INTERNATIONAL STANDARD

ISO 8000-1

First edition
2022-04

Data quality —

Part 1: Overview

Qualité des données —

Partie 1: Aperçu

Reference number
ISO 8000-1:2022(E)

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

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 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 first edition of ISO 8000-1 cancels and replaces ISO/TS 8000-1:2011, which has been technically revised.

The main changes are as follows:

— updates to cover the complete published parts in the ISO 8000 series;

— general editorial changes.

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.

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.

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.

1) https://sdgs.un.org/goals

2) ISO 8000-2 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”.

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

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.

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 8000-2[3] specifies the single, common vocabulary for the ISO 8000 series. This vocabulary is a foundation for understanding the overall subject matter of data quality. ISO 8000-2 presents the vocabulary structured by a series of topic areas (for example, terms relating to quality and terms relating to data and information).

ISO has identified this document, ISO 8000-2 and ISO 8000-8 as horizontal deliverables[4].

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

3) The content is available on the ISO Online Browsing Platform. https://www.iso.org/obp

4) Deliverable dealing with a subject relevant to a number of committees or sectors or of crucial importance to ensure coherence across standardization deliverables.

Data quality —

Part 1: Overview

1 Scope

This document provides an overview of the ISO 8000 series.

The following are within the scope of this document:

— stating the scope of the ISO 8000 series as a whole;

— establishing the principles of information and data quality;

— describing the path to data quality;

— describing the structure of the ISO 8000 series;

— providing a summary of the content of each part in the ISO 8000 series;

— establishing the relationship of the ISO 8000 series to other international standards.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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 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 apply.

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/

4 Scope of the ISO 8000 series

The ISO 8000 series provides frameworks for improving data quality for specific kinds of data. The series defines which characteristics of data are relevant to data quality, specifies requirements applicable to those characteristics, and provides guidelines for improving data quality. The series is applicable within all stages of the data life cycle.

NOTE The ISO 8000 series can be used either in conjunction with or independently of standards for quality management systems.

The following are within the scope of the ISO 8000 series:

— general aspects of data quality, including principles, vocabulary and measurement of information and data quality;

— data governance;

— data quality management, including processes, roles, responsibilities and maturity assessment;

— data quality assessment, including profiling and data rules;

— quality of master data, including exchange of characteristic data and identifiers;

— quality of industrial data, including product shape data.

The following are outside the scope of the ISO 8000 series:

— quality of the things represented by data;

EXAMPLE 1 AF Industries makes fasteners. AF publishes an electronic catalogue of its products. The quality of the catalogue (the data) is within the scope of the ISO 8000 series. The quality of the fasteners (the things represented by the data) is outside of the scope of the ISO 8000 series.

— quality management principles;

EXAMPLE 2 ISO 9000 identifies eight quality management principles: customer focus, leadership, engagement of people, process approach, improvement, evidence-based decision making and relationship management.

— systems and software product quality.

EXAMPLE 3 ISO/IEC 25000, ISO/IEC 25010, ISO/IEC 25012 and ISO/IEC 25024 address systems and software product quality requirements and evaluation.

The ISO 8000 series contains requirements that are intended to be applicable to:

— all organizations, regardless of type and size;

— organizations at each point in the data supply chain.

5 Principles of information and data quality

The following principles of information and data quality underlie the ISO 8000 series:

— data are the reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing;

— an agreed level of data quality enables the right people to make the right decisions at the right time;

NOTE 1 Making the right decisions validates that the data are fit for purpose. Decision making can depend on many different characteristics including location and sequence.

NOTE 2 Agreement is necessary between the stakeholders participating in and affected by the decision.

— effective data quality management builds on the fundamental concepts and principles of ISO 9000;

— data quality is a function of the inherent characteristics of the data under consideration;

EXAMPLE 1 A data set contains only characters from the character set specified by ISO/IEC 10646. This fact is an inherent characteristic of the data.

EXAMPLE 2 One item in a data set has a length of 13 (i.e. is represented by 13 characters). This length is an inherent characteristic of the data.

EXAMPLE 3 One column in a database table contains only values that are members of the set {"cm", "inches", "mm", "m"}. This fact is an inherent characteristic of the data.

EXAMPLE 4 ISO/IEC 27001 addresses confidentiality, integrity and availability of data as the primary considerations for management of information security. Confidentially, integrity and availability are functions of the whole system (including hardware, software and people) and, thus, are not inherent characteristics of the data.

— organizations gain greater value when digital data are computer processable;

— sustained co-ordinated activity by an organization protects and realises the value of data as an asset.

NOTE 3 This activity requires levels of investment that are appropriate to the value delivered by the data. The co-ordination addresses the influence that everyone in the organization has on the quality and value of data.

