International Standard

ISO 8000-114

First edition
2024-03

Data quality —

Part 114: Master data: Application of ISO/IEC 21778 and ISO 8000-115 to portable data

Reference number
ISO 8000-114:2024(en)

COPYRIGHT PROTECTED DOCUMENT

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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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*, SC 4, *Industrial data*.

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

0.1 Foundations of the ISO 8000 series

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

— operational effectiveness and efficiency;

— safety and security;

— 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. This 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) 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”.

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

EXAMPLE 4 ISO 8000-1 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 for 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.

0.2 Understanding more about the ISO 8000 series

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

ISO 8000-2[3] 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. 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 ISO 8000-1, ISO 8000-2 and ISO 8000-8 as horizontal deliverables[4].

0.3 Role of this document

As a contribution to this overall capability of the ISO 8000 series, this document specifies requirements that make data portable. These requirements support the data to represent unambiguously the information that organizations use. Portable data are important to organizations because these data can exist independently of the application that created the data. This independence ensures other applications can use the information represented by the data. Portable data are necessary for interoperability of systems and for the protection of intellectual property in data.

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.

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.

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

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

0.4 Benefits of the ISO 8000 series

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.

Data quality —

Part 114:
Master data: Application of ISO/IEC 21778 and ISO 8000-115 to portable data

1 Scope

This document specifies requirements for data to be portable data. The following are within the scope of this document:

— fundamental principles and assumptions for portable data;

— the role of semantic encoding in making data portable;

— the syntax of a structured format for portable data;

— the metadata and the reference data to include as part of portable data.

The following are outside the scope of this document:

— the information to be represented by the interoperable data format.

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*

ISO 8000-115, *Data quality — Part 115: Master data: Exchange of quality identifiers: Syntactic, semantic and resolution requirements*

ISO/IEC 10646, *Information technology — Universal coded character set (UCS)*

ISO/IEC 21778, *Information technology — The JSON data interchange syntax*

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 Fundamental principles and assumptions

4.1 General

Application software is a kind of information. This software consists of instructions to a computer system to perform a specific function. Application software is different from data, which serve as the input to such software.

For data to be portable, the characteristics of the data include the following:

— being identifiable by an explicit identifier;

EXAMPLE A filename is an identifier.

— representing information using characters that conform to a stated character encoding;

EXAMPLE ISO/IEC 10646 specifies an encoding system that applies to representation of the written form of the languages of the world as well as additional symbols.

— having a stated syntax;

— having explicit meaning, where either the metadata and reference data are accessible for the data or the data are semantically encoded using an open concept dictionary.

EXAMPLE The ISO 22745 series specifies requirements for open concept dictionaries that are technical dictionaries.

4.2 Filename

In digital computers, information is represented as processable data in binary form, which consists of a collection of bits of ones and zeros (on and off). Such data can be managed and exchanged by computers when the data are in a file, which has a filename.

Application software also has a filename when stored by a computer.

The general naming conventions for an individual file is a base file name and an optional file extension. The name and extension are separated by a period.

Different file systems have specific rules for filename length and allowable characters. Each system also has differing rules about naming the location, which involves the formation of individual components in a path to a directory or file to create a namespace where filenames are unique.

4.3 Character encoding

When representing information as data, character encoding standards specify how to convert human-readable characters to computer-processable binary form.

ISO/IEC 10646 specifies the Universal Coded Character Set and provides three allowable encoding forms. Of these encoding forms, UTF-8 is the most prevalent on the internet.

4.4 Syntax

The general syntax rules for data are fundamentally simple. These rules include requiring values to have an accompanying label. This labelling results in a tuple, which is the smallest meaningful construct of data to represent information.

NOTE 1 The typical names for such tuples are “attribute-value pair”, “property-value pair” or “name-value pair”.

Similar to how in certain languages, a valid sentence has at least a verb and a noun, meaningful data are always the combination of a named attribute and a named value.

EXAMPLE 1 `20201001` is a meaningless value on its own.

EXAMPLE 2 `Date = "20201001"` is a meaningful attribute-value pair, given that the true meaning requires an implicit understanding that the date is in ISO 8601-1 format.

EXAMPLE 3 `Date = "First day of the month of October of the year 2020"` is also a meaningful attribute-value pair that only requires the implicit understanding of the meaning of the terms “First”, “day”, “month” “October” and “year”. These meanings appear in a dictionary.

ISO/IEC 21778 specifies JSON as a lightweight, text-based, language-independent syntax for defining data interchange formats. This syntax is suitable for data consisting of attribute-value pairs including values that form an array.

NOTE 2 In this case, “language-independent” means independent of any particular computer programming language.

NOTE 3 By being language independent, JSON is suitable for moving data between systems that cover a wide range of different implementation technologies.

4.5 Semantic encoding

Naming an attribute and providing the corresponding value is necessary but not sufficient to represent information as meaningful data. In order to use such information, the decoder must be able to reliably convert such data back to original content. This conversion can only be done if the meanings of the attributes and of the coded values are available to the decoder.

