Committee Draft ISO/IEC CD2 11179-1	
Date: **2013-07-17**	Reference number: ISO/JTC 1/SC 32**N2413**
Supersedes document: 32N2224	

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)	Circulated to P- and O-members, and to technical committees and organizations in liaison for voting (P-members only) by: **2013-10-17** Please return all votes and comments in electronic form directly to the SC 32 Secretariat by the due date indicated.

ISO/IEC CD2 11179-1:2013(E)

Title: Information Technology -- Metadata Registries (MDR) - Part 1: Framework Ed 3

Project: 1.32.15.03.01.00

Introductory note:

The attached document is hereby submitted for a 3-month letter ballot to the NBs of ISO/IEC JTC 1/SC 32. The ballot starts 2013-07-17. Disposition of comments on the previous CD is 32N2414. New edition authorized at Berlin 2012-06-08.

Medium: E

No. of pages: 47

Dr. Timothy Schoechle, Secretary, ISO/IEC JTC 1/SC 32
Farance Inc *, 3066 Sixth Street, Boulder, CO, United States of America
Telephone: +1 303-443-5490; E-mail: Timothy@Schoechle.org
available from the JTC 1/SC 32 WebSite http://www.jtc1sc32.org/
*Farance Inc. administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI

ISO/IEC JTC1 SC32 N

Date: 2013-07-15

ISO/IEC CD2 11179-1

ISO/IEC JTC1 SC32 WG2

Secretariat: ANSI

Information technology — Metadata registries (MDR) — Part 1: Framework

Technologies de l'information —Registre de métadonnées (RM) — Partie 1: Cadre de reference

Document type: International standard
Document subtype:
Document stage: (20) Committee
Document language: E

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

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

The main task of technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member 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 shall not be held responsible for identifying any or all such patent rights.

ISO/IEC 11179-1 was prepared by Technical Committee ISO/IEC JTC1, *Information Technology*, Subcommittee SC32, *Data Management and Interchange*.

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 principles*

— *Part 6: Registration*

Introduction

The International Standard 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 the standard 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 this International Standard 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 help to 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 kind 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 concepts, not necessarily finite, where the concepts represent the meaning of the permissible values in a value domain. A value domain contains the allowed values for a data element.

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

New to Edition 3 of ISO/IEC 11179 is the introduction of concepts and concept systems in the description of the semantics of data. Object classes, properties, DECs, value meanings, and CDs are concepts. Therefore, they have definitions and may be designated by names or codes. They may also be organized through the use of relations among them into concept systems. A classification scheme is a concept system that is used for classifying some objects, and classification of an object adds meaning to that object.

Features needed for formal reasoning are also new to Edition 3. Applying the rules of some form of formal logic (1st order logic, predicate calculus, description logic, etc) may add additional abilities to query and reason with concept systems. Ontologies are concept systems that allow the application of formal logic, and Edition of ISO/IEC 11179 provides for their use.

The representational component is about the permitted values a data element may use. Each such permissible value is a designation of one of the concepts in the CD. The set of these permissible 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 Part 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 – *Metadata registries*, 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 among 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. This International Standard 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/IEC Guide 2, *Standardization and related activities — General vocabulary*

ISO 704:2000, *Terminology work – Principles and methods*

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

ISO 2382 (all parts), *Information processing systems*

ISO/IEC 10241:1992, *International Terminology Standards – Preparation and layout*

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

ISO/IEC 11404: 1996, *Information technology – Language independent datatypes*

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

3 Terms and definitions

For the purposes of this document, the following terms, abbreviations, 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 1 A name should be used as an identifier because it is not linguistically neutral.

NOTE 2 It is possible to define an identifier from the point of view of terminology as defined in ISO 1087 and described in ISO 704, as follows: representation of an object by a sign which denotes it, and is intended for dereferencing that object. Note the parallel with the definition of designation (3.2.9), except this applies to any object rather than just for concepts.

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 International Standard, 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 1 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 1 Data can be processed by humans or by automatic means.

[ISO 2382-1:1993, 01.01.02]

NOTE 2 Data may also be defined using the terminological notions defined in ISO 1087-1:2000 and the computational notions defined in ISO/IEC 11404 (General purpose datatypes). Define datum as follows: designation of a concept with a notion of equality defined for that concept.

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 Please observe that 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 **designation**s 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

NOTE A classification scheme is a **concept system** used for classifying some **object**s.

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 The definition is not the same as in 11179-3. The term is used in this Part as defined here.

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
described conceptual domain
conceptual domain that is not specified by a list of all valid **value meaning**s

3.3.13
described conceptual domain description
description or specification of a rule, reference, or range for a set of all **value meaning**s for the **conceptual domain**

3.3.14
described value domain
value domain that is specified by a description rather than a list of all **permissible value**s

3.3.15
described value domain description
description or specification of a rule, reference, or range for a set of all permissible value**s for the** value domain

3.3.16
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 $C^o = (5/9)*(F^o - 32)$ and $F^o = (9/5)*(C^o) + 32$.

3.3.17
enumerated conceptual domain
conceptual domain that is specified by a list of all its **value meaning**s

3.3.18
enumerated value domain
value domain that is specified by a list of all its **permissible value**s

3.3.19
international code designator
ICD
identifier of an organization identification scheme

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

3.3.20
item identifier
identifier for an item

3.3.21
item registration authority identifier
identifier of the **registration authority** registering the item

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**

[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**

NOTE A permissible value, the pairing of a **value** and **value meaning**, is a **designation**. The **value** is the **sign** and the **value meaning** is the **concept**.

