

<b>Final Committee Draft ISO/IEC FCD 11179-4</b>	
Date: <b>2002-11-26</b>	Reference number: ISO/JTC 1/SC <b>32N0909</b>
Supersedes document SC 32N0755	

THIS DOCUMENT IS STILL UNDER STUDY AND SUBJECT TO CHANGE. IT SHOULD NOT BE USED FOR REFERENCE PURPOSES.

ISO/IEC JTC 1/SC 32 Data Management and Interchange  Secretariat: USA (ANSI)	<p>Circulated to P- and O-members, and to technical committees and organizations in liaison for voting (P-members only) by:</p> <p style="text-align: center;"><b>2003-04-13</b></p> <p>Please return all votes and comments in electronic form directly to the SC 32 Secretariat by the due date indicated.</p>
--	--

ISO/IEC FCD 11179-4:200x(E)  Title: Information technology — Metadata Registries (MDR) - Part 4: Formulation of data definitions  Project: 1.32.15.02.04.00
---

<p>Introductory note: The attached document is hereby submitted for a four-month letter ballot to the National Bodies of ISO/IEC JTC 1/SC 32. The ballot starts 2002-12-13; ITTF must have the document two weeks before the balloting can start.</p> <p>Medium: E</p> <p>No. of pages: 15</p>
--

Address Reply to: Douglas Mann, Secretariat, ISO/IEC JTC 1/SC 32, Pacific Northwest National Laboratory, 13600 Angelica Court, Chantilly, VA, 20151-3360, United States of America

Telephone: +1 202-566-2126; Facsimile: +1 202-566-1639; E-mail: [MannD@battelle.org](mailto:MannD@battelle.org)

# ISO/IEC JTC 1/SC 32 N **FCD 11179-4**

Date: 2002-08-01

**ISO/IEC WD 11179-4**

ISO/IEC JTC 1/SC 32/WG FCD

Secretariat: ANSI

## **Information technology — Metadata Registries (MDR) — Part 4: Formulation of data definitions**

*Élément introductif — Élément central — Partie 4: Élément complémentaire*

### **Warning**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Document type: International Standard  
Document subtype:  
Document stage: (20) Preparatory  
Document language: E

### Copyright notice

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:

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.ch](mailto:copyright@iso.ch)  
Web [www.iso.ch](http://www.iso.ch)

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

**Contents**

Page

**Information technology — Metadata Registries (MDR) — Part 4: Formulation of data definitions**

<b>Foreword</b> .....	<b>iv</b>
<b>Introduction</b> .....	<b>v</b>
<b>1 Scope</b> .....	<b>1</b>
<b>2 Normative references</b> .....	<b>1</b>
<b>3 Terms and definitions</b> .....	<b>2</b>
<b>4 Summary of Data Definition Rules and Guidelines</b> .....	<b>2</b>
<b>4.1 Rules</b> .....	<b>2</b>
<b>4.2 Guidelines</b> .....	<b>3</b>
<b>5 Requirements</b> .....	<b>3</b>
<b>5.1 Premises</b> .....	<b>3</b>
<b>5.2 Rules</b> .....	<b>3</b>
<b>5.3 Guidelines (Guiding Principles)</b> .....	<b>5</b>
<b>6 Conformance</b> .....	<b>8</b>

## Foreword

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1. 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.

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.

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 (ISO/IEC 11179-4:1994), clauses and subclauses of which have been technically revised.

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

- *Part 1: Framework*
- *Part 2: Classification*
- *Part 3: Registry metamodel*
- *Part 4: Formulation of data definitions*
- *Part 5: Naming and identification principles*
- *Part 6: Registration*

## 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 provides for guidance on the formulation of data definitions that comprise Metadata Registries. The purpose of these definitions is to define a data element, to describe, to explain, or to clarify its meaning, 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.



# Information technology — Metadata Registries (MDR) — Part 4: Formulation of data definitions

## 1 Scope

This Part of the International Standard specifies rules and guidelines 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 this standard. While especially applicable to the content of metadata registries as specified in Part 3 of this International Standard, this Part is useful broadly for developing definitions for data and metadata.

These definitional rules and guidelines 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.

The definitional rules and guidelines in this Part of the International standard do not always apply to terminological definitions found in glossaries and language dictionaries. Differences exist between the rules that apply in a language dictionary, and the rules that apply in a metadata registry. The rules for this International Standard 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.

## 2 Normative references

The following referenced documents are indispensable for the application 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 704:2000, Principles and methods of terminology

ISO 1087-1:2000 Terminology work - Vocabulary - Part 1: Theory and Application

ISO/IEC 2382-1:1998, Information technology – Vocabulary – Part 1 Fundamental terms

ISO/IEC 11179-3:2002, Information technology – Metadata registries (MDR) – Part 3: Registry metamodel

### 3 Terms and definitions

For the purposes of this Part of this International Standard, the following definitions apply.

#### 3.1

##### **data**

a 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 2382-1:1998 (01.01.02)]

#### 3.2

##### **data element**

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

#### 3.3

##### **definition**

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

[ISO 1087-1:2000 (3.3.1)]

#### 3.4

##### **Metadata Registry**

MDR

an information system for registering metadata

#### 3.5

##### **name**

the **designation** of an object by a linguistic expression

### 4 Summary of Data Definition Rules and Guidelines

A listing of the rules and guidelines 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 rules and guidelines is achieved. Clause 5 describes each rule and guideline with an explanation and examples to ensure their exact meaning is understood.

#### 4.1 Rules

##### **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 Guidelines

### 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, domain information, or procedural information
- f) avoid circular reasoning
- g) use the same terminology and consistent logical structure for related definitions

## 5 Requirements

### 5.1 Premises

Data is used for specific purposes. Differences in use require different operational manifestations of some rules and guidelines. For example, different levels of specificity for data definitions are generally required in different contexts. Guideline 5.3.a) below, provides an example of this need for varying levels of specificity for different definitions. The implementation of Guideline 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 Rules

To facilitate understanding of the rules for construction of well-formed data definitions, explanations and examples are provided below. Each rule 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 rule 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 Guidelines (Guiding Principles)**

**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.

**c) be concise**

EXPLANATION - The definition should be brief and comprehensive. Extraneous qualifying phrases such as "for the purpose of this metadata registry," "terms to be described," shall be avoided.

EXAMPLE - "Character Set Name"

1) good definition: The name given to the set of phonetic or ideographic symbols in which data is encoded.

2) poor definition: The name given to the set of phonetic or ideographic symbols in which data is encoded, for the purpose of this metadata registry, or, as used elsewhere, the capability of systems hardware and software to process data encoded in one or more scripts.

REASON - In the poor definition, all the phrases after "...data is encoded" are extraneous qualifying phrases.

**d) be able to stand alone**

EXPLANATION - The meaning of the concept should be apparent from the definition. Additional explanations or references should not be necessary for understanding the meaning of the definition.

EXAMPLE - "School Location City Name"

1) good definition: Name of the city where a school is situated.

2) poor definition: See "school site".

REASON - The poor definition does not stand alone, it requires the aid of a second definition (school site) to understand the meaning of the first.

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

EXPLANATION - Although they are often necessary, such statements do not belong in the definition proper because they contain information extraneous to the definition. If deemed useful, such expressions may be placed in other metadata attributes (see ISO/IEC 11179, Part 3). It is, however, permissible to add examples after the definition.

1) The rationale for a given definition should not be included as part of the definition (e.g. if a data element uses miles instead of kilometers, the reason should not be indicated in the definition).

2) Functional usage such as: "this data element should not be used for ..." should not be included in the definition proper.

3) Remarks about procedural aspects. For example, "This data element is used in conjunction with data element 'xxx'", should not appear in the definition; instead use "Related data reference" and "Type of relationship" as specified in ISO/IEC 11179, Part 3.

EXAMPLE - "Data Field Label"

1) good definition: Identification of a field in an index, thesaurus, query, database, etc.

2) poor definition: Identification of a field in an index, thesaurus, query, database, etc., which is provided for units of information such as abstracts, columns within tables.

REASON - The poor definition contains remarks about functional usage. This information starting with "which is provided for..." must be excluded from the definition and placed in another attribute, if it is necessary information.

**f) avoid circular reasoning**

EXPLANATION - Two definitions shall not be defined in terms of each other. A definition should not use another concept's definition as its definition. This results in a situation where a concept is defined with the aid of another concept that is, in turn, defined with the aid of the given concept.

EXAMPLE - two data elements with poor definitions:

1) Employee ID Number - Number assigned to an employee.

2) Employee - Person corresponding to the employee ID number.

REASON - Each definition refers to the other for its meaning. The meaning is not given in either definition.

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

EXPLANATION - A common terminology and syntax should be used for similar or associated definitions.

EXAMPLE - The example for rule 5.2.a) above also illustrates this idea. Both definitions pertain to related concepts and therefore have the same logical structure and similar terminology.

- 1) "Goods Dispatch Date" - Date on which goods were dispatched by a given party.
- 2) "Goods Receipt Date" - Date on which goods were received by a given party.

REASON - Using the same terminology and syntax facilitates understanding. Otherwise, users wonder whether some difference is implied by use of synonymous terms and variable syntax.

## **6 Conformance**

This Part of International Standard 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 rules and guidelines have been followed in developing the definitions.

Where used in the context of an International Standard 11179 metadata registry, Part 4 shall constitute the criteria for definitions when establishing the registration status as specified in International Standard 11179, Part 6.