INTERNATIONAL STANDARD

ISO/IEC 11179-1

Second edition
2004-09-15

Information technology — Metadata registries (MDR) —

Part 1: **Framework**

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

Partie 1: Cadre

Reference number
ISO/IEC 11179-1:2004(E)

PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

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 11179-1 was prepared by 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 11179-1:1999), 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 2: Classification*

— *Part 3: Registry metamodel and basic attributes*

— *Part 4: Formulation of data definitions*

— *Part 5: Naming and identification principles*

— *Part 6: Registration*

Introduction

ISO/IEC 11179 - *Metadata registries (MDR)*, addresses the semantics of data, the representation of data, and the registration of the descriptions of that data. It is through these descriptions that an accurate understanding of the semantics and a useful depiction of the data are found.

The purposes of ISO/IEC 11179 are to promote the following:

— Standard description of data

— Common understanding of data across organizational elements and between organizations

— Re-use and standardization of data over time, space, and applications

— Harmonization and standardization of data within an organization and across organizations

— Management of the components of data

— Re-use of the components of data

ISO/IEC 11179 is six part standard. Each part is devoted to addressing a different aspect of the needs listed above. The parts and a short description follow:

— Part 1 – *Framework* – Contains an overview of the standard and describes the basic concepts

— Part 2 – *Classification* – Describes how to manage a classification scheme in a metadata registry

— Part 3 – *Registry metamodel and basic attributes* – Provides the basic conceptual model, including the basic attributes and relationships, for a metadata registry

— Part 4 – *Formulation of data definitions* – Rules and guidelines for forming quality definitions for data elements and their components

— Part 5 – *Naming and identification principles* – Describes how to form conventions for naming data elements and their components

— Part 6 – *Registration* – Specifies the roles and requirements for the registration process in an ISO/IEC 11179 metadata registry

Generally, descriptive data is known as metadata. That is, metadata is data that is used for describing other data. As the use of the term has evolved, metadata now refers, generally, to data that is used for describing some other objects. We limit the scope of the term as it is used here in ISO/IEC 11179 to descriptions of data - the more traditional use of the term.

An MDR is a database of metadata that supports the functionality of registration. Registration accomplishes three main goals: identification, provenance, and monitoring quality. Identification is accomplished by assigning a unique identifier (within the registry) to each object registered there. Provenance addresses the source of the metadata and the object described. Monitoring quality ensures that the metadata does the job it is designed to do.

An MDR manages the semantics of data. Understanding data is fundamental to its design, harmonization, standardization, use, re-use, and interchange. The underlying model for an MDR is designed to capture all the basic components of the semantics of data, independent of any application or subject matter area.

MDR's are organized so that those designing applications can ascertain whether a suitable object described in the MDR already exists. Where it is established that a new object is essential, its derivation from an existing description with appropriate modifications is encouraged, thus avoiding unnecessary variations in the way similar objects are described. Registration will also allow two or more administered items describing identical objects to be identified, and more importantly, it will identify situations where similar or identical names are in use for administered items that are significantly different in one or more respects.

In ISO/IEC 11179 the basic container for data is called a data element. It may exist purely as an abstraction or exist in some application system. In either case, the description of a data element is the same in ISO/IEC 11179. Data element descriptions have both semantic and representational components. The semantics are further divided into contextual and symbolic types.

The contextual semantics are described by the data element concept (DEC). The DEC describes the kinds of objects for which data are collected and the particular characteristic of those objects being measured. The symbolic semantics are described by the conceptual domain (CD). A CD is a set of categories, not necessarily finite, where the categories represent the meaning of the permissible values in a value domain - the allowed values for a data element.

The names, definitions, datatype, and related objects that are associated with a particular object in an MDR give that object meaning. The depth of this meaning is limited, because names and definitions convey limited information about an object. The relationships that object has with semantically related objects in a registry provides additional information, but the additional information is dependent on how many semantically related objects there are.

The representational component is about the permitted values a data element may use. Each value corresponds to one of the categories in the CD. The set of these permitted values is called a value domain (VD). A VD specifies all the values that are allowed either through an enumeration, a rule, or a combination of these. The computational model the values follow is given by their datatype.

The semantic and representational components are described through attributes contained in the conceptual model of a metadata registry as specified in ISO/IEC 11179-3. A metadata registry that conforms to ISO/IEC 11179 can describe a wide variety of data. In fact, the attributes described in ISO/IEC 11179-3 are data elements, and they can be registered in an ISO/IEC 11179 metadata registry. Moreover, any set of descriptors or metadata attributes may be interpreted as data elements and registered in the metadata registry.

There are two main consequences to this:

— The metadata registry can describe itself

— Metadata layers or levels are not defined in ISO/IEC 11179

As a result, ISO/IEC 11179 is a general description framework for data of any kind, in any organization, and for any purpose. This standard does not address other data management needs, such as data models, application specifications, programming code, program plans, business plans, and business policies. These need to be addressed elsewhere.

The increased use of data processing and electronic data interchange heavily relies on accurate, reliable, controllable, and verifiable data recorded in databases. One of the prerequisites for a correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and descriptive characteristics (e.g., representation) of that data. To guarantee this shared view, a number of basic attributes has to be defined.

The basic attributes specified are applicable for the definition and specification of the contents of data dictionaries and interchanging or referencing among various collections of administered items. The "basic" in basic attributes means that the attributes are commonly needed in specifying administered items completely enough to ensure that they will be applicable for a variety of functions, such as

— design of information processing systems

— retrieval of data from databases

— design of EDI-messages for data interchange

— maintenance of metadata registries

— data management

— dictionary design

— dictionary control

— use of information processing systems

Basic also implies that they are independent of any

— application environment

— function of an object described by an administered item

— level of abstraction

— grouping of administered items

— method for designing information processing systems or data interchange messages

— MDR system

Basic does not imply that all attributes specified in ISO/IEC 11179-3 are required in all cases. Distinction is made between those attributes that are mandatory, conditional, or optional.

Information technology — Metadata registries (MDR) —

Part 1:
Framework

1 Scope

ISO/IEC 11179 specifies the kind and quality of metadata necessary to describe data, and it specifies the management and administration of that metadata in a metadata registry (MDR). It applies to the formulation of data representations, concepts, meanings, and relationships between them to be shared among people and machines, independent of the organization that produces the data. It does not apply to the physical representation of data as bits and bytes at the machine level.

In ISO/IEC 11179, metadata refers to descriptions of data. ISO/IEC 11179 does not contain a general treatment of metadata. This part of ISO/IEC 11179 provides the means for understanding and associating the individual parts and is the foundation for a conceptual understanding of metadata and metadata registries.

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, *Terminology work — Principles and methods*

ISO 1087-1:2000, *Terminology work — Vocabulary — Part 1: Theory and application*

ISO/IEC 11179 (all parts), *Information technology — Metadata registries (MDR)*

3 Terms and definitions

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

3.1 Definitions of modeling constructs

This sub-clause defines the modeling constructs used in this part of ISO/IEC 11179.

3.1.1
attribute
characteristic of an **object** or **entity**

3.1.2
class
description of a set of **object**s that share the same **attribute**s, operations, methods, **relationship**s, and semantics

[ISO/IEC 19501-1:2001, 2.5.2.9].

3.1.3
identifier (in **Metadata Registry**)
sequence of characters, capable of uniquely identifying that with which it is associated, within a specified **context**

NOTE A name should be used as an identifier because it is not linguistically neutral.

3.1.4
relationship
connection among model elements

[ISO/IEC 19501-1:2001, 2.5.2.36].

3.2 General terms used in this part of ISO/IEC 11179

This sub-clause defines terms that have general usage beyond the specific needs of this part of ISO/IEC 11179, but are not modeling constructs defined in 3.1.

3.2.1
basic attribute
attribute of a **metadata item** commonly needed in its specification

3.2.2
characteristic
abstraction of a property of an **object** or of a set of objects

NOTE Characteristics are used for describing **concept**s.

[ISO 1087-1:2000, 3.2.4].

3.2.3
concept
unit of knowledge created by a unique combination of **characteristic**s

[ISO 1087-1:2000, 3.2.1].

3.2.4
concept system
set of **concept**s structured according to the relations among them

[ISO 1087-1:2000, 3.2.11]

3.2.5
conceptual data model
conceptual model
data model that represents an abstract view of the real world

NOTE A conceptual model represents the human understanding of a system.

3.2.6
data
re-interpretable representation of information in a formalized manner suitable for communication, interpretation, or processing

NOTE Data can be processed by humans or by automatic means.

[ISO 2382-1:1993, 01.01.02].

3.2.7
data model
graphical and/or lexical representation of **data**, specifying their properties, structure and inter-relationships

3.2.8
definition
representation of a **concept** by a descriptive statement which serves to differentiate it from related **concept**s

[ISO 1087-1:2000, 3.3.1].

3.2.9
designation
representation of a concept by a sign which denotes it

[ISO 1087-1:2000, 3.4.1].

3.2.10
entity
any concrete or abstract thing that exists, did exist, or might exist, including associations among these things

EXAMPLE A person, object, event, idea, process, etc.

NOTE An entity exists whether data about it are available or not.

[ISO/IEC 2382-17:1999, 17.02.05].

3.2.11
essential characteristic
characteristic which is indispensable to understanding a **concept**

[ISO 1087-1:2000, 3.2.6].

3.2.12
extension
<terminology>
totality of **object**s to which a **concept** corresponds

[ISO 1087-1:2000, 3.2.8].

NOTE This term has a different meaning in ISO/IEC 11179-3.

3.2.13
general concept
concept which corresponds to two or more **objects**, which form a group by reason of common properties

NOTE Examples of general concepts are 'planet', 'tower'.

[ISO 1087-1:2000, 3.2.3]

3.2.14
individual concept
concept which corresponds to only one **object**

NOTE Examples of individual concepts are: 'Saturn', 'the Eiffel Tower'.

[ISO 1087-1:2000, 3.2.2].

3.2.15
intension
<terminology>
set of **characteristic**s which makes up the **concept**

[ISO 1087-1:2000, 3.2.9].

3.2.16
metadata
data that defines and describes other **data**

3.2.17
metadata item
instance of a **metadata object**

3.2.18
metadata object
object type defined by a metamodel

3.2.19
metadata registry
MDR
information system for registering **metadata**

3.2.20
metamodel
data model that specifies one or more other data models

3.2.21
name
designation of an **object** by a linguistic expression

3.2.22
object
anything perceivable or conceivable

NOTE Objects may also be material (e.g. an engine, a sheet of paper, a diamond), immaterial (e.g. a conversion ratio, a project plan), or imagined (e.g. a unicorn).

[ISO 1087-1:2000, 3.1.1].

3.2.23
registry item
metadata item recorded in a **metadata registry**

3.2.24
registry metamodel
metamodel specifying a **metadata registry**

3.2.25
terminological system
concept system with **designations** for each **concept**

3.3 Alphabetical list of terms used in the metamodel

This sub-clause provides definitions for terms used in this part of ISO/IEC 11179, which are the names of metadata objects in the metamodel specified in ISO/IEC 11179-3.

3.3.1
administered item
registry item for which administrative information is recorded in an **administration record**

3.3.2
administration record
collection of administrative information for an **administered item**

3.3.3
administrative status
designation of the status in the administrative process of a **registration authority** for handling registration requests

NOTE The values and associated meanings of "administrative status" are determined by each **registration authority**. C.f. "**registration status**".

3.3.4
classification scheme
descriptive information for an arrangement or division of **object**s into groups based on **characteristic**s, which the objects have in common

3.3.5
classification scheme item
CSI
item of content in a **classification scheme**.

NOTE This may be a node in a taxonomy or ontology, a term in a thesaurus, etc.

3.3.6
conceptual domain
CD
set of valid **value meaning**s

NOTE The **value meaning**s in a **conceptual domain** may either be enumerated or expressed via a description.

3.3.7
context
circumstance, purpose, and perspective under which an **object** is defined or used

NOTE This term has a different meaning in 11179-3.

3.3.8
data element
DE
unit of **data** for which the **definition**, identification, representation and **permissible value**s are specified by means of a set of **attribute**s

3.3.9
data element concept
DEC
concept that can be represented in the form of a **data element**, described independently of any particular representation

3.3.10
data identifier
DI
unique **identifier** for an **administered item** within a **registration authority**

3.3.11
datatype
set of distinct values, characterized by properties of those values and by operations on those values

[ISO/IEC 11404:1996, 4.11].

3.3.12
dimensionality
expression of measurement without units

NOTE A quantity is a value with an associated unit of measure. 32º Fahrenheit, 0º Celsius, $100 USD, and 10 reams (of paper) are quantities. Equivalence between two units of measure is determined by the existence of a quantity preserving one-to-one correspondence between values measured in one unit of measure and values measured in the other unit of measure, independent of context, and where characterizing operations are the same. Equivalent units of measure in this sense have the same dimensionality. The equivalence defined here forms an equivalence relation on the set of all units of measure. Each equivalence class corresponds to a dimensionality. The units of measure "temperature in degrees Fahrenheit" and "temperature in degrees Celsius" have the same dimensionality, because for each value measured in degrees Fahrenheit there is a value measured in degrees Celsius with the same quantity, and vice-versa. The same operations may be performed on quantities in each unit of measure. Quantity preserving one-to-one correspondences are the well-known equations Cº = (5/9)*(Fº - 32) and Fº = (9/5)*(Cº) + 32.

