WP.31
ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD work session on statistical metadata (METIS)
(Geneva, 9-11 February 2004)

Topic (iii): Metadata models and terminology

ISO/IEC 11179: Framework for a Metadata Registry

Contributed Paper

Submitted by the Bureau of Labor Statistics, United States[1]

1.0 Introduction

ISO/IEC 11179 – *Metadata registries* [1], is a standard for describing any kind of data and organizing the descriptions in a registry. Many NSO's are implementing this standard as part of their metadata management strategy, as describing data is one of the most important functions an NSO provides.

This paper provides a detailed description of the concepts defined and used in the standard. The purpose of this is to increase the general understanding of the specification and to promote its use. When NSO's describe their data in compatible ways, this increases the chance that an outside user will be able to contrast and compare data from these different sources. This in turn makes the data produced by the NSO much more useful.

After the introduction, the paper provides a list of special terms and definitions used in the paper and the standard. A detailed description of data elements and their components follows, and finally a metadata registry is described.

2.0 Terms and Definitions

2.1 Terminology

We describe some useful concepts from the theory of terminology here. They are taken from ISO 704 [2] and ISO 1087-1 [3]. The terms and definitions follow in a list below:

- **object** - something conceivable or perceivable
- **property** - attribute used to describe or distinguish an object (e.g., "Dan has blue-gray eyes" means "blue-gray eyes" is the property of Dan associated with the characteristic "eye color" of people.)

[1] Prepared by Daniel Gillman, Gillman.Daniel@BLS.Gov. The opinions expressed in this paper are those of the author and do not necessarily reflect the official policy of the Bureau of Labor Statistics.

- **concept** - mental constructs, units of thought, or unit of knowledge created by a unique combination of characteristics
- **characteristic** - abstraction of a property of a set of objects
- **essential characteristic** - characteristic that is indispensable to understanding a concept
- **delimiting characteristic** - essential characteristic used for distinguishing a concept from related concepts
- **intension** - sum of characteristics that constitute a concept
- **extension** - set of objects a concept refers to
- **definition** - expression of a concept through natural language, which specifies a unique intension and extension
- **designation** - representation of a concept by a sign, which denotes it
- **general concept** - concept with two or more objects that correspond to it (e.g., planet, tower)
- **individual concept** - concept with one object that corresponds to it (e.g., Saturn, Eiffel Tower)
- **generic relation** - relation between two concepts where the intension of one of the concepts includes that of the other and at least one additional delimiting characteristic
- **generic concept** - broader concept (with a smaller intension) in a generic relation
- **specific concept** - narrower concept (with a greater intension) in a generic relation
- **concept system** - set of concepts structured according to the relations among them.

2.2 ISO/IEC 11179

In this section we define terms that are related to ISO/IEC 11179. The terms and definitions follow in a list below:

- **value domain** - set of permissible values
- **permissible value** - ordered pair of a value and its value meaning (e.g., <*value*, *value meaning*>)
- **conceptual domain** - set of value meanings
- **value meaning** - meaning of a value
- **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
- **property** - characteristic of an object class
- **data element concept** - specialization of an object class by a property
- **data element** - association of a data element concept and a value domain

In ISO/IEC 11179, the object class, property, data element concept, conceptual domain, and value meaning are all concepts from the terminological point of view. A property in ISO/IEC 11179 is a characteristic in terminology.

3.0 Fundamentals

3.1 Value Domains

A value domain is a set of permissible values. A permissible value is a combination of a value and the meaning for that value. The associated meaning is called the value meaning. A value domain is the set of valid or allowed values for one or more data elements.

A conceptual domain is a set of value meanings. It is a concept, and its value meanings are its characteristics. The value meanings may also be concepts. Every value domain is in the extension of some conceptual domain.

Value domains and conceptual domains come in two (non-exclusive) sub-types:

- Enumerated – A domain specified as a list of its elements

- Non-enumerated – A domain specified by a description of its elements

An enumerated value domain contains a list of all its permissible values. An enumerated conceptual domain contains a list of all its value meanings. Non-enumerated value domains and non-enumerated conceptual domains are specified by descriptions. The non-enumerated value domain description describes precisely which permissible values belong and which do not belong to the value domain. The non-enumerated conceptual domain description describes precisely which value meanings belong and which do not belong to 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 are in the extension of one conceptual domain. The following two examples taken from ISO/IEC 11179-1 - *Framework* illustrate several things:

- Examples of non-enumerated value domains and a non-enumerated conceptual domain
- Examples of enumerated value domains and an enumerated conceptual domain
- Use of conceptual domains to manage similarities between value domains

EXAMPLE 1 – Similar non-enumerated value domains

Conceptual domain name: Probabilities

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

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

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

Precision: 2 digits to the right of the decimal point

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

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

Precision: 5 digits to the right of the decimal point

EXAMPLE 2 – Similar enumerated value domains

Conceptual domain name: Countries of the world (ISO 3166-1, see [4])

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

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

Permissible values:

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

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

. . .

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

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

Permissible values:

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

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

. . .

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

A classification as used by statistical offices may be described through the use of enumerated conceptual and value domains. In an example taken from ISO/IEC TR 20943-3 [5], part of the ISIC is described below.

The International Standard Industrial Classification[2] (ISIC, rev. