry.]

3.3 Terms relating to identifier

3.3.1
identifier
string of characters created by an organization to reference a *data set* (3.2.4)

3.3.2
identifier resolution
process (3.1.1) that, when applied to an *identifier* (3.3.1), returns an associated *data set* (3.2.4)

3.3.3
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.]