3.3.13
enumerated conceptual domain
conceptual domain that is specified by a list of all its **value meaning**s

3.3.14
enumerated value domain
value domain that is specified by a list of all its **permissible value**s

3.3.15
international code designator
ICD
identifier of an organization identification scheme

NOTE Based on ISO/IEC 6523-1:1998, 3.8.

3.3.16
item identifier
identifier for an item

3.3.17
item registration authority identifier
identifier of the **registration authority** registering the item

3.3.18
non-enumerated conceptual domain
conceptual domain that is not specified by a list of all valid **value meaning**s

3.3.19
non-enumerated conceptual domain description
description or specification of a rule, reference, or range for a set of all **value meaning**s for the **conceptual domain**

3.3.20
non-enumerated value domain
value domain that is specified by a description rather than a list of all **permissible value**s

3.3.21
non-enumerated value domain description
description or specification of a rule, reference, or range for a set of all **permissible value**s for the **value domain**

3.3.22
object class
set of ideas, abstractions, or things in the real world that are identified with explicit boundaries and meaning and whose properties and behavior follow the same rules

3.3.23
organization
unique framework of authority within which a person or persons act, or are designated to act, towards some purpose

[ISO/IEC 6523-1:1998, 3.1].

3.3.24
organization identifier
identifier assigned to an **organization** within an organization identification scheme, and unique within that scheme

[ISO/IEC 6523-1:1998, 3.10].

3.3.25
organization part
any department, service, or other **entity** within an **organization** which needs to be identified for information exchange

[ISO/IEC 6523-1:1998, 3.2].

3.3.26
organization part identifier
OPI
identifier allocated to a particular **organization part**

[ISO/IEC 6523-1:1998, 3.11].

3.3.27
organization part identifier source
source for the **organization part identifier**

NOTE Based on ISO/IEC 6523-1:1998, 3.12.

3.3.28
permissible value
expression of a **value meaning** allowed in a specific **value domain**

3.3.29
property
characteristic common to all members of an **object class**

3.3.30
registrar
representative of a **registration authority**

3.3.31
registration
relationship between an **administered item** and the **registration authority**

3.3.32
registration authority
RA
organization responsible for maintaining a register

3.3.33
registration authority identifier
identifier assigned to a **registration authority**

3.3.34
registration status
designation of the status in the registration life-cycle of an **administered item**

3.3.35
representation class
classification of types of representations

3.3.36
unit of measure
actual units in which the associated values are measured

NOTE The **dimensionality** of the associated **conceptual domain** must be appropriate for the specified **unit of measure**.

3.3.37
value
data value

3.3.38
value domain
VD
set of **permissible values**

3.3.39
value meaning
meaning or semantic content of a **value**

NOTE Given a **permissible value**, representation of its **value meaning** shall be independent of (and shall not constrain) the representation of its corresponding **value**.

3.3.42
value meaning description
description of a **value meaning**

3.3.43
version
unique version **identifier** of the **administered item**

3.4 Specific terms used in this part of ISO/IEC 11179

This sub-clause defines terms that have specific usage in this part of ISO/IEC 11179 and are not used in the other parts.

3.4.1
data construct
object a **metadata item** describes

NOTE Individual data elements, value domains, data element concepts, conceptual domains, object classes, and properties are data constructs.

3.4.2
quantity
value associated with a unit of measure

4 Abbreviations and acronyms

NOTE Some of the abbreviations or acronyms in this section represent terms defined in Clause 3.

CD -- Conceptual Domain

DE -- Data Element

DEC -- Data Element Concept

DI -- Data Identifier

EDI -- Electronic Data Interchange

IEC -- International Electrotechnical Commission

ICD -- International Code Designator

ISO -- International Organization for Standardization

JTC1 -- Joint Technical Committee 1

MDR -- Metadata Registry

OPI -- Organization Part Identifier

RA -- Registration Authority

SC32 -- ISO/IEC JTC1/Sub-committee 32

5 Theory of terminology

The concepts from the theory of terminology that are used in ISO/IEC 11179 shall be in conformity with ISO 704 and ISO 1087-1. A short description of the necessary theory follows.

In the theory of terminology, an object is something conceivable or perceivable. **Concepts** are mental constructs, units of thought, or unit of knowledge created by a unique combination of characteristics. Concepts are organized or grouped by common elements, called **characteristics**. **Essential characteristics** are indispensable to understanding a concept. Other characteristics are **inessential**. The sum of characteristics that constitute a concept is called its **intension**. The set of objects a concept refers to is its **extension**.

In natural language, concepts are expressed through **definitions**, which specify a unique intension and extension.

A **designation** (term, appellation, or symbol) represents a concept.

A **general concept** has two or more objects that correspond to it. An **individual concept** has one object that corresponds to it. That is, a general concept has two or more objects in its extension, and an individual concept has one object in its extension.

A **concept system** is set of concepts structured according to the relations among them. A **terminological system** is a concept system with designations for each concept.

6 Metadata

6.1 Concepts

For ISO/IEC 11179, **metadata** is defined to be data that defines and describes other data. This means that metadata are data, and data become metadata when they are used in this way. This happens under particular circumstances, for particular purposes, and with certain perspectives, as no data are always metadata. The set of circumstances, purposes, or perspectives for which some data are used as metadata is called the **context**. So, metadata are data about data in some context.

Since metadata are data, then metadata can be stored in a database and organized through the use of a model. Some models are very application specific, and others are more general. The model presented and described in ISO/IEC 11179-3 (*Registry metamodel and basic attributes*) is general. It is a representation of the human understanding of the metadata needed to describe **data constructs**, including the relationships that exist among that metadata, and not necessarily how the metadata will be represented in an application of an MDR. A model of this kind is called a **conceptual model**. Conceptual models are meant for people to read and understand.

Models that describe metadata are often referred to as **metamodels**. The conceptual model presented in ISO/IEC 11179-3 is a metamodel in this sense.

6.2 Fundamental model of data elements

Figure 1 illustrates the ideas conveyed in this sub-clause. The figure itself is not normative, but it is used to illustrate the basic ideas.

For the purposes of ISO/IEC 11179, a **data element** is composed of two parts:

— **Data element concept** – A DEC is **concept** that can be represented in the form of a **data element**, described independently of any particular representation.

— **Representation** – The representation is composed of a value domain, datatype, units of measure (if necessary), and representation class (optionally).

From a data modeling perspective and for the purposes of ISO/IEC 11179, a data element concept may be composed of two parts:

— The **object class** is a set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behavior follow the same rules

— The **property** is a characteristic common to all members of an object class

Object classes are the things for which we wish to collect and store data. They are concepts, and they correspond to the notions embodied in classes in object-oriented models and entities in entity-relationship models. Examples are cars, persons, households, employees, and orders. Properties are what humans use to distinguish or describe objects. They are characteristics, not necessarily essential ones, of the object class and form its intension. They are also concepts, and they correspond to the notions embodied in attributes (without associated datatypes) in object-oriented or entity-relationship models. Examples of properties are color, model, sex, age, income, address, or price.

An object class may be a **general concept**. This happens when the set of objects corresponding to the object class has two or more members. The examples in the previous paragraph are of this type. Record level data are described this way. On the other hand, an object class may be an **individual concept**. This happens when the set of objects corresponding to the object class has one member. Examples are concepts corresponding to single objects, such as "the set of persons in the US" or "the set of service sector establishments in Australia". Aggregate data are described this way. Examples of properties are average income or total earnings.

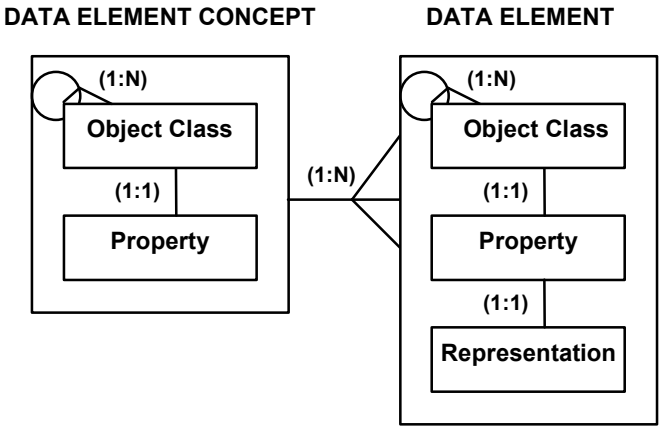
It is important to distinguish an actual object class or property from its name. This is the distinction between concepts and their designations. Object classes and properties are concepts; their names are designations. Complications arise because people convey concepts through words (designations), and it is easy to confuse a concept with the designation used to represent it. For example, most people will read the word income and be certain they have unambiguously interpreted it. But, the designation income may not convey the same concept to all readers, and, more importantly, each instance of income may not designate the same concept.

Not all ideas are simply expressed in a natural language, either. For example, "women between the ages of 15 and 45 who have had at least one live birth in the last 12 months" is a valid object class not easily named in English. Some ideas may be more easily expressed in one language than in another. The German word *Götterdämmerung* has no simple English equivalent.

A data element is produced when a representation is associated with a data element concept. The representation describes the form of the data, including a value domain, datatype, representation class (optionally), and, if necessary, a unit of measure. **Value domains** are sets of permissible values for data elements. For example, the data element representing annual household income may have the set of non-negative integers (with units of dollars) as a set of valid values. This is its value domain.

A data element concept may be associated with different value domains as needed to form conceptually similar data elements. There are many ways to represent similar facts about the world, but the concept for which the facts are examples is the same. Take the DEC country of person's birth as an example. ISO 3166-1 – *Country Codes* contains seven different representations for countries of the world. Each one of these seven representations contains a set of values that may be used in the value domain associated with the DEC. Each one of the seven associations is a data element. For each representation of the data, the permissible values, the datatype, the representation class, and possibly the units of measure, are altered.

See ISO/IEC 20943-1:2003, *Information technology — Procedures for achieving metadata registry content consistency — Part 1: Data elements* for details about the registration and management of descriptions of data elements.



This figure is for informational purposes only. It is not normative.

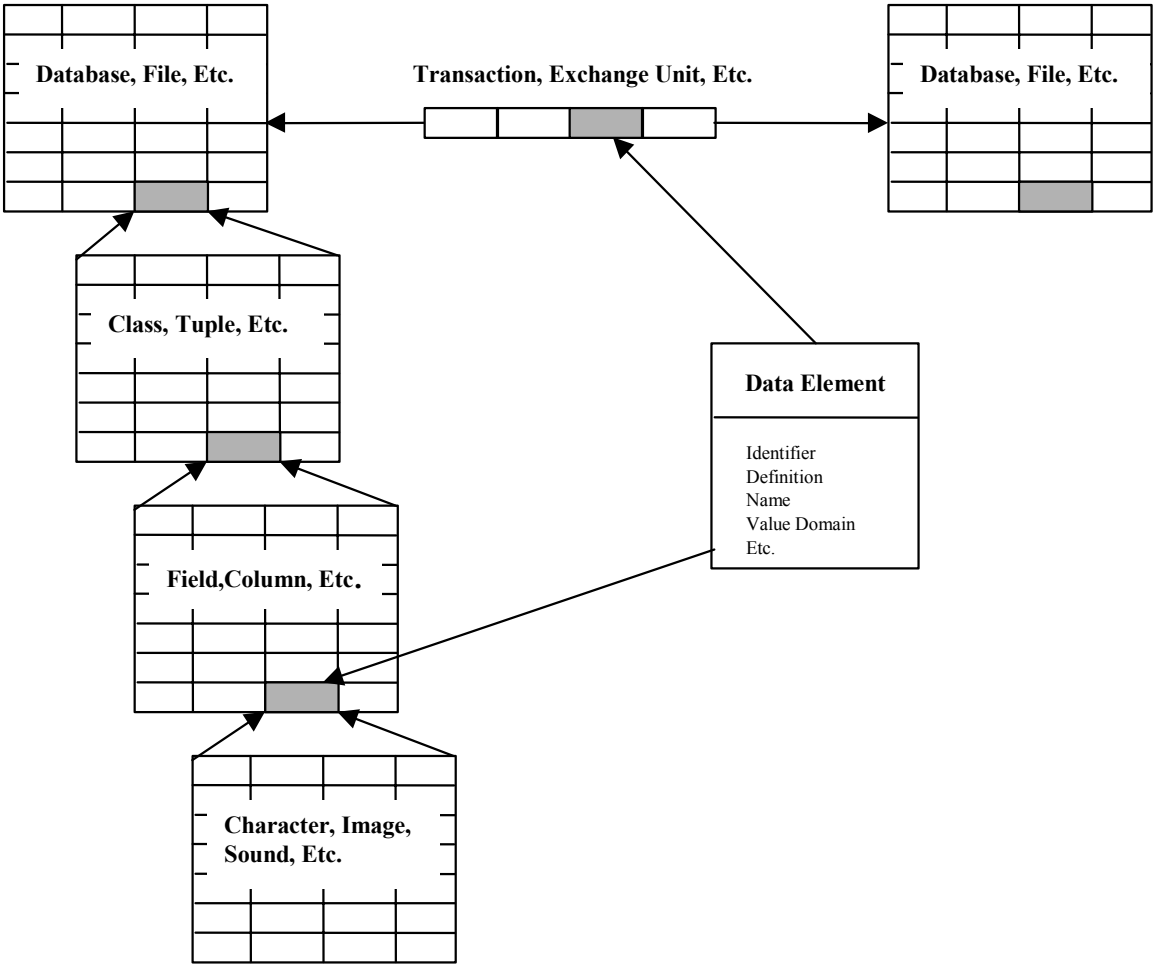
Figure 1 — Fundamental model for data elements

6.3 Data elements in data management and interchange

Figure 2 provides a simplified picture to illustrate those situations in which data elements lie. Data elements appear in databases, files, and transaction sets. Data elements are the fundamental units of data an organization manages, therefore they must be part of the design of databases and files within the organization and all transaction sets the organization builds to communicate data to other organizations.

Within the organization, databases or files are composed of records, segments, tuples, etc., which are composed of data elements. The data elements themselves contain various kinds of data that include characters, images, sound, etc.

When the organization needs to transfer data to another organization, data elements are the fundamental units that make up the transaction sets. Transactions occur primarily between databases or files, but the structure (i.e. the records or tuples) of the files and databases don't have to be the same across organizations. So, the common unit for transferring information (data plus understanding) is the data element.



This figure is for informational purposes only. It is not normative.

Figure 2 — Data elements and other data concepts

6.4 Fundamental model of value domains

Figure 3 illustrates the ideas conveyed in this sub-clause. The figure itself is not normative, but it is used to illustrate the basic ideas.

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 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 (non-exclusive) sub-types:

— **Enumerated value domain** – A value domain specified as a list of permissible values (values and their meanings)

— **Non-enumerated value domain** – A value domain specified by a description

An enumerated value domain contains a list of all its values and their associated meanings. Each value and meaning pair is called a **permissible value**. The meaning for each value is called the **value meaning**.

A non-enumerated value domain is specified by a description. The **non-enumerated value domain description** describes precisely which permissible values belong and which do not belong to the value domain. An example of a description is the phrase "Every real number greater than 0 and less than 1".

Each value domain is a member of the extension of a concept, called the **conceptual domain**. A conceptual domain is a set of value meanings. The intension of a conceptual domain is its value meanings. Many value domains may be in the extension of the same conceptual domain, but a value domain is associated with one conceptual domain. Conceptual domains may have relationships with other conceptual domains, so it is possible to create a concept system of conceptual domains. Value domains may have relationships with other value domains, which provide the framework to capture the structure of sets of related value domains and their associated concepts.

Conceptual domains, too, come in two (non-exclusive) sub-types:

— **Enumerated conceptual domain** – A conceptual domain specified as a list of value meanings

— **Non-enumerated conceptual domain** – A conceptual domain specified by a description

The value meanings for an enumerated conceptual domain are listed explicitly. This conceptual domain type 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 conceptual domain type corresponds to the non-enumerated type for value domains. See ISO/IEC TR 20943-3, *Information technology — Procedures for achieving metadata registry content consistency — Part 3: Value domains* for detailed examples.

A unit of measure is sometimes required to describe data. If temperature readings are recorded in a database, then the temperature scale (e.g., Fahrenheit or Celsius) is necessary to understand the meaning of the values. Another example is the mass of rocks found on Mars, measured in grams. However, units of measure are not limited to physical quantities, as currencies (e.g., US dollars, Lire, British pounds) and other socio-economic measures are units of measure, too.

Some units of measure are equivalent to each other in the following sense: Any quantity in one unit of measure can be transformed to the same quantity in another unit of measure. All equivalent units of measure are said to have the same dimensionality. For example, currencies all have the same dimensionality. Measures of speed, such as miles per hour or meters per second, have the same dimensionality. Two units of measure that are often erroneously seen as having the same dimensionality are pounds (as in weight) and grams. Pounds is a measure of force, and grams is a measure of mass.

A unit of measure is associated with a value domain, and the dimensionality is associated with the conceptual domain.

Some value domains contain very similar values from one domain to another. Either the values themselves are similar or the meanings of the values are the same. When these similarities occur, the value domains may be in the extension of one conceptual domain. The following examples illustrate this and the other ideas in this sub-clause:

EXAMPLE 1 – Similar non-enumerated value domains

Conceptual domain name: Probabilities

Conceptual domain definition: Real numbers greater than 0 and less than 1.

Value domain name (1): Probabilities – 2 significant digits

Value domain description: All real numbers greater than 0 and less than 1 represented with 2-digit precision.

Unit of measure precision: 2 digits to the right of the decimal point

---

Value domain name (2): Probabilities – 5 significant digits

Value domain description: All real numbers greater than 0 and less than 1 represented with 5-digit precision

Unit of measure precision: 5 digits to the right of the decimal point.

EXAMPLE 2 – Similar enumerated value domains

Conceptual domain name: Countries of the world

Conceptual domain definition: Lists of current countries of the world represented as names or codes.

Value domain name (1): Country codes – 2 character alpha

Permissible values:

<AF, The primary geopolitical entity known as "Democratic Republic of Afghanistan">

<AL, The primary geopolitical entity known as "People's Socialist Republic of Albania">

. . .

<ZW, The primary geopolitical entity known as "Republic of Zimbabwe">

Value domain name (2): Country codes – 3 character alpha

Permissible values:

<AFG, The primary geopolitical entity known as "Democratic Republic of Afghanistan">

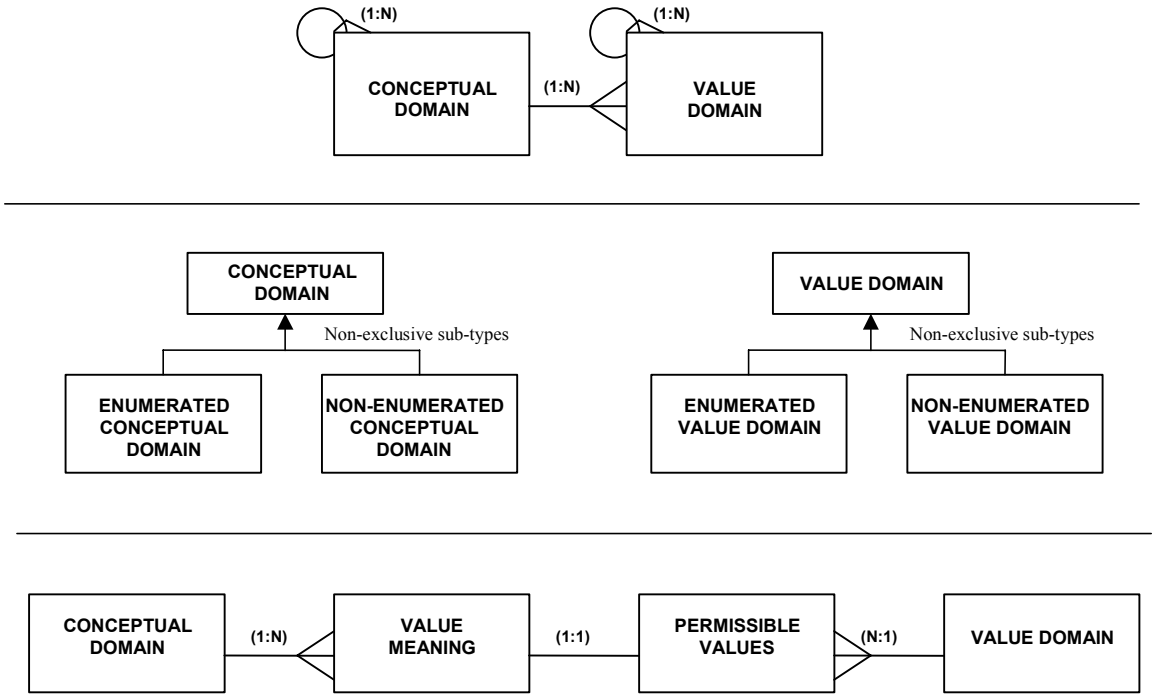
<ALB, The primary geopolitical entity known as "People's Socialist Republic of Albania">

. . .

<ZWE, The primary geopolitical entity known as "Republic of Zimbabwe">

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.

See ISO/IEC TR 20943-3, *Information technology — Procedures for achieving metadata registry content consistency — Part 3: Value domains* for detailed examples about the registration and management of value domains.



This figure is for informational purposes only. It is not normative.

Figure 3 — Fundamental model for value domains

6.5 Fundamentals of classification schemes

For the purposes of ISO/IEC 11179, a **classification scheme** is a concept system intended to classify objects. It is organized in some specified structure, limited in content by a scope, and designed for assigning objects to concepts defined within it. Concepts are assigned to an object, and this process is called classification. The relationships linking concepts in the concept system link objects that the related concepts classify. In general, any concept system is a classification scheme if it is used for classifying objects.

The content scope of the classification scheme circumscribes the subject matter area covered by the classification scheme. The scope of the classification scheme is the broadest concept contained in the concept system of the scheme. It determines, theoretically, whether an object can be classified within that scheme or not.

Concept systems, and classification schemes in particular, can be structured in many ways. The structure defines the types of relationships that may exist between concepts, and each classification scheme can be used for the purpose of linking concepts to objects. In a particular classification scheme, the linked concepts together with the other concepts related to the linked concept in the scheme provide a conceptual framework in which to understand the meaning of the object. The framework is limited by the scope of the classification scheme.

A concept system may be represented by a terminological system. The designations are used to represent each of the concepts in the system and are used as key words linked to objects for searching, indexing, or other purposes.

A special kind of concept system is a relationship system. There, the concepts are relationship types. A relationship type has N arguments, and it is called an n-ary relationship type. The statement "a set of N objects is classified by an n-ary relationship type" means that the N objects have a relationship among them of the given relationship type.

7 Metadata registries

7.1 General

Metadata is also data, so metadata may be stored in a database. A database of metadata that supports the functionality of registration is a **metadata registry** (MDR). A conceptual model of an MDR for describing data is provided in ISO/IEC 11179-3. The requirements and procedures for the ISO/IEC 11179 aspects of registration are described in ISO/IEC 11179-6. For actual metadata registries, there may be additional requirements and procedures for registration, which are outside the scope of ISO/IEC 11179. Rules and guidelines for providing good definitions and developing naming conventions are described in ISO/IEC 11179-4 and ISO/IEC 11179-5, respectively. The role of classification is described in ISO/IEC 11179-2. Recommendations and practices for registering data elements are described in ISO/IEC TR 20943-1. Recommendations and practices for registering value domains are described in ISO/IEC TR 20943-3.

An MDR contains metadata describing data constructs. The attributes for describing a particular data construct (e.g., data elements) are known, collectively, as a metadata object. When the attributes are instantiated with the description of a particular data construct, they are known as a metadata item. Registering the metadata item (i.e., entering the metadata into the MDR) makes it a registry item. If the registry item is also subject to administration (as in the case of a data element), it is called an administered item.

NOTE In common parlance, registering a metadata item describing a data construct is known as registering that data construct. Actually, the data construct is not stored in the MDR, its description is. This is analogous to the registries maintained by governments to keep track of motor vehicles. A description of each motor vehicle is entered in the registry, but not the vehicle itself. However, people say they have registered their motor vehicles, not the descriptions.

7.2 Overview model for an ISO/IEC 11179 MDR

The conceptual model for an ISO/IEC 11179 MDR contains two main parts: the conceptual level and the representational (or syntactical) level. The conceptual level contains the classes for the *data element concept* and *conceptual domain*. Both classes represent concepts. The representational level contains the classes for *data element* and *value domain*. Both classes represent containers for data values.

Clause 6 contains descriptions of each of the classes represented in Figure 4.

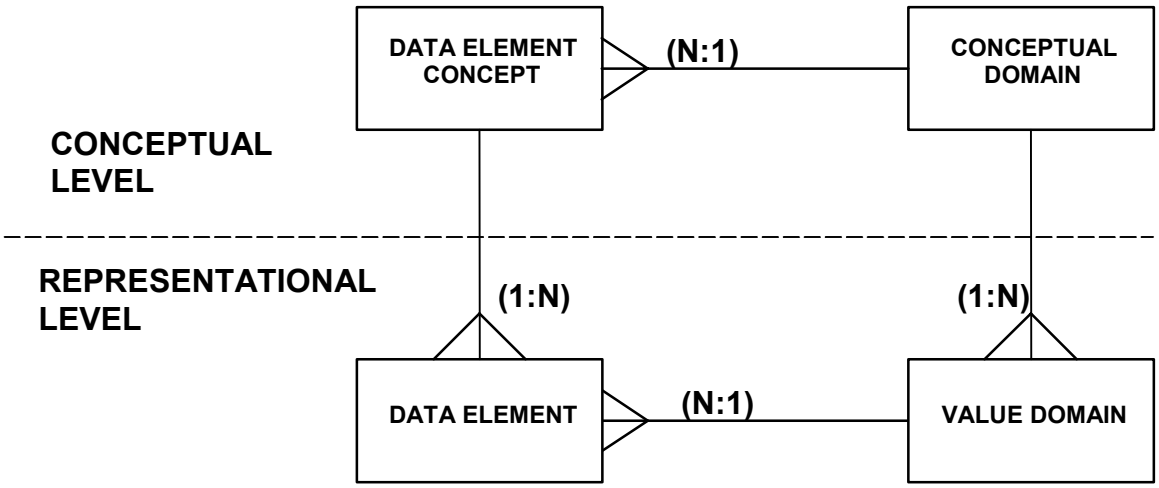


Figure 4 — Overview Model for ISO/IEC 11179 Metadata Registry

Figure 4 pictorially represents several fundamental facts about the four classes:

