TECHNICAL REPORT

ISO/IEC TR 20943-3

First edition
2004-03-01

Information technology — Procedures for achieving metadata registry content consistency —

Part 3: Value domains

Technologies de l'information — Procédures pour réaliser la consistance du contenu de l'enregistrement des métadonnées —

Partie 3: Domaines de valeur

Reference number
ISO/IEC TR 20943-3:2004(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.

In exceptional circumstances, the joint technical committee may propose the publication of a Technical Report of one of the following types:

— type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;

— type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;

— type 3, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example).

Technical Reports of types 1 and 2 are subject to review within three years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

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 TR 20943-2, which is a Technical Report of type 3, was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 32, *Data management and interchange.*

ISO/IEC 20943 consists of the following parts, under the general title *Information technology — Procedures for achieving metadata registry content consistency*:

— *Part 1: Data elements* [Technical Report]

— *Part 3: Value domains* [Technical Report]

The following parts are under preparation:

— *Part 2: XML structured data*

— *Part 4: Overview*

Introduction

The exchange of metadata between metadata registries based on ISO/IEC 11179, *Information technology — Metadata registries* (all parts), depends not only on registry software that conforms to the standard, but also on metadata contents that are comparable between registries. While the standard has provisions for data specification and registration, there are pragmatic issues pertaining to populating the registries with content. Based on the experiences of organizations that are implementing the standard, technical reports to explore content issues will help current and future users.

Metadata registries can be used to register data elements, value domains, other objects, and associated attributes for many kinds of organizational data resource collections. Metadata registries can store information describing value domains used to specify the allowed values of a data element, the codes in a standard list, and classification schemes.

This technical report is based on ISO/IEC 11179-3:2003 of the six-part ISO/IEC 11179 International Standard that describes the organization of a registry for managing the semantics of data. The standard specifies the structure of a registry in the form of a conceptual model. The conceptual model is not intended to be a logical or physical data model for a computer system.

ISO/IEC 11179-3:2003, models a value domain and an associated conceptual domain. Conceptualization and articulation of rules and relationships are needed in the creation of conceptual domains and value domains. Reuse of value domains should be enabled and regularized. *Elementarily equivalent domains* have a relationship between their values that needs to be captured in a metadata registry. Some *conceptually equivalent domains* have relationships between their values, too. These also need to be captured. This Technical Report describes how this can be accomplished.

While metadata registries can be used for storing information about a variety of metadata items, this Technical Report addresses only value domains, conceptual domains, and their associated attributes and relationships. The goal of this paper is to ensure that there is a common understanding of the content of the value domain attributes so that metadata can be shared between registries, despite their differences.

Information technology — Procedures for achieving metadata registry content consistency —

Part 3: **Value domains**

1 Scope

1.1 Background

An ISO/IEC 11179 metadata registry (MDR) is a tool for the management of shareable data; a comprehensive, authoritative source of reference information about data. It supports the standardization and harmonization processes by recording and disseminating descriptions of data, which facilitates data sharing among organizations and users. It provides links to documents that refer to specific data elements, value domains, and classification schemes and to information systems where those objects are used. When used in conjunction with a database, the registry enables users to understand any information obtained from the database better.

A registry does not contain data itself. It contains the metadata that is necessary to clearly describe, inventory, analyse, and classify data. It provides an understanding of the meaning, representation, and identification of units of data. This International Standard identifies the information elements that need to be available for determining the meaning of data to be shared between systems.

1.2 Purpose

The purpose of this Technical Report is to describe a set of procedures for the consistent registration of value domains and their attributes in a registry. This Technical Report is not a data entry manual, but a user's guide for conceptualizing a value domain and its components for the purpose of consistently establishing good quality metadata. An organization may adapt and/or add to these procedures as necessary.

1.3 Limits of this Technical Report

The scope of this Technical Report is limited to value domains, conceptual domains, and their associated attributes and relationships. Examples are used throughout the TR to illustrate the concepts described.

1.4 Registration approach — value domains and data elements

There is a choice when registering value domains in an MDR. Some Registration Authorities treat these sets as value domains, and others treat them as data elements. For the purposes of this Technical Report, the choice will always be to treat the sets as value domains unless explicitly stated. This choice is made to help illustrate the way to register many different kinds of value domains.