6 The ISO 8000 series path to data quality

The ISO 8000 series specifies a comprehensive capability to achieve and sustain high quality data. This capability (see Figure 1):

— builds on the principles of data quality (see Clause 5);

— responds to the information need arising from the decision-making requirements of an organization;

— consists of processes that deliver effective data quality management and achieve governance and assurance;

NOTE 1 These processes follow the fundamental structure of the Plan-Do-Check-Act cycle as also adopted by ISO 9001.

— makes use of data specifications to address information need and to establish a rigorous basis for determining that data meet applicable requirements in respect of syntactic, semantic and pragmatic considerations;

— supports verification and validation of conformance to requirements;

— ensures alignment with the strategic objectives of the organization;

— delivers sustainable organizational and behavioural changes to improve data quality;

— is applicable to any type of data.

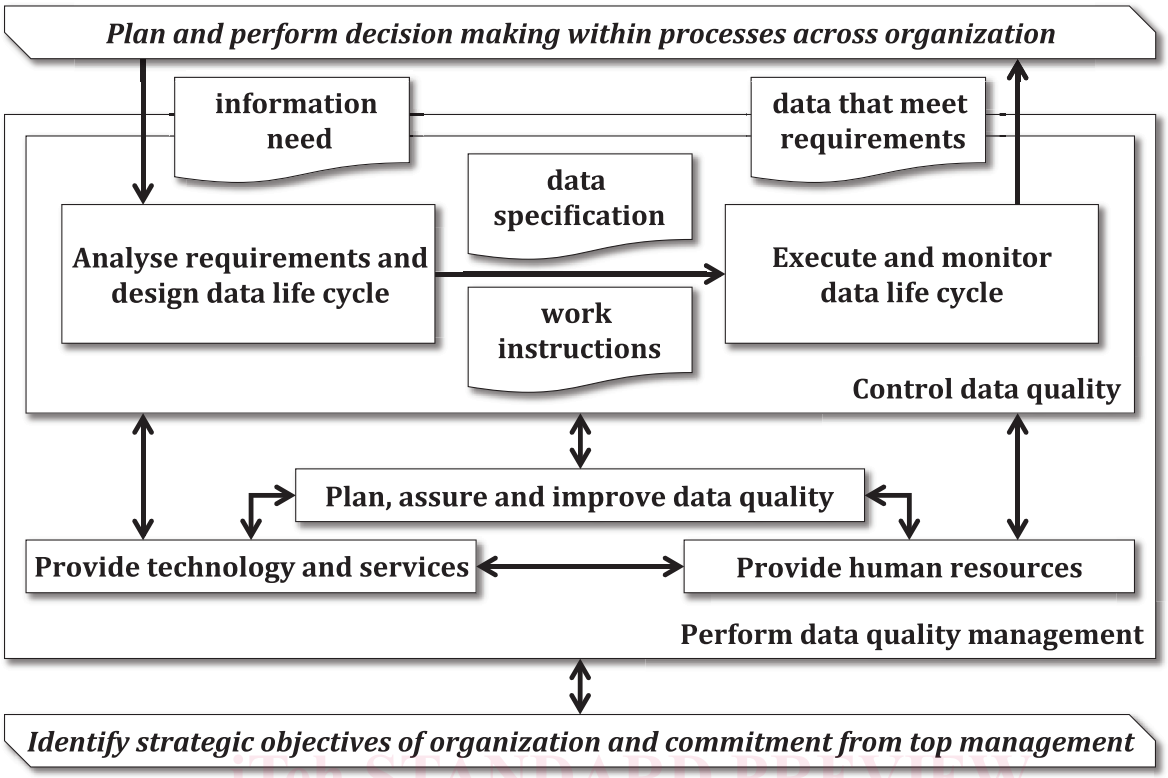

Figure 1 — ISO 8000 series capability to achieve and sustain high quality data

Although the ISO 8000 series is applicable to any type of data, the series does address some of the different considerations that distinguish between types. These considerations include the:

— role of the data;

EXAMPLE 1 Roles include, but are not limited to: master data; transaction data; reference data.

EXAMPLE 2 The role of master data is to describe the entities that are both independent and fundamental for an organization. The data are referenced in order to perform transactions. The entities include customers, products, employees, materials, suppliers, services, shareholders, facilities, equipment, rules and regulations.

— component within the data architecture;

EXAMPLE 3 Components include, but are not limited to: dictionary; exchange file; database, data schema.

EXAMPLE 4 A dictionary consists of data that uniquely identify and describe entities such as part classifications, types of property and units of measurement. This dictionary is a component of the data architecture within an organization. This component enables consistent understanding of other data managed by the organization.

— scope of information represented by the data;

EXAMPLE 5 Scope includes, but is not limited to: products (for example, product shape data); projects; financial records; identifiers; performance characteristics; physical characteristics; safety. These scopes are not necessarily distinct from each other.

EXAMPLE 6 The subject matter of product data is all the different aspects that determine the functions, performance and how to conduct activities on the product (including creation, maintenance and disposal).

— intended area of application of the data;

EXAMPLE 7 Areas of application include, but are not limited to: industry; healthcare; banking.