— A data element is an association of a data element concept and a representation (primarily a value domain)

— Many data elements may share the same data element concept, which means a DEC may be represented in many different ways

— Data elements may share the same representation, which means that a value domain can be reused in other data elements

— Value domains do not have to be related to a data element and may be managed independently

— Value domains that share all the value meanings of their permissible values are conceptually equivalent, so share the same conceptual domain

— Value domains that share some of the value meanings of their permissible values are conceptually related, so share the same conceptual domain in the concept system of conceptual domains that contain their respective conceptual domains

— Many value domains can share the same conceptual domain

— A data element concept is related to a single conceptual domain, so all the data elements sharing the same data element concept share conceptually related representations

Many other facts are not illustrated in Figure 4, but some of these are described in Clause 6. Two facts not described in Figure 4 are worth stating:

— Relationships among data element concepts may be maintained in an MDR, which implies that a concept system of data element concepts may be maintained

— Relationships among conceptual domains may be maintained in an MDR, which implies that a concept system of conceptual domains may be maintained

Some fundamental issues of registration and administration of metadata in an MDR are described later in this clause.

7.3 Fundamentals of registration

The registration and administration functions specified in ISO/IEC 11179-6 are what separate an MDR from a database of metadata. The means to accomplish these functions are a large part of the design of the metamodel specified in ISO/IEC 11179-3.

Registration is the set of rules, operations, and procedures that apply to an MDR. A detailed description of registration as it applies in ISO/IEC 11179 is found in ISO/IEC 11179-6. The three most important outcomes of registration are the ability to monitor the quality of metadata, provenance (the source of the metadata), and the assignment of an identifier to each object described in an MDR. Registration also requires a set of procedures for managing a registry, submitting metadata for registration of objects, and maintaining subject matter responsibility for metadata already submitted. For actual implementations of a metadata registry, there may be additional requirements, which are outside the scope of ISO/IEC 11179.

Each administered item is maintained in a uniform and prescribed manner. Identifiers, quality measures, responsible organizations, names, and definitions are all part of the general metadata that falls under administration. Registration is the process of creating or maintaining administrative and other detailed metadata.

Metadata quality is monitored through the use of a **registration status**. The status records the level of quality. Each level is specified in ISO/IEC 11179-6. Every administered item is assigned a registration status, and this status may change over time. In addition, metadata quality is multi-faceted. That is, there are several purposes to monitoring metadata quality. The main purposes are

— Monitoring adherence to rules for providing metadata for each attribute

— Monitoring adherence to conventions for forming definitions, creating names, and performing classifications

— Determining whether an administered item still has relevance

— Determining the similarity of related administered items and harmonizing their differences

— Determining whether it is possible to ever get higher quality metadata for some administered items

The rules for creating and assigning identifiers are described in ISO/IEC 11179-6. Each administered item within an MDR is assigned a unique identifier.

The **registration authority** is the organization responsible for setting the procedures, administering, and maintaining an MDR. The **submitting organization** is responsible for requesting that a new metadata item be registered in the registry. The **steward** is responsible for the subject matter content of each registered item. Each of these roles is described in ISO/IEC 11179-6.

8 Overview of ISO/IEC 11179

8.1 Introduction of parts

This sub-clause introduces each part of the multi-part standard ISO/IEC 11179. It summarizes the main points and discusses the importance of each.

8.1.1 Part 1

ISO/IEC 11179-1, *Framework*, introduces and discusses fundamental ideas of data elements, value domains, data element concepts, conceptual domains, and classification schemes essential to the understanding of this set of standards and provides the context for associating the individual parts of ISO/IEC 11179.

8.1.2 Part 2

ISO/IEC 11179-2, *Classification*, provides a conceptual model for managing concept systems used as classification schemes. Concepts from these schemes are associated with administered items through the process of classification. Librarians, terminologists, linguists, and computer scientists are perfecting the classification process, so it is not described here. The additional semantic content derived from classification is the important point.

Associating an object with one or more concepts from one or more classification schemes provides

— Additional understanding of the object

— Comparative information across similar objects

— Understanding of an object within the context of a subject matter field (defined by the scope of a classification scheme)

— Ability to determine slight differences of meaning between similar objects

Therefore, managing classification schemes is an important part of maximizing the information potential within an MDR. ISO/IEC 11179-2 provides the framework for this.

8.1.3 Part 3

ISO/IEC 11179-3, *Registry metamodel and basic attributes*, specifies a conceptual model for an MDR. It is limited to a set of basic attributes for data elements, data element concepts, value domains, conceptual domains, classification schemes, and other related classes. The basic attributes specified for data elements in ISO/IEC 11179-3:1994 are included in this revision.

The registry metamodel is expressed in the Unified Modeling Language. It is divided into six regions for readability. All the provisions represented in the model are described in the text. Several provisions represented in comment boxes in the diagrams are described in the text.

The document contains a dictionary of all the modeling constructs (classes, attributes, and relationships) used in the model. This collection of attributes are known as the "basic attributes". All the attributes described in Parts 2, 4, 5, and 6 are contained in the registry metamodel.

The registry metamodel is not a complete description of all the metadata an organization may wish to record. So, the model is designed to be extended if required. However, extensions are, by their nature, not part of the standard.

A clause describing conformance criteria is provided. Conformance is described as either strictly conforming (all provisions met) or conforming (all provisions met, but additional provisions may exist).

8.1.4 Part 4

ISO/IEC 11179-4, *Formulation of data definitions*, provides guidance on how to develop unambiguous data definitions. A number of rules and guidelines are presented in ISO/IEC 11179-4 that specify exactly how a data definition should be formed. A precise, well-formed definition is one of the most critical requirements for shared understanding of data; well-formed definitions are imperative for the exchange of information. Only if every user has a common and exact understanding of the data can it be exchanged trouble-free.

The usefulness of definitions is one aspect of metadata quality. Following the rules and guidelines provided in Part 4 helps establish this usefulness.

8.1.5 Part 5

ISO/IEC 11179-5, *Naming and identification principles*, provides guidance for the designation of administered items. Designation is a broad term for naming or identifying a particular data construct.

Names are applied to data constructs through the use of a naming convention. Naming conventions are algorithms for generating names within a particular context. There are semantic, syntactic, and lexical rules used to form a naming convention. Names are a simple means to provide some semantics about data constructs, however the semantics are not complete. Syntactic and lexical rules address the constituents (e.g., allowable characters), format, and other considerations.

Data constructs may be assigned multiple names, and one may be identified as preferred. Usually, each assigned name is used within the context for which it was created.

Identifiers are designations meant for dereferencing data constructs for metadata management and exchange. An RA assigns a unique identifier for each data construct, unique within the context of the registry. Thus, identifiers are assigned to data constructs for use in unambiguously locating data constructs.

8.1.6 Part 6

ISO/IEC 11179-6, *Registration*, provides instruction on how a registration applicant may register a data construct with an RA and the assignment of unique identifiers for each data construct. Maintenance of administered items already registered is also specified in this document. Registration mainly addresses identification, quality, and provenance of metadata in an MDR.

An administered item identifier is formed by the combination of the RA Identifier, the unique identifier assigned to the administered item within an RA, and the version. Each registry is maintained by an RA, to which data constructs logically and functionally belong. For example, data constructs related to chemical matter would likely be registered under a Chemical Manufacturer Registration Authority

Registration is more complex than a simple indication whether a metadata item is either registered or not. Although it is tempting to insist that only "good" metadata may be registered, that is not practical. Therefore, improvement in the quality of administered items is divided into levels called registration status. In addition, there are status levels for administration between each of these quality levels. Collectively, these status levels are called administrative status. They indicate the point in the registration life cycle currently attained for an administered item.

The provenance of metadata and the chain of responsibility are managed in an MDR, too. The tasks and roles of the registration authority, data steward, responsible organization, and submitting organization are described. A framework for the registration process to be used in an MDR is provided.

Registration is both a process and a goal. The assignment of an identifier, quality status, life-cycle status, and describing provenance are goals. The rules by which these goals are accomplished is the process.

8.2 Basic principles for applying ISO/IEC 11179

Each part of ISO/IEC 11179 assists in a different aspect of metadata creation, organization, and registration; and each part shall be used in conjunction with the other parts. ISO/IEC 11179-1 establishes the relationships among the parts and gives guidance on their usage as a whole. ISO/IEC 11179-3 specifies metadata items a registration applicant shall provide for each object to be registered. Detailed characteristics of each basic attribute are given. Because of their importance in the administration of metadata describing data constructs, three of the attributes (name, definition, and identification) are given special and extensive treatment in two documents. ISO/IEC 11179-4 shall be followed when constructing data definitions. Identification and naming shall follow principles set forth in ISO/IEC 11179-5. ISO/IEC 11179-2 specifies a set of attributes for use in the registration and administration of classification schemes and their components. Metadata items are registered as registry items and administered as administered items in an MDR. ISO/IEC 11179-6 provides guidance on these procedures.

9 Conformance

There are no specific conformance criteria for this part of ISO/IEC 11179. It is a framework that ties the other parts of ISO/IEC 11179 together. As such, conformance is not an issue for this part. Each of the other parts of ISO/IEC 11179 has its own conformance clause.

Bibliography

[1] Barr, Avron & Feigenbaum, Edward A., *The Handbook of Artificial Intelligence* (3 Volumes), William Kaufman, Inc., 1981

[2] Gosh, Sakti, P., *Data Base Organization for Data Management*, Second Edition, New York: Academic Press, Inc., 1986

[3] Herman, G.T., "Theory of Algorithms", in *Encyclopedia of Computer Science and Engineering*, by Ralston, Anthony & Reilly, Jr., Edwin D., Second Edition, Van Nostrand Reinhold Company Inc., 1983.

[4] ISO/IEC Guide 2, *Standardization and related activities — General vocabulary*

[5] ISO 639 (all parts), *Codes for the representation of names of languages*

[6] ISO/IEC 646, *Information technology — ISO 7-bit coded character set for information interchange*

[7] ISO/IEC 2375, *Information technology — Procedure for registration of escape sequences and coded character sets*

[8] ISO/IEC 2382 (all parts), *Information technology — Vocabulary*

[9] ISO 2788, *Documentation — Guidelines for the establishment and development of monolingual thesauri*

[10] ISO 3166:1988, *Codes for the representation of names of countries*

[11] ISO/IEC 4873, *Information technology — ISO 8-bit code for information interchange — Structure and rules for implementation*

[12] ISO 5964, *Documentation — Guidelines for the establishment and development of multilingual thesauri*

[13] ISO 6093, *Information processing — Representation of numerical values in character strings for information interchange*

[14] ISO 6862*, Information and documentation — Mathematical coded character set for bibliographic information interchange*

[15] ISO/IEC 6937, *Information technology — Coded graphic character set for text communication — Latin alphabet*

[16] ISO/IEC 7498-1, *Information technology — Open Systems Interconnection — Basic Reference Model: The Basic Model*

[17] ISO 8601, *Data elements and interchange formats — Information interchange — Representation of dates and times*

[18] ISO/IEC 8824 (all parts), *Information technology — Abstract Syntax Notation One (ASN.1)*

[19] ISO/IEC 8859 (all parts), *Information technology — 8-bit single-byte coded graphic character sets*

[20] ISO/TR 9007, *Information processing systems — Concepts and terminology for the conceptual schema and the information base*

[21] ISO/IEC TR 9789, *Information technology — Guidelines for the organization and representation of data elements for data interchange — Coding methods and principles*

[22] ISO/IEC 10027, *Information technology — Information Resource Dictionary System (IRDS) framework*

[23] ISO/IEC TR 10032, *Information technology — Reference Model of Data Management*

[24] ISO/IEC 10241:1992, *International terminology standards — Preparation and layout*

[25] ISO/IEC 10646, *Information technology — Universal Multiple-Octet Coded Character Set (UCS)*

[26] ISO/IEC 11404:1996, *Information technology — Programming languages, their environments and system software interfaces — Language-independent datatypes*

[27] ISO/IEC TR 20943 (all parts), *Information technology — Procedures for achieving metadata registry content consistency*

[28] *Manual for Data Administration*, edited by Judith J. Newton and Daniel C. Wahl, National Institute of Standards and Technology, 1993

[29] Tasker, Dan, *Fourth Generation Data*, Prentice Hall, 1989

ICS 35.040

Price based on 23 pages