Words or abbreviations are potentially ambiguous attribute names, while coded values can be confusing.

EXAMPLE 1 The value “CA” can represent the ISO 3166 series code for the country of Canada, the state code for the State of California or be an abbreviation of “cost analysis” or “cardiac arrest”.

A metadata registry is a list of the attribute names and their corresponding meanings. Such registries rarely, however, contain code lists.

EXAMPLE 2 The ISO/IEC 11179 series specifies requirements for metadata registries.

A vocabulary, glossary or dictionary contains both metadata and reference data or code lists.

A concept dictionary explicitly describes and also, by using a concept identifier, identifies each concept. This identifier can be converted by semantic decoding to generate the corresponding attribute name or code value.

Attribute names and code values are concepts described by a dictionary. These concepts can be unambiguously identified by an identifier that conforms to ISO 8000-115. Such identifiers provide the link to the corresponding name and description in a dictionary.

By requiring explicit association of the domain and the owner of the domain with each identifier, ISO 8000-115 establishes how to create globally unique dictionary concept identifiers. This uniqueness ensures the consistent resolution of each identifier to a name and a description of the corresponding concept.

The owner of the domain and the domain are the two main elements that form a prefix for the identifier in a concept dictionary. This prefix also includes a full stop character “.” (UTF-8 character 2E, see ISO/IEC 10646) between the two elements.

Examples of semantic encoding are given in Table 1. In these examples, the domain is represented by the short name for the concept dictionary and the owner of the domain by the short name for the legal owner of the dictionary.

NOTE The content of prefixes is agreed by the organizations who are intending to exchange the identifiers. This exchange will only be successful if there is no duplication of identifiers. The individual prefixes in these examples are not specified by any standard.

Table 1 — Examples of semantic encodings in different concept dictionaries

Item	Example concept dictionary	
	ECCMA[a] Open Technical dictionary (eOTD)	IEC 61360-4 – IEC/SC 3D – Common Data Dictionary (CDD)[14]
Legal owner short name	ECCMA	IEC
Concept dictionary short name	eOTD	CDD
Dictionary concept identifier in accordance with ISO 8000-115	ECCMA.eOTD:061-1#01-068756#1	IEC.CDD:0112/2///61360_4#AAA002#002
[a] Electronic Commerce Code Management Association.		

5 Structured format for portable data

To be portable, data shall:

a) be stored by a computer as a file that has a base file name and the file extension ".idf";

b) consist of characters represented using the UTF-8 encoding form in accordance with ISO/IEC 10646;

c) be a valid JSON text in accordance with ISO/IEC 21778;

d) have metadata that are available in a concept dictionary where identifiers have a format in accordance with ISO 8000-115;

e) have reference data that are available in a concept dictionary where identifiers have a format in accordance with ISO 8000-115.

6 Conformance

Data conform to this document when 5 a), 5 b), 5 c), 5 d) and 5 e) are met.

NOTE Annex B provides some examples of portable data.

Annex A
(informative)

Document identification

To provide for unambiguous identification of an information object in an open system, the following object identifier is assigned to this document. The meaning of this value is defined in ISO 10303-1.

```
{ iso standard 8000 part(114) version(1) }
```

Annex B
(informative)

Examples of portable data

B.1 General

This annex provides simple examples to show alternative forms of data to represent the following information:

— a party having the legal name “Code Management Association”;

— the party having the trade names (or DBA, doing business as) “ECCMA” and “Electronic Commerce Code Management Association”;

— the party having the date of formation “1999-04-20”;

— the party having the country and state of incorporation of “USA” and “Delaware”, respectively.

The alternative forms of data are:

— valid JSON text but without semantic encoding, which is not portable (see B.2);

— valid JSON text with only semantic encoding, which is portable (see B.3);

— valid JSON text with semantic encoding and accompanying text labels, which is portable (see B.4).

B.2 JSON without semantic encoding

The following data represent the example information (see B.1) as valid JSON text but without semantic encoding. The data contains keys and values in human-readable form. These data are not portable.

```
{
  "PARTY":{
    "LEGAL NAME": "CODE MANAGEMENT ASSOCIATION",
    "TRADE NAME(S) (DBA)": "ECCMA, ELECTRONIC COMMERCE CODE MANAGEMENT ASSOCIATION",
    "DATE OF FORMATION": "1999-04-20",
    "COUNTRY OF INCORPORATION": "USA",
    "STATE OF INCORPORATION": "DELAWARE"
  }
}
```

B.3 JSON with only semantic encoding

The following data represent the example information (see B.1) as valid JSON text with semantic encoding. The data contains keys and values in semantic-encoded form (i.e. using concept identifiers from a concept dictionary). These data are portable.

The applicable data specification determines the format of each value in the data, where the format can be either semantic encoded (i.e. a concept identifier that conforms to ISO 8000-115) or a primitive data type (e.g. a text string or a date).

Any organization receiving the data is only able to decode the semantics of the data if a relevant concept dictionary is available to the organization.

In this case, the originating concept dictionary is the ECCMA Open Technical Dictionary.