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**

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

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

3.3.40
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 this International Standard 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
permissible value associated with a unit of measure

4 Abbreviations and acronyms

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

This clause describes the concepts from the theory of terminology that are used in this International Standard. They are mostly taken from ISO 704 - *Principles and methods of terminology* and ISO 1087-1 – *Terminology work – Vocabulary* – Part 1: *Theory and application*. 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 **characteristics**, which are also concepts. Any concept may be a characteristic; being a characteristic is a role for a concept. **Essential characteristics** are indispensable to understanding a concept, and they differentiate them, though which characteristics are essential depends on context. For instance, the concept *person* has sex, age, marital status, educational attainment, and race/ethnicity as essential characteristics in demography; however, it has name, sex, date/time of birth, height, weight, and mother's name as essential characteristics in a birth records system. For zoology, the essential characteristics of a person are different still. Other characteristics are **inessential**. The sum of characteristics for a concept is called its **intension**. The totality of objects a concept corresponds 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) is the representation of a concept by a sign, which denotes it.

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 Introduction

For this International Standard, **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 Part 3 (*Registry metamodel and basic attributes*) of this International Standard 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.

Informative Annex A contains a detailed description of the relationships among data, their descriptions, information, metadata, and meta-models.

6.2 Concepts

6.2.1 General

New to the 3rd edition of ISO/IEC 11179 is the notion of concepts; their definitions, designations, and relationships; their uses in the description of data; and their management in a MDR. This sub-clause gives a small introduction to the uses of concepts in describing data. Several data constructs used in ISO/IEC 11179 are concepts. They are data element concept, object class, property, conceptual domain, and value meaning. These are discussed in more detail in sub-clauses 6.3 and 6.5.

The semantics of data come from the concepts used in their descriptions. The meanings of all the concepts used to describe a datum are combined into a story, sometimes called a fact. This is equivalent to the information conveyed by some datum.

As ISO/IEC 11179-5 describes, the names for data elements, which may convey some of the semantics of their underlying data, can be constructed from the designations of their constituent concepts. So, for some datum, the story it conveys might be written as “The temperature in Washington, DC at the bottom of the Washington Monument on 14 June 2013 at 1600 ET was 78°F”. The designations of concepts (temperature; Washington, DC; Washington Monument, 1600 ET, and 78°F) are interspersed with English words to create a sentence, which contains the story.

Finally, the relationships some concepts have with others, as defined in a concept system and described in ISO/IEC 11179-2, add semantics to data. For instance, the concept of a temperature measurement is different if it is a measure of the kinetic activity of molecules of air in some location on Earth versus a measure of ambient infra-red radiation in inter-planetary space between Jupiter and Saturn. In both cases, instances of temperature are ultimately measures of infra-red radiation, but they are obtained far differently. The temperature of air is directly determined by the motion of molecules. There are far too few molecules in inter-planetary space for the same kind of measurement to be meaningful. A different sort of measurement is required.

6.2.2 Management

Looking across all the data elements found in an organization or across organizations, one finds many concepts that are the same. For instance, in statistical survey organizations, data are collected and estimates produced for some population. But surveys are often conducted on a regular basis – monthly, quarterly, yearly – so the population is repeated. Moreover, many surveys might be conducted on the same population, each for its own specialized purpose. A similar situation applies in a scientific research lab, where in a large program, the same scientific experiments are conducted repeatedly.

Since some of the purposes of the MDR are understanding, re-use, harmonization, and standardization of data, then managing meanings is critical for those needs. In the case of re-use in particular, where the same meanings are applied in different situations, it is inefficient, error prone, redundant, and inhibitory to store one concept multiple times. If the same concept is used to describe many data elements, describe it once, and re-use it.

This concept management capability is an important addition to the 3rd edition of ISO/IEC 11179. The case for why concept management is important is provided in this sub-clause.

6.3 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 a **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, and units of measure (if necessary).

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

The totality of objects for which we wish to collect and store data is the extension of an object class. Object classes 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 object classes. 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 totality 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 totality of objects corresponding to the object class has one member. Examples are concepts corresponding to single objects, such as "the collection of all persons" or "the collection of service sector establishments". Aggregate data are described this way. Examples of properties for these object classes are average income or total earnings.

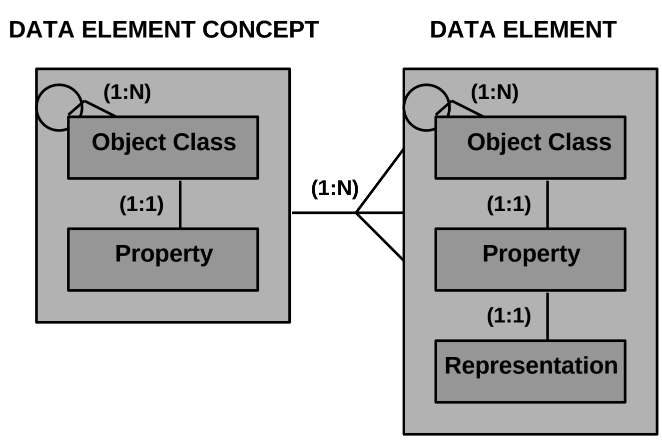
It is important to distinguish an actual object class or property from its name. This is the distinction between concepts and their designations. Object classes and properties are concepts; their names are designations. Complications arise because people convey concepts through words (designations), 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, for inistance.

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 – *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, and possibly the units of measure, are altered.

See ISO/IEC TR 20943-1:2002 – *Procedures for achieving metadata registry content consistency* – Part 1: *Data elements* for details about the registration and management of descriptions of data elements.



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

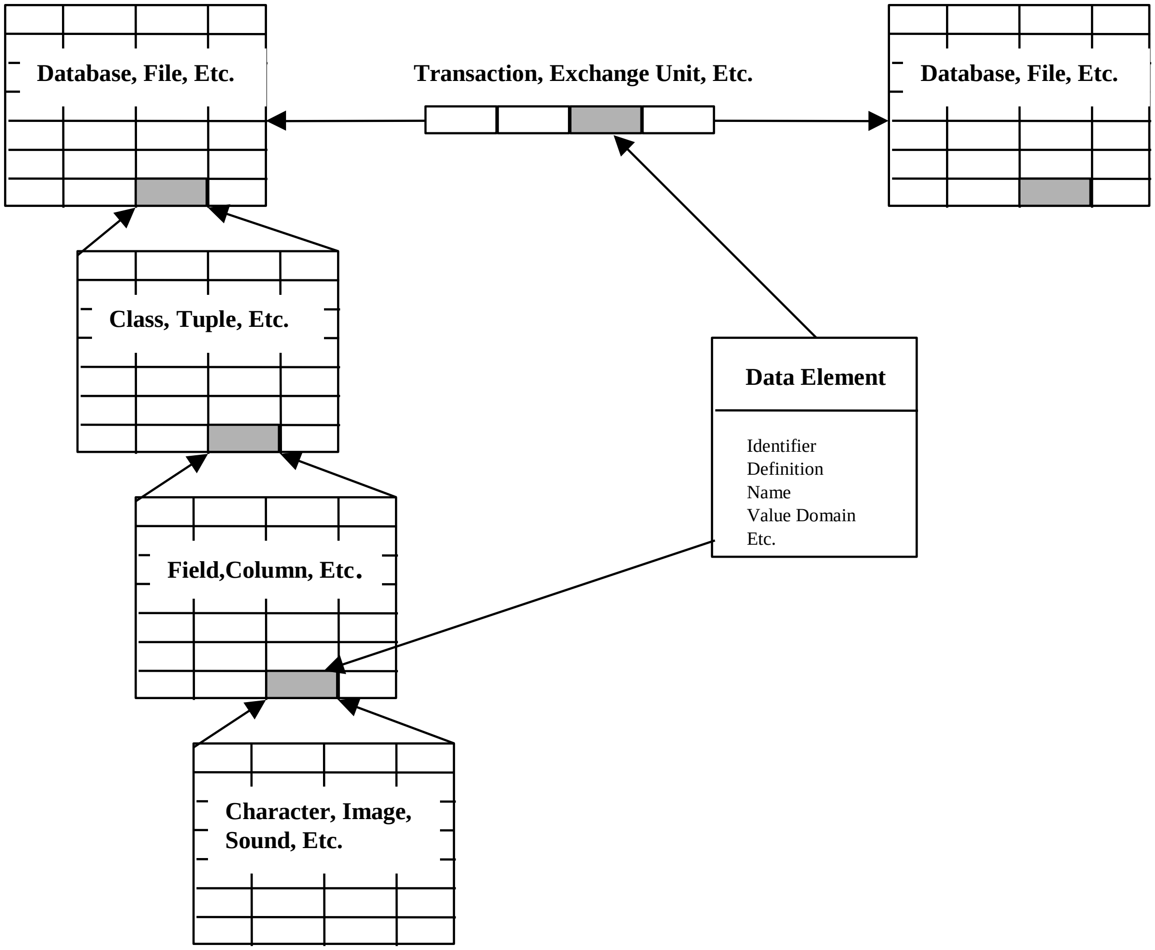
Figure 1: Fundamental model of data elements

6.4 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.



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

Figure 2: Data elements and other data concepts

6.5 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 the association 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)

— **Described 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 described value domain is specified by a description. The **described 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".

A **conceptual domain** is a set of value meanings. Each value domain is linked to a conceptual domain in the following way: the value meaning from each permissible value in the value domain is one of the value meanings in the linked conceptual domain. The intension of a conceptual domain is its value meanings. Many value domains may be linked to 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

— **Described 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 described conceptual domain are expressed using a rule, called a **described conceptual domain description**. Thus, the value meanings are listed implicitly. This rule describes the meaning of permissible values in a described value domain. This conceptual domain type corresponds to the described type for value domains. See ISO/IEC TR 20943-3 – *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. A pound is a measure of force, and a gram 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 described 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 expressed in decimal numerals 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 decimal digits

Value domain description: All real numbers expressed in decimal numerals 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.

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:

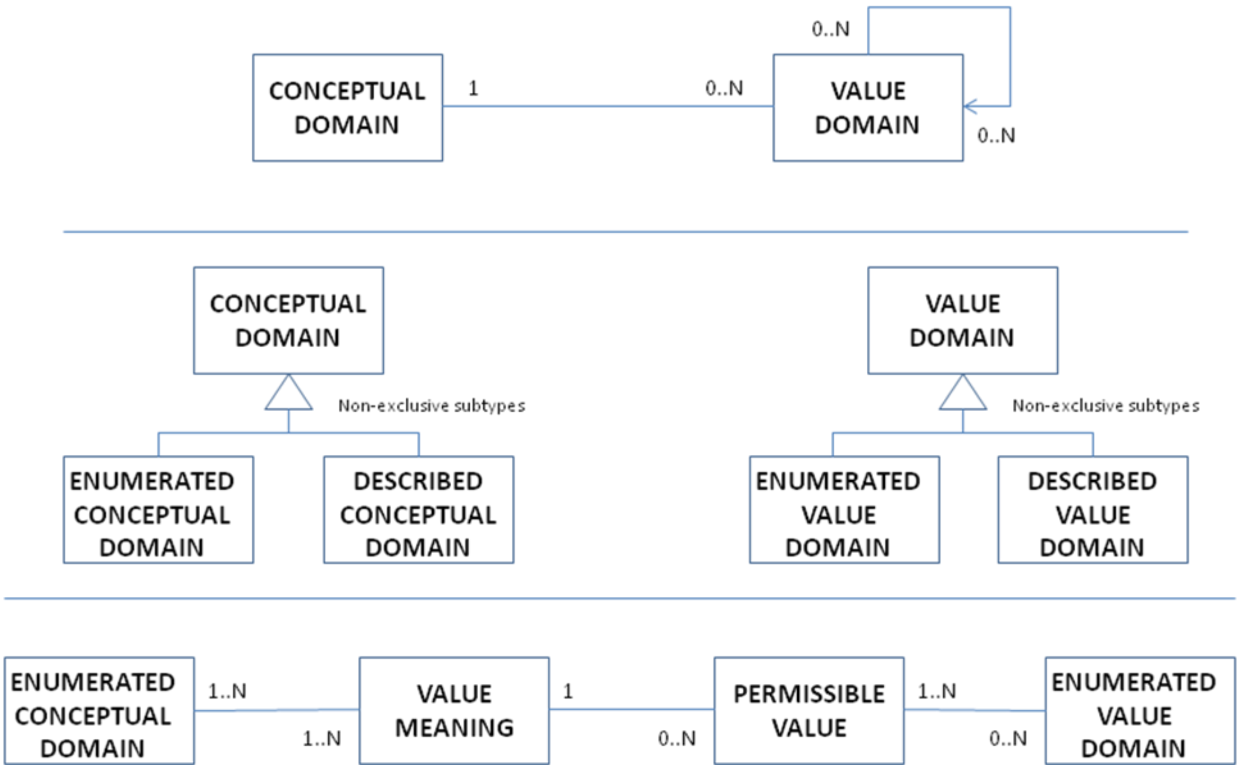
<AFG, The primary geopolitical entity known as "Democratic Republic of Afghanistan">

<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 part of 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.



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

Figure 3: Fundamental model of value domains

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

6.6 Fundamental model of concept systems

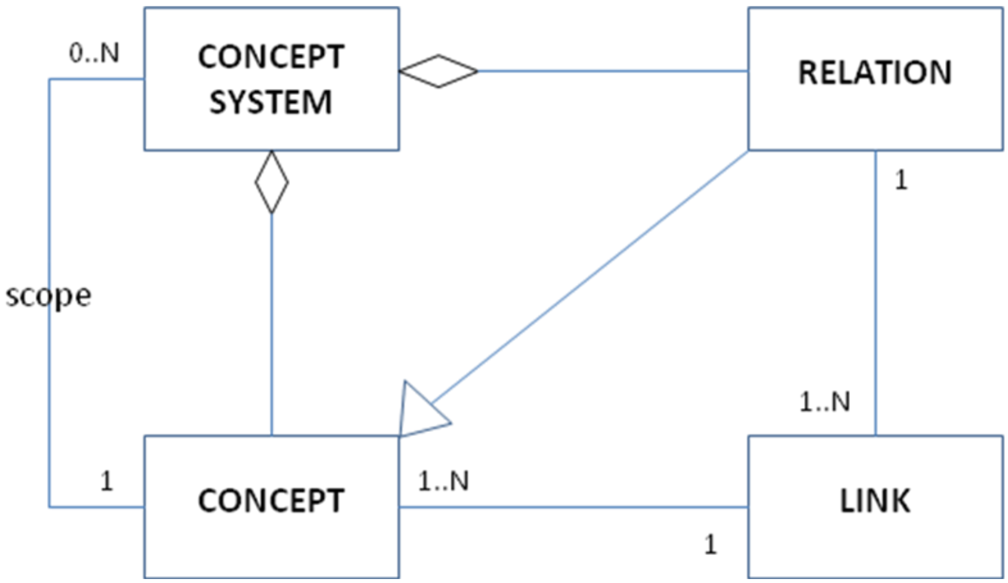
For the purposes of ISO/IEC 11179, a **classification scheme** is a concept system intended to classify objects. It is organized in some specified structure, limited in content by a scope, and designed for assigning objects to concepts defined within it. Concepts are assigned to an object, and this process 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.

Figure 4 illustrates the ideas conveyed in the following three paragraphs in this sub-clause. The figure itself is not normative, but it is used to illustrate the basic ideas.

Concept systems consist of concepts and relations among the concepts. The relations are a kind of concept, and they are types for the relationships that are established among particular sets of concepts. In ISO/IEC 11179-3, the relationships between concepts in a concept system are called links. Concept systems, and classification schemes in particular, can be structured in many ways. The structure is defined by the types of relationships that may exist between concepts.

A special kind of concept system is a relationship system. The statement "a set of N objects is classified by an n-ary relation" means that the N objects have a relationship among them of the given relationship type, where the relationship of that type takes N arguments.

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.



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

Figure 4: Fundamental model of concept systems

A 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.

7 Metadata registries

7.1 Introduction

Metadata is also data, so metadata might 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 this International Standard. 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, data element concept, conceptual domain, and value domain) 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 5.

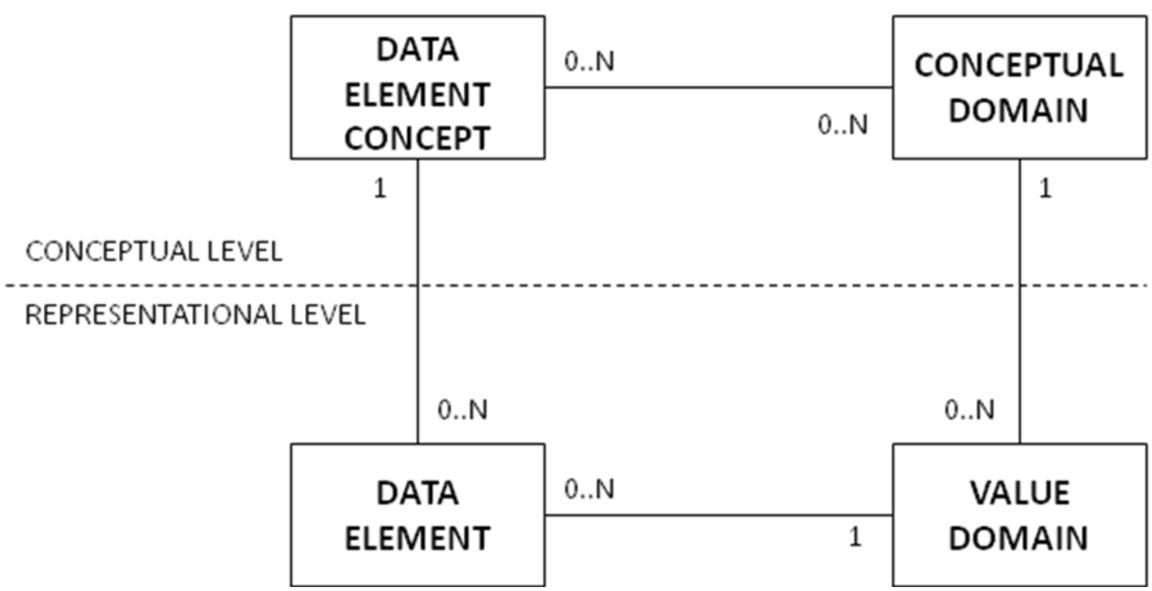


Figure 5: Overview Model for ISO/IEC 11179 Metadata Registry

Figure 5 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 may be related to many conceptual domains

There is one important rule the Figure 5 does not depict: Given a data element, the conceptual domain related to its data element concept shall be the conceptual domain of its value domain.

Many other facts are not illustrated in Figure 5, but some of these are described in Clause 6. Two facts not described in Figure 5 are worth stating:

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

— Relationships among conceptual domains may be maintained in an MDR, which implies that a concept system of conceptual domains might 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 this International Standard.

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 includes 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, Parts 1- 6

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, concepts, and concept systems 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, which might be 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, concept systems, 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 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 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.

The aim for any naming convention is to allow development of names for items that are clear and transparent in meaning, concise, demanding minimal effort of interpretation by the end user, and subject to the constraints of the system under which the items are processed. A naming convention can be used to form names by which information about the data is expressed. Ideally, the names resemble short summaries of the formal definition of the information being named.

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, the chain of responsibility is 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, Parts 1-6

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 this International Standard. ISO/IEC 11179-1 is a framework that ties the other parts of the standard together. As such, conformance is not an issue for this Part. Each of the other Parts 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 639:1988, *Code for the representation of names of languages*

[5] ISO 646, *ISO seven-bit coded character set for information interchange*

[6] ISO 646:1983, *Information interchange - ISO 7-bit coded character set for information interchange*

[7] ISO 2375, *Procedures for registration of escape sequence*

[8] ISO 2788:1986, *Documentation - Guidelines for the establishment and development of monolingual thesauri*

[9] ISO 3166:1988, *Code for the representation of names of countries ISO 4873, ISO eight-bit code for information interchange - structure and rules for implementation*

[10] ISO 5985:1985 *Documentation - Guidelines for the establishment and development of multilingual thesauri*

[11] ISO 6093, *Representation of numeric values in character strings*

[12] ISO 6093:1985, *Information processing - Representation of numerical values in character strings for information interchange*

[13] ISO 6862:1996*, Mathematical coded character set for bibliographic information interchange*

[14] ISO 6937, *Coded character sets for text communication*

[15] ISO 7498:1984, *Reference model of open systems interconnection*

[16] ISO 8601:1988, *Data elements and interchange formats; - Representation of dates and times*

[17] ISO 8824:1989, *Abstract Syntax Notation One*

[18] ISO 8859:1992, *Eight-bit single-byte coded graphic characters sets*

[19] ISO 9007:1987, *Information processing systems - Concepts and terminology for the conceptual schema and the information base*

[20] ISO/IEC TR 9789:1993, *Information technology - Guidelines for the organization and representation of data elements for data interchange -- Coding methods and principles*

[21] ISO/IEC 10027:1990, *Information technology - Information Resource Dictionary System (IRDS) Framework*

[22] ISO/IEC TR 10032:2003, *Information technology - Reference model of data management*

[23] ISO/IEC 10646-1:1993, *Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and basic multilingual plane*

[24] Langefors, Börje, *Essays on Infology*, Studentlitteratur, 1995

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

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

Annex A

(Informative)

Data, Metadata, and Meta-Models

A.1 Introduction

ISO/IEC 11179 specifies the classes of metadata needed to describe data, and these specified classes are organized into a model, called the meta-model. This Informative Annex describes the relationships between data and metadata and between data and the MDR meta-model, and these relationships provide a deeper understanding of ISO/IEC 11179.

Since metadata are defined as data defining and describing other data, then an understanding of data and how metadata are related to data will enhance the understanding and usages of ISO/IEC 11179.

A.2 Data

A.2.1 Definition

ISO/IEC 2382, sub-clause 01.01.02, defines data as “reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing”. Upon inspection of this definition, the fundamental characteristic of data is that they are representations of information. The other phrases and words in the definition are modifiers and behave as distinguishing characteristics.

Consider a typical example of a table of data from a database. See Table 1 for an illustration.

Name	Sex	Education	Age	Weight
Joe	M	5	52	81.6
Bill	M	2	27	68.4
Mary	F	1	33	56.7

Table 1: Illustration of Data in Database Table

Each cell in Table 1 contains a datum of some kind.

In the rest of Clause A.2, the definition of data in ISO/IEC 2382 is analyzed.

A.2.2 Representation of information

Since data are representations of information, then each cell in Table 1 contains such a representation, because each cell contains a datum. The representation in each cell in this case is in the form of a string of characters, depending on which column the cell is in; and these representations stand for, or denote, some information, which is the meaning of the datum in each cell.