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/IEC 11179-1, *Information technology — Metadata registries (MDR) — Part 1: Framework for the specification and standardization of data elements*

ISO/IEC 11179-2, *Information technology — Specification and standardization of data elements — Part 2: Classification for data elements*

ISO/IEC 11179-3, *Information technology — Metadata registries — Part 3: Registry metamodel and basic attributes*

ISO/IEC 11179-4, *Information technology — Metadata registries (MDR) — Part 4: Rules and guidelines for the formulation of data definitions*

ISO/IEC 11179-5, *Information technology — Specification and standardization of data elements — Part 5: Naming and identification principles for data elements*

ISO/IEC 11179-6, *Information technology — Metadata registries (MDR) — Part 6: Registration of data elements*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 11179 and the following apply.

3.1
conceptually equivalent domains
value domains that represent the same conceptual domain

3.2
elementarily equivalent domains
domains that are elementarily equivalent if there exists a one-to-one correspondence between their permissible values such that given any pair of corresponding permissible values their value meanings are equal

NOTE 1 See Example in 4.2.5.

NOTE 2 All elementarily equivalent domains are conceptually equivalent. This follows from the fact that elementarily equivalent domains have the same set of value meanings, therefore they represent the same conceptual domain.

NOTE 3 Elementarily equivalence is an equivalence relation on the set of all enumerated value domains. So, any number of enumerated value domains may be elementarily equivalent to each other. See Examples in 5.4.1.

4 Understanding value domains

4.1 Introduction

This section is devoted to describing several things about value domains:

1) Some general principles about value domains

2) The structure or relationships that exist in some value domains

3) Code sets as value domains

4) Classification schemes as value domains

5) The relationship of data types to value domains

6) Use of units of measure

7) The importance of dimensionality

8) Classifying value domains

Examples are used throughout to illustrate the ideas. See Annex A for a detailed model (from ISO/IEC 11179-3) illustrating the relationships among all the constructs described herein.

4.2 General principles

4.2.1 Introduction

A *Value Domain* is a set of permissible values. A *Permissible Value* is a combination of some value and the meaning for that value. The associated meaning is called the *Value Meaning*. A permissible value is represented in this Technical Report as an ordered pair delimited by angle brackets as follows: <*value, value meaning*>. A value domain is the set of valid values for one or more data elements. It is used for validation of data in information systems and in data exchange. It is also an integral part of the metadata needed to describe a data element. In particular, a value domain is a guide to the content, form, and structure of the data represented by a data element.

Value domains come in two main types: enumerated and non-enumerated. An *Enumerated Value Domain* is a value domain where all the permissible values are listed explicitly. Examples of types of enumerated value domains include code sets, standard classifications, and categorizations. A *Non-enumerated Value Domain* is a value domain where the permissible values are expressed using a rule, called a *Non-enumerated Value Domain Description*. Thus, the permissible values are listed implicitly. This rule specifies precisely which values belong to the value domain and which do not. Examples of types of non-enumerated value domains include intervals of numbers, character strings, and bit maps.

A *Conceptual Domain* is a set of value meanings. It is a concept for which the extension is a collection of value domains. Conceptual domains, too, come in two main types: enumerated and non-enumerated. The value meanings for an *Enumerated Conceptual Domain* are listed explicitly. This type of conceptual domain corresponds to the enumerated type for value domains. The value meanings for a *Non-enumerated Conceptual Domain* are expressed using a rule, called a *Non-enumerated Conceptual Domain Description*. Thus, the value meanings are listed implicitly. This rule describes the meaning of permissible values in a non-enumerated value domain. This type of conceptual domain corresponds to the non-enumerated type for value domains.

Every value domain represents two kinds of concepts: data element concept (indirectly) and conceptual domain (directly). The *Data Element Concept* is the concept associated with a data element. The value domain is the representation for the data element, and, therefore, indirectly represents the data element concept, too. However, the value domain is directly associated with a conceptual domain, so represents that concept, independent of any data element.

An example will help to illustrate the distinctions in the discussion, which is shown below:

EXAMPLE

Data element name:	Sex of employee – code
Data element concept name:	Sex of employee
Data element concept definition:	The sex of the employee of an organization.
Conceptual domain name:	Human sex categories
Conceptual domain definition:	Enumerations of human sexes.
Value domain name:	Human sex codes (1)
Value domain definition:	Codes for the human sexes.
Permissible values:	<1, Male>
	<2, Female>
	<0, Unknown>

The codes used in the value domain above are taken from ISO/IEC 5218. Using standardized codes ensures interoperability between metadata registries and application systems. However, in general, the choice of codes for a value domain may be arbitrary. In this case, the MDR is the source for obtaining the values and their meanings for a non-standard value domain.

Several points about value domains need to be made here.

4.2.2 Choice of codes

The choice of codes used in the value domain above is arbitrary. Another code set might work just as well, but the set is a different value domain. Which value domain to use is determined by the needs of the application and the organization. The following example is another code set for human sex codes:

EXAMPLE

Value domain name:	Human sex codes (2)
Value domain definition:	Codes for the human sexes.
Permissible values:	<M, Male>
	<F, Female>
	<U, Unknown>

4.2.3 Number of permissible values

The number of permissible values (3 in our example) may also be different. We might want a code for representing hermaphrodites or a code for representing transsexuals. Each time new permissible values are added or subtracted, a new value domain, or value domain version, is created. Determining whether a change to a value domain merits the creation of a new value domain or just a new version of an existing value domain is up to the individual registration authority. The following example shows an expanded value domain considered as a new one, not a version of an old one, as in the example in 4.2.2.

EXAMPLE

Value domain name:	Human sex codes (3)
Value domain definition:	Codes for the human sexes.
Permissible values:	<M, Male>
	<F, Female>
	<H, Hermaphrodite>
	<T, Transsexual>
	<U, Unknown>

4.2.4 Conceptual domain hierarchies

All the value domains for human sex codes can be viewed as being conceptually equivalent. There is no requirement that each of the value meanings in a conceptual domain be associated with a value. However, some Registration Authorities may decide that to adequately differentiate the concept, for example, of using five categories of human sex codes instead of three, separate conceptual domains must be created. See example below. At the highest level, all the value domains (examples in 4.2.1, 4.2.2, and 4.2.3) represent the idea of categories of human sexes. So, the super-ordinate conceptual domain captures the concept represented by a class of value domains (e.g., human sex codes) needed within a registry. The subordinate conceptual domains provide the enumeration of value meanings to be mapped to the corresponding value domains.

EXAMPLE

Super-ordinate conceptual domain	(non-enumerated)
Conceptual domain name:	Human sex categories
Conceptual domain definition:	Categorizations of human sexes.
Subordinate conceptual domain (1)	(enumerated)
Conceptual domain name:	Human sex categories: 3 values
Conceptual domain definition:	Enumerations of human sexes with 3 categories.
Subordinate conceptual domain (2)	(enumerated)
Conceptual domain name:	Human sex categories: 5 values
Conceptual domain definition:	Enumerations of human sexes with 5 categories

4.2.5 Sharing value meanings across permissible values

The value meaning is used to link equivalent permissible values across conceptually equivalent domains. In elementarily equivalent domains, each value meaning links equivalent codes between a unique pair of permissible values, one from each value domain, as the following example illustrates:

EXAMPLE

Conceptual domain name:	Human sex categories
Conceptual domain definition:	Enumerations of human sexes.
Value domain names:	Human sex codes (1) (See Example, 4.2.1)
	Human sex codes (2) (See Example, 4.2.2)
Value domain definition:	Codes for the human sexes.

A one-to-one correspondence that preserves value meanings between these two enumerated value domains is defined as follows:

HSC(1)		→	HSC(2)
<1, Male>	↔		<M, Male>
<2, Female>	↔		<F, Female>
<0, Unknown>	↔		<U, Unknown>

Each pair of corresponding permissible values has the same value meaning. So, these two enumerated value domains are elementarily equivalent and, therefore, conceptually equivalent.

Each permissible value in one of the two value domains listed above shares its value meaning with that of a permissible value in the other value domain. So, through the use of value meanings, equivalence of values across value domains is achievable, e.g., the values 1 and M mean Male or the values 2 and F mean Female. These two value domains are elementarily equivalent domains.

4.2.6 Sharing value domains across data elements

Sex of employee (the idea that employees are classified or characterized by sex) and *sex of student* (the idea that students are classified or characterized by sex) are different data element concepts, but they could use the same value domain to represent them. So, a value domain (e.g., Human Sex Codes (1)) may be associated with many data element concepts, and, therefore, data elements.

4.2.7 Associating value domains with concepts (data element concepts and conceptual domains)

A data element concept is associated with different value domains as needed to describe similar, but different, data elements, and those value domains are conceptualized by the same conceptual domain (e.g., Human Sex Codes (1), Human Sex Codes (2), Human Sex Codes (3) in the examples). However, the converse is not true: two value domains under the same conceptual domain do not need to be associated with the same data element concept. The following two examples (1 and 2) are of this type:

EXAMPLE 1

Conceptual domain name:	Human sex categories
Conceptual domain definition:	Enumerations of human sexes.
Value domain name:	Human sex codes (1) (See Example in 4.2.1)
Value domain definition:	Codes for the human sexes.
Data element concept name:	Sex of employee
Data element concept definition:	The biological sex of the employee of an organization.

EXAMPLE 2

Conceptual domain name:	Human sex categories
Conceptual domain definition:	Enumerations of human sexes.
Value domain name:	Human sex codes (2) (See Example in 4.2.2)
Value domain definition:	Codes for the human sexes.
Data element concept name:	Sex of student
Data element concept definition:	The biological sex of the student of an educational institution.

4.2.8 Value domains not associated with data elements

Value domains do not have to be associated with a data element concept at all. They can be managed independently, such as code sets sometimes are. For instance, the maintenance agency, which is the authority for maintaining the code values for a standard code set, might make the code set freely available even though it published no data using that value domain.

4.2.9 Contrasting conceptual domains and data element concepts

There are two kinds of semantics associated with data: symbolic and contextual. Symbolic semantics refers to the meaning of symbols, i.e. values. The conceptual domain captures this kind of semantics, as it is a set of value meanings.

Contextual semantics addresses the interrogatives (who, what, when, where, why, and how) as they relate to data. Fundamentally, data represent some observation of a property of a member of some set of objects, an object class. The term observation usually implies some human action, but mechanical or electrical instruments can make or record observations, too.

The object class describes who is being observed, the property is the distinguishing or describing feature of an object that is observed, and the instrument (including humans) is how the observation is made. Both the object class and property are concepts, and they form the basic contextual meaning associated with data. Without them, data have no reason to be. The data element concept captures this meaning.

Details about the instrument (how) and why observations were made (the experimental design) are out of scope for the MDR. The time (when) an observation was made and where it was made are specific to individual observations. This case level metadata is often found in a data record in companion data elements in a database.

4.2.10 Non-enumerated value domains

The examples provided so far are for enumerated value domains. Non-enumerated value domains are used to represent and constrain data through a rule rather than through an enumerated list of permissible values. The following example is of a non-enumerated value domain and associated data element and data element concept:

EXAMPLE

Conceptual domain name:	Industry descriptions
Conceptual domain definition:	Text describing an industry.
Non-enumerated conceptual domain description:	Limited length text describing an industry
Value domain name:	Textual English industry descriptions
Value domain definition:	Textual descriptions of an industry.
Non-enumerated value domain description:	English text up to 60 characters
Data element name:	Industry description for person's job - text
Data element concept name:	Industry description for person's job
Data element concept definition:	The description of the industry within which a person works.

4.2.11 Value domains with enumerated and non-enumerated components

It is possible, although rare, for a domain to have enumerated and non-enumerated components. This situation may occur when values fall within a certain range, have a minimum (or maximum) value, and discrete values below the minimum (or above the maximum) are used for special cases. The following example (1) will illustrate this:

EXAMPLE 1

Data element name:	Volume of household monthly water usage - gallons
Data element concept name:	Volume of household monthly water usage
Data element concept definition:	Volume of water used by a household each month.
Conceptual domain name:	Volumes (liquid)
Conceptual domain definition:	Measures of liquid volume with additional special values.
Non-enumerated conceptual domain description:	Liquid volume measures
Value domain name:	Volume in gallons (US, liquid) x 1000
Value domain definition:	Liquid volume in thousands of US gallons.
Non-enumerated value domain description:	Integers greater than or equal to 0, and special values -1 and -2
Permissible values:	<-1, Not reported> <-2, Not measurable (e.g., use of well or stream)>

Another situation where a value domain has both enumerated and non-enumerated parts is illustrated in the following example (2). Here, special meaning is attached to a finite number of values in a range.

EXAMPLE 2

Data element name:	Likelihood of voting in the next election, on a scale of 0 to 1
Data element concept name:	Likelihood of voting in the next election
Data element concept definition:	A measure of how likely the person will vote in the next election.
Conceptual domain name:	Rating scales
Conceptual domain definition:	A measure of effort, attitude, preference, difficulty, etc.
Non-enumerated conceptual domain description:	Rating scale expressed numerically
Value domain name:	Preference scale, 0 to 1
Value domain definition:	Preferences measured on a scale of 0 to 1.
Non-enumerated value domain description:	All real numbers between 0 and 1.
Permissible values	<0, Definitely not> <0.5, No preference> <1, Definitely>

NOTE This example requires some further explanation. The data may be obtained by reading a mark made by a respondent on a scale printed on a form. The value is interpolated through finding the ratio of the distance from the mark to the "0" end divided by the length of the entire scale. In addition, the use of both enumerated and non-enumerated