INTERNATIONAL STANDARD

ISO/IEC 11179-4

Second edition
2004-07-15

Information technology — Metadata registries (MDR) —

Part 4: Formulation of data definitions

Technologies de l'information — Registres de métadonnées (RM) —

Partie 4: Formulation des définitions de données

Reference number
ISO/IEC 11179-4:2004(E)

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
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 11179-4 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange*.

This second edition cancels and replaces the first edition (ISO/IEC 11179-4:1995), which has been technically revised.

ISO/IEC 11179 consists of the following parts, under the general title *Information technology — Metadata registries (MDR)*:

— *Part 1: Framework*

— *Part 3: Registry metamodel and basic attributes*

— *Part 4: Formulation of data definitions*

— *Part 5: Naming and identification principles*

— *Part 6: Registration*

The following part is under preparation:

— *Part 2: Classification*

Introduction

Data processing and electronic data interchange rely heavily on accurate, reliable, controllable and verifiable data recorded in databases. A prerequisite for correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and representation of the data. To facilitate this common understanding, a number of characteristics, or attributes of the data have to be defined. These characteristics of data are known as “metadata”, that is, “data that describes data”. This part of ISO/IEC 11179 specifies requirements and recommendations on the formulation of data definitions that are specified in Metadata Registries. The purpose of these definitions is to specify, describe, explain, and clarify the meaning of data, to promote the standardization or reuse of data elements, and to promote data sharing and integration of information systems.

The structure of a *Metadata Registry* is specified in the form of a conceptual data model. The *Metadata Registry* is used to keep information about data elements and associated concepts, such as “data element concepts”, “conceptual domains”, and “value domains”. Generically, these are all referred to as “metadata items”. Such metadata are necessary to clearly describe, record, analyse, classify, and administer data.

The definitional requirements and recommendations specified in this part of ISO/IEC 11179 do not always apply to terminological definitions found in glossaries and language dictionaries. Differences exist between the requirements that apply in a language dictionary, and the requirements that apply in a metadata registry. The requirements for ISO/IEC 11179 are more restrictive than those for a natural language dictionary. For example, a language dictionary may have multiple definitions covering multiple senses of a term or word, whereas data definitions are developed for particular contexts and should not have multiple senses within any context. Data definitions are intended to explicate the concept or concepts, which are represented by a collection of data, a data value, a data element, or other metadata item. A single definition may be established as the reference definition, with other definitions asserted to be equivalent (e.g., a definition in one language may be established as a reference definition, with definitions in other languages asserted to be equivalent). Metadata items may have a single preferred definition within a particular context, with other deprecated definitions.

Many data definitions include terms that themselves need to be defined (e.g., “charge”, “allowance”, “delivery”). Some of these terms may have different definitions in different industrial sectors. Therefore, there is a need for most metadata registries to establish an associated *glossary or terminology reference* of terms used in the definitions.

Information technology — Metadata registries (MDR) —

Part 4:
Formulation of data definitions

1 Scope

This part of ISO/IEC 11179 specifies requirements and recommendations for constructing definitions for data and metadata. Only semantic aspects of definitions are addressed; specifications for formatting the definitions are deemed unnecessary for the purposes of ISO/IEC 11179. While especially applicable to the content of metadata registries as specified in ISO/IEC 11179-3, this part of ISO/IEC 11179 is useful broadly for developing definitions for data and metadata.

These definitional requirements and recommendations pertain to formulating definitions for data elements and other types of data constructs such as entity types, entities, relationships, attributes, object types (or classes), objects, composites, code entries, metadata items, and the data referred to by XML tags.

2 Conformance

This part of ISO/IEC 11179 may be used independently, e.g., for defining data elements outside the context of a metadata registry. In such cases, compliance may be claimed if the requirements and recommendations have been followed in developing the definitions.

Where used in the context of an ISO/IEC 11179 metadata registry, this part of ISO/IEC 11179 shall constitute the criteria for definitions when establishing the registration status as specified in ISO/IEC 11179-6.

3 Terms and definitions

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

3.1
Conceptual Domain
set of valid value meanings

NOTE The value meanings may either be enumerated or expressed via a description.

[ISO/IEC 11179-3:2003, 3.3.21]

3.2
Concept
unit of knowledge created by a unique combination of characteristics

[ISO 1087-1:2000, 3.2.1]

3.3
data
re-interpretable representation of information in a formalized manner suitable for communication, interpretation or processing

NOTE Data can be processed by human or automatic means.

[ISO/IEC 2382-1:1993, 01.01.02]

3.4
data element
unit of **data** for which the definition, identification, representation and permissible values are specified by means of a set of attributes

[ISO/IEC 11179-3:2003, 3.3.36]

3.5
Data Element Concept
concept that can be represented in the form of a data element, described independently of any particular representation

[ISO/IEC 11179-3:2003, 3.3.38]

3.6
definition
representation of a concept by a descriptive statement which serves to differentiate it from related concepts

[ISO 1087-1:2000, 3.3.1]

3.7
metadata
data that defines and describes other data

[ISO 1087-1:2000, 3.2.18]

3.8
Metadata Registry
information system for registering metadata

[ISO/IEC 11179-3:2003, 3.2.22]

3.9
metadata item
instance of a **metadata object**

NOTE 1 In all parts of ISO/IEC 11179, this term is applied only to instances of metadata objects described by the metamodel in Clause 4 of ISO/IEC 11179-3:2003. Examples include instances of data elements, data element concepts, permissible values, etc.

NOTE 2 A metadata item has associated attributes, as appropriate for the metadata object it instantiates.

[ISO/IEC 11179-3:2003, 3.2.19]

3.10
metadata object
object type defined by a metamodel

NOTE In all parts of ISO/IEC 11179, this term is applied only to metadata objects described by the metamodel in Clause 4 of ISO/IEC 11179-3:2003. Examples include data elements, data element concepts, permissible values etc. See Clause 3.3 of ISO/IEC 11179-3:2003 for a complete list.

[ISO/IEC 11179-3:2003, 3.2.20]

3.11
name
designation of an object by a linguistic expression

[ISO/IEC 11179-3:2003, 3.2.26]

3.12
Value Domain
set of permissible values

NOTE 1 The value domain provides representation, but has no implication as to what data element concept the values may be associated with nor what the values mean.

NOTE 2 The permissible values may either be enumerated or expressed via a description.

[ISO/IEC 11179-3:2003, 3.3.140]

4 Summary of data definition requirements and recommendations

A listing of the requirements and recommendations without explanations is provided in this clause for convenience of the user. The intent is to facilitate ease of use of this document once an understanding of the requirements and recommendations is achieved. Clause 5 describes each requirement and recommendation with an explanation and examples to ensure their exact meaning is understood.

4.1 Requirements

A data definition shall:

a) be stated in the singular

b) state what the concept is, not only what it is not

c) be stated as a descriptive phrase or sentence(s)

d) contain only commonly understood abbreviations

e) be expressed without embedding definitions of other data or underlying concepts

4.2 Recommendations

A data definition should:

a) state the essential meaning of the concept

b) be precise and unambiguous

c) be concise

d) be able to stand alone

e) be expressed without embedding rationale, functional usage, or procedural information

f) avoid circular reasoning

g) use the same terminology and consistent logical structure for related definitions

h) be appropriate for the type of metadata item being defined

5 Provisions

5.1 Premises

Data is used for specific purposes. Differences in use require different operational manifestations of some requirements and recommendations. For example, different levels of specificity for data definitions are generally required in different contexts. Recommendation 5.3.a) below, provides an example of this need for varying levels of specificity for different definitions. The implementation of Recommendation 5.3.a), "state the essential meaning of the concept" is context dependent. The primary characteristics deemed necessary to convey the essential meaning of a particular definition will vary according to the level of generalization or specialization of the data. Primary and essential characteristics for defining concepts such as "airport" in the commercial air transportation industry might be specific, where a more general definition may be adequate in a different context. Within a metadata registry, multiple equivalent definitions may be written in different languages or, within a single language, for different audiences such as children, general public, or subject area specialists. For a discussion of relationships between concepts in different contexts and how characteristics are used to differentiate concepts, see ISO 704, Clause 5. Definitions should be written to facilitate understanding by any user and by recipients of shared data.

5.2 Requirements

To facilitate understanding of the requirements for construction of well-formed data definitions, explanations and examples are provided below. Each requirement is followed by a short explanation of its meaning. Examples are given to support the explanations. In all cases, a good example is provided to exemplify the explanation. When deemed beneficial, a poor, but commonly used example is given to show how a definition should **NOT** be constructed. To further explain the differences between the good and poor examples, examples are followed by a statement of rationale behind them. Note that the examples below are definitions for data elements and these definitions are illustrative.

A data definition shall:

a) **be stated in the singular**

EXPLANATION - The concept expressed by the data definition shall be expressed in the singular. (An exception is made if the concept itself is plural.)

EXAMPLE - "Article Number"

1) good definition: A reference number that identifies an article.

2) poor definition: Reference number identifying articles.

REASON - The poor definition uses the plural word "articles," which is ambiguous, since it could imply that an "article number" refers to more than one article.

b) **state what the concept is, not only what it is not**

EXPLANATION - When constructing definitions, the concept cannot be defined exclusively by stating what the concept is **not.**

EXAMPLE - "Freight Cost Amount"

1) good definition: Cost amount incurred by a shipper in moving goods from one place to another.

2) poor definition: Costs which are not related to packing, documentation, loading, unloading, and insurance.

REASON - The poor definition does not specify what is included in the meaning of the data.

c) **be stated as a descriptive phrase or sentence(s)** (in most languages)

EXPLANATION - A phrase is necessary (in most languages) to form a precise definition that includes the essential characteristics of the concept. Simply stating one or more synonym(s) is insufficient. Simply restating the words of the name in a different order is insufficient. If more than a descriptive phrase is needed, use complete, grammatically correct sentences.

EXAMPLE - “Agent Name”

1) good definition: Name of party authorized to act on behalf of another party.

2) poor definition: Representative.

REASON - “Representative” is a near-synonym of the data element name, which is not adequate for a definition.

d) **contain only commonly understood abbreviations**

EXPLANATION - Understanding the meaning of an abbreviation, including acronyms and initialisms, is usually confined to a certain environment. In other environments the same abbreviation can cause misinterpretation or confusion. Therefore, to avoid ambiguity, full words, not abbreviations, shall be used in the definition.

Exceptions to this requirement may be made if an abbreviation is commonly understood such as “i.e.” and “e.g.” or if an abbreviation is more readily understood than the full form of a complex term and has been adopted as a term in its own right such as “radar” standing for “radio detecting and ranging.”

All acronyms must be expanded on the first occurrence.

EXAMPLE 1 - “Tide Height”

1) good definition: The vertical distance from mean sea level (MSL) to a specific tide level.

2) poor definition: The vertical distance from MSL to a specific tide level.

REASON - The poor definition is unclear because the acronym, MSL, is not commonly understood and some users may need to refer to other sources to determine what it represents. Without the full word, finding the term in a glossary may be difficult or impossible.

EXAMPLE 2 - “Unit of Density Measurement”

1) good definition: The unit employed in measuring the concentration of matter in terms of mass per unit (m.p.u.) volume (e.g., pound per cubic foot; kilogram per cubic meter).

2) poor definition: The unit employed in measuring the concentration of matter in terms of m.p.u. volume (e.g., pound per cubic foot; kilogram per cubic meter).

REASON - m.p.u. is not a common abbreviation, and its meaning may not be understood by some users. The abbreviation should be expanded to full words.

e) **be expressed without embedding definitions of other data or underlying concepts**

EXPLANATION - As shown in the following example, the definition of a second data element or related concept should not appear in the definition proper of the primary data element. Definitions of terms should be provided in an associated glossary. If the second definition is necessary, it may be attached by a note at the end of the primary definition's main text or as a separate entry in the dictionary. Related definitions can be accessed through relational attributes (e.g., cross-reference).

EXAMPLE 1- “Sample Type Code”

1) good definition: A code identifying the kind of sample.

2) poor definition: A code identifying the kind of sample collected. A sample is a small specimen taken for testing. It can be either an actual sample for testing, or a quality control surrogate sample. A quality control sample is a surrogate sample taken to verify results of actual samples.

REASON - The poor definition contains two extraneous definitions embedded in it. They are definitions of “sample” and of “quality control sample.”

EXAMPLE 2 - "Issuing Bank Documentary Credit Number"

1) good definition: Reference number assigned by issuing bank to a documentary credit.

2) poor definition: Reference number assigned by issuing bank to a documentary credit. A documentary credit is a document in which a bank states that it has issued a documentary credit under which the beneficiary is to obtain payment, acceptance, or negotiation on compliance with certain terms and conditions and against presentation of stipulated documents and such drafts as may be specified.

REASON - The poor definition contains a concept definition, which should be included in a glossary.

5.3 Recommendations

A data definition should:

a) **state the essential meaning of the concept**

EXPLANATION - All primary characteristics of the concept represented should appear in the definition at the relevant level of specificity for the context. The inclusion of non-essential characteristics should be avoided. The level of detail necessary is dependent upon the needs of the system user and environment.

EXAMPLE 1 - “Consignment Loading Sequence Number” (Intended context: any form of transportation)

1) good definition: A number indicating the sequence in which consignments are loaded in a means of transport or piece of transport equipment.

2) poor definition: A number indicating the sequence in which consignments are loaded in a truck.

REASON - In the intended context, consignments can be transported by various transportation modes, e.g., trucks, vessels or freight trains. Consignments are not limited to trucks for transport.

EXAMPLE 2 - “Invoice Amount”

1) good definition: Total sum charged on an invoice.

2) poor definition: The total sum of all chargeable items mentioned on an invoice, taking into account deductions on one hand such as allowances and discounts, and additions on the other hand, such as charges for insurance, transport, handling, etc.

REASON - The poor definition includes extraneous material.

b) **be precise and unambiguous**

EXPLANATION -The exact meaning and interpretation of the defined concept should be apparent from the definition. A definition should be clear enough to allow only one possible interpretation.

EXAMPLE - “Shipment Receipt Date”

1) good definition: Date on which a shipment is received by the receiving party.

2) poor definition: Date on which a specific shipment is delivered.

REASON - The poor definition does not specify what determines a “delivery.” “Delivery” could be understood as either the act of unloading a product at the intended destination or the point at which the intended customer actually obtains the product. It is possible that the intended customer never receives the product that has been unloaded at his site or the customer may receive the product days after it was unloaded at the site.