The Table 1 provides several ways to look at how a representation encodes meaning. For instance, consider the row with “Joe” in the name column. There, the character M in the sex column denotes the male sex. The numeric character 5 denotes an educational attainment of a graduate degree. The numeral 52 means the person is fifty-two years old. The numeral 81.6 means the person weighs eighty-one and six-tenths kilograms. As ISO/IEC 11179-3 shows, much more meaning might be attached than what is illustrated here, and 11179-3 may not provide for all the meaning an application needs.

As described in Sub-Clause 6.2.1, the information conveyed by a datum is (partially) contained in the meanings of ISO/IEC 11179 data constructs that are concepts. In particular, each datum is a permissible value from some value domain as described in Sub-Clause 6.5, and the meaning part of a permissible value includes the value meaning, which is a kind of concept. So, the information conveyed by a datum is in the form of meanings of concepts, and the representation of that information is the other part of a permissible value, called value in 11179-3. This means a datum in Table 1 is a representation of a concept (value meaning) by some alphanumeric string which denotes it.

More generally, representations might be alphanumeric strings, bit-maps, or any other perceivable object (see 3.2.22). This is what is meant by a sign; see Clause 5. Therefore, substituting sign for “alphanumeric string” in the last sentence of the previous paragraph, we see that any datum is a designation, as defined in ISO 1087-1. See also Clause 5.

A.2.3 Caveat

A datum cannot be just any designation, however. There must be delimiting characteristics that distinguish data from terminology in general. Going back to the ISO/IEC 2382 definition, data are “suitable for communication, interpretation, or processing”, they are “reinterpretable”, and they are representations in a “formalized manner”. Analyzing these parts of the definition will uncover the delimiting characteristics.

A.2.4 Interpretation

Data are interpretable, i.e., capable of being understood, because interpretation is the process of going from a representation to its underlying meaning, as described further here. See also Clause A.3. A datum results from the determination of a property of an object, where the term property is understood by how the term is used in ISO 704, not ISO/IEC 11179. The property is itself a concept, and a designation for this concept is recorded. The sign for this designation is the representation discussed in Sub-Clause A.2.2. Because the representation is tied to a concept, it is capable being understood. The context under which a determination is made provides the extra meaning beyond that of the property. This will be discussed further in this Annex.

NOTE 1: The use of the word determination here is purposeful. Often, data are said to be observations, but many data are calculated or estimated from others (e.g., statistical estimates), measured by an instrument not in human control (e.g., the altitude or airspeed of an airplane), or generated by the application of some law, policy, or administrative program (e.g., US Social Security numbers). There may be others as well.

NOTE 2: In applications of ISO/IEC 11179, it is expected that the representations, meanings, and context for some data are recorded in an MDR.

Data are reinterpretable, because the interpretation process does not change anything about them. That is, a representation is unaffected by interpretation. Finally, interpretation essentially concerns the terminological character of a datum. So, no new characteristics are uncovered.

A.2.5 Communication and processing

Communication and processing require a different kind of understanding, and here is where additional characteristics for data lie. Communication is about the conveyance of information through being able to move a datum from one computer storage medium to another. This is fundamental to the operations of almost any process carried out on a computer, and it requires the ability to make faithful copies of data. For example, one copies a datum from a flash drive to main memory to perform calculation on it. A faithful copy of a datum is determined by whether there is equality between the original and the copy. Therefore, the ability to determine equality is a necessary characteristic of data.

Processing, at its core, refers to some kind of manipulation of data. The ability to perform basic arithmetic and string operations are the fundamental building blocks of any operation that is allowed on data. These manifest themselves in the definition of a datatype for data. See ISO/IEC 11404 – *General purpose datatypes* for a deeper discussion. However, using the columns in Table 1, some typical datatypes for data are illustrated. These examples show the kinds of assumptions and operations that may be allowed for data. Again, this is not a general treatment, as the details can be found in ISO/IEC 11404.

There are 5 columns in Table 1, each with a label: name, sex, education (educational attainment), age, and weight. The 11404 datatype families appropriate for each column are as follows:

- <u>Column</u> <u>11404 Datatype Family</u>
- Name Character string
- Sex State
- Educational attainment Enumerated
- Age Natural number
- Weight Real

Each of these datatype families is defined through a set of axioms, called properties in ISO/IEC 11404, and a set of allowed operations, called characterizing operations in 11404. The operations follow from the axioms.

The equality axiom is true for every datatype family defined in 11404. Other axioms are added to allow for more complex operations. For example, State types are finite lists with only equality possible. Enumerated types are ordered finite lists. Character string types provide typical string manipulations. Natural number types provide for the operations allowed on the Natural Numbers (i.e., no division), and Real types allow all arithmetic operations and taking roots.

The bottom line is each datatype family defined in 11404 provides a model or rules for the kinds of operations and processing allowed for some data.

A.2.6 Suitable formalized manner

What gets manipulated during processing? It is the signs used in the designations of data that are manipulated. The kinds of operations allowed are determined by the underlying concepts the signs represent. For instance, if the sign **1** designates the concept of the male sex, then not much can be done with it. The concepts male and female do not have any obvious arithmetic associated with them. If, on the other hand, it designates the real number one, then many operations and arithmetic properties can be assumed for it. This implies that computers process signs rather than data, but the manner of that processing depends on the datatype.

NOTE: The use of the word property in the paragraph above is intended to have its common English meaning. This represents the 4th different usage of the sign in this Annex.

The signs themselves need to be regularized (i.e., formalized) in some sense so that processing happens consistently. Otherwise, computers will not be able to make sense of them. Humans recognize that 3 and 3 are the same in some sense – they both commonly designate the number three. Such regularized signs in computers are, for example, 16-bit strings used to encode any character in a character set in use in the world. The character set supplies the underlying concept for each of the allowed bit combinations. So, for example, the simple arithmetic problem of two plus two is visualized through use of signs as 2+2, and the human familiarity with that notation makes it easy to arrive at the answer, four (or 4 visualized).

A.2.7 Signs

Children are taught early in school that a numeral and a number are not the same. Numerals are what are written down or perceived. Numbers are concepts; they are units of knowledge or thought. Therefore, numerals are signs used to designate numbers. For the number three, the signs 3 and 3 both designate it. They each could designate other concepts as well. In any case, they are examples of the same (Arabic) numeral. What is it that allows people to say these signs are the same?

Without getting into a deep philosophical discussion about signs, it should be clear the idea of a particular numeral is a concept as well. The concept of the Arabic numeral 3 might be defined, roughly, as two approximately semi-circular shapes, both open to the left, placed vertically so the bottom end of the upper one merges with the top end of the lower one. Other numerals, including Roman numerals, might be like-wise defined. In fact, this idea generalizes, and all signs are really concepts with perceivable objects belonging to their extensions. The perceivable objects are what are used to refer to concepts.

Plainly, 3 and 3 are perceivable objects and are both in the extension of the concept of the numeral 3. Other signs behave similarly. In fact, every numeral, letter, and word of text in this International Standard is a sign, yet each could be written in a different font or font size. These alphanumeric strings are signs designating concepts.

A.2.8 Examples

The following examples illustrate how the ideas presented in this Clause provide a rich description of data. Here, we depict a hierarchy of signs and concepts to describe how computers and the humans that use them encode and make use of data.

Example 1 – Computers are electronic machines that operate through the use and detection of voltages. Voltages are perceivable objects, as they are detectable. What follows might not be the actual way any computer works, but the principle is the important point. Let the idea of a binary digit (bit) “0” (similar to a decimal numeral) be denoted by a voltage of zero volts and the idea of a bit “1” be denoted by the voltage of five volts. Thus, the voltages are signs and the bits they designate are concepts. Therefore, the permissible values in some value domain might be defined as follows:

<0V, bit “0”>

<5V, bit “1”>

Example 2 – From the definitions of bits in Example 1, let 0 denote the binary number zero and 1 denote the binary number one. Therefore, a set of permissible values in some new value domain might be defined as follows:

<0, binary number zero>

<1, binary number one>

Example 3 – Strings of bits may represent any number in base-2 notation. The range of numbers is limited by the number of bits available. As with the Arabic decimal notation, the least significant bit is written on the right, and each place subsequent to the left denotes the next higher power of two beginning with power zero. For instance, the number designated “55” in decimal notation has a binary representation of “110111”, which is interpreted as 1*32 + 1*16 + 0*8 + 1*4 + 1*2 + 1*1 (by means of the usual practice of inferring numbers from the decimal Arabic numeral notation). Therefore, the description in some value domain might be written as follows: Natural numbers designated by binary representation, with least significant bits on the right.

Example 4 – Strings of bits may also represent a character in some character set. It is outside the scope of this international standard to explain character sets, but it suffices to note that each character is assigned a natural number within some fixed range. For instance, in the ASCII character set, the number sixty five denotes the character “A”, and number ninety nine denotes the character “c”. Therefore, the *permissible values* in some *value domain* might be defined as given in http://en.wikipedia.org/wiki/ASCII (The page devoted to a description of ASCII at the Wikipedia web site.).

Example 5 – Strings of characters constitute words in natural language and terms in special languages. Underlying concepts are their meanings. For instance, the reserved words in programming languages are examples of such terms, such as while and switch in the C language.

A.3 Information

Again, ISO/IEC 2382 defines information as “knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning”. In this Annex, information that is conveyed by a datum is the limit of the discussion. Even so, the definition states information is a kind of knowledge concerning objects, and this knowledge has a particular meaning within some context. So, information is about the meaning of some objects, data in this case.

In sub-clause A.2.2, the interpretability of data was discussed. The result of interpretation is the meaning behind data. In this sub-clause, information is the meaning of some objects in context, data are objects, and data are observed under certain conditions (i.e., context). Therefore the interpretation of data leads to information.

In [24], information conveyed by data is described as the result of an interpretation of that data under certain circumstances. This is expressed by a function, the infological equation, defined as follows:

$I = i(D, S, t)$

where

I = information

i = the interpretation function

D = data

S = pre-knowledge, i.e., what an interpreter knows in advance

t = time

Here, the context under which information is interpreted is the time (t) and the pre-knowledge of the interpreter (S).

The infological equation, then, closes the circle between, data, meaning, and information. Knowing some information, it is possible to extract a meaning, and it is possible to record this meaning as a datum. Now, it is possible to go from the datum back to the information it conveys, via an interpretation using the infological equation.

A.4 Metadata

ISO/IEC 11179 defines metadata as “data that defines and describes other data”. However, this does not say how metadata arises or where it comes from. The infological equation provides the answer.

Suppose, the following string appears in some cell in a table of data:

2013-02-29-14-31-35

Further, suppose this cell means the following:

- The takeoff date/time of the president’s aircraft on February 29, 2013 was 35 seconds past 2:31 PM.

NOTE - In fact-based modeling, this statement is referred to as a fact.

This fact is the interpretation of, the meaning behind, the datum, and it is the result of applying the infological equation to the datum above. However, at any particular time, this might not reflect all that can be interpreted about this datum, and this why S (pre-knowledge) and t (time) are parameters to the function.

The meaning of the fact is information the datum conveys, and one interprets the data to obtain it. However, meaning and information are ideas humans carry around in their heads. The string of words whose meaning is called a fact above is actually a reification (a realization) of that meaning.

It is also true that the fact given above is a description of the datum. A description conveys the meaning of some object. This might not be all that one wants or needs to know about the datum, but more pre-knowledge will help uncover missing pieces. For our fact above, the place of the event, the type of aircraft, the local weather, the instrument used to record the time, and the reason the recording is made might all be relevant details in the interpretation. Each can be added to the fact as they are uncovered. For example, adding these additional pieces of information to the description might lead to the following:

- In clear, windless, cold (35° F) conditions, the takeoff date/time of the president's Boeing 747 aircraft as read (to the nearest 5 seconds) from the official clock in the control tower at Andrews Air Force Base near Washington, DC USA on February 29, 2013 was 35 seconds past 2:31 PM.

This statement is a description and meaning of the datum presented above. The statement is a sentence typed into this document, so it is rendered as data itself. Therefore, the statement is metadata.

The question then comes to mind as to how this statement could be organized. ISO/IEC 11179-3 provides a meta-model for organizing metadata (specific portions of the reified information). The following is a small subset of the attributes 11179-3 provides and the values the fact above provides:

Object class	takeoff of president's aircraft
Property	date/time
Value domain description	date and time using Arabic numerals
Format	yyyy-mm-dd-hh-mm-ss, where the first mm refers to month and the second refers to minutes, hours are in 24 hour format
Precision	nearest 5 seconds
Datatype	date-time

Some of the possible metadata in the description has no obvious attribute in the 11179-3 meta-model, such as the weather conditions, aircraft type, measuring device, and geographic location. This problem, and the general issue of selecting attributes, is discussed in the next sub-clause.

A.5 Factoring

Factoring is the process of taking a complex idea and breaking it into manageable conceptual pieces. In this sub-clause, factoring and how it relates to the use of meta-models is discussed.

A.5.1 Factoring data descriptions

First, the issue of how to factor a description so that it will meaningfully fit into the classes and attributes of a meta-model is discussed.

As seen in the previous sub-clause, the reified fact may contain many ideas (concepts and combinations of concepts) strung together. These ideas are instances (objects) of some classes; however the problem is to determine which ideas are instances of which classes.

Returning to the example in the last sub-clause, here is a list of the ideas in the detailed description:

Takeoff of president's aircraft

Date/time

cold (35° F)

windless

official clock

control tower at Andrews Air Force Base near Washington, DC USA

Boeing 747

February 29, 2013 35 seconds past 2:31 PM

Not all these ideas are atomic in the sense that they might meaningfully be broken into two or more other broader ideas.

NOTE – The use of the word "broader" here might be confusing. An idea that has many descriptors to it is highly specialized. It has a narrowed intension, as described in ISO 704. For instance, "takeoff of president's aircraft" is more specialized than "takeoff of aircraft". Removing extra descriptors broadens the concept left over.

Examples of non-atomic ideas from the list above and how they might be factored further follow here:

- Takeoff of president's aircraft
  - Takeoff; president's aircraft
  - Aircraft takeoff; president
  - Aircraft; takeoff; president
- February 29, 2013 was 35 seconds past 2:31 PM
  - Year; month; day; hour; minute; seconds
  - Date; time
- control tower at Andrews Air Force Base near Washington, DC USA
  - Andrews Air Force Base Control Tower; Washington, DC; USA
  - Control tower; Andrews Air Force Base; Washington; DC; USA

The main point is there is no canonical way to divide the ideas in a description. The information needs of analysts may determine how best to do this.

To explore this further, take the Object Class and Property from the 11179-3 meta-model and determine which part of the example description fits there. The Object Class is generally a description of the collection of objects for which data are determined. Typical examples are persons, business establishments, or educational institutions. The Property is generally a characteristic of the Object Class that can be determined. For instance, date and time or the date and time of takeoff are examples. Here, the relevant part of the description seems to be "takeoff date/time of president's aircraft". Some other details, such as Boeing 747, could be added, but this will be left out for simplicity.

There are several possibilities for both:

- Object Class　　　　　　　　Property
  - takeoff of president's aircraft　　　　date/time

o takeoff of aircraft	date/time (with “president’s” as a specialization)
o president’s aircraft	takeoff date/time
• aircraft	president’s takeoff date/time

As previously described, there is no right answer. It depends on what the use of the data will be as to how it should be factored. But, factoring, in the sense of this sub-clause, is the process of deciding which ideas belong to what classes and attributes in the meta-model.

A.5.2 Factoring, meta-models, and classification

The MDR meta-model provides classes and attributes in which to include descriptions of data. A description of some data, a fact as described in sub-clause A.4, is a set of ideas (including concepts) strung together. Fitting these ideas meaningfully into the classes and attributes of the meta-model is called factoring, as described in sub-clause A.5.1.

Classification is the process of assigning some object to the extension of a concept. The process is successful if the object really does correspond to the assigned concept. Usually, classification is done against a concept system, so objects of many different kinds can be compared once they are classified. For instance, the biological classification of living things is used in this way.

Since the classes and attributes in the MDR meta-model are concepts, then that meta-model is a concept system. Since, concepts are conceivable objects (as described in ISO 704), then they are objects, too, and subject to classification. Finally, since facts contain concepts and other ideas, then the factors constituting those facts might be classified as well.

This means that factoring as described in sub-clause A.5.1 is a kind of classification (a process, not a concept system). As a result, one can apply the same kinds of analyses to factoring that are applied to classification in general. In particular, there may be errors; there may be ambiguous situations with more than one adequate answer; and situations may arise with no adequate answer given the available concept system (i.e., the MDR).

All this means care must be applied when using the MDR, both from the metadata management perspective and the perspective of the user of data described by it. Two different organizations might register descriptions of equivalent data in the form of data elements, yet those data elements might look substantially different. Looked at in a different way, just because two data element descriptions differ does not mean they cannot be describing similar if not equivalent data.

ISO/IEC TR 20943-5 describes a means of comparing and harmonizing data by analyzing descriptions in MDRs. This Annex shows how these descriptions can fail to look the same and the situations that might cause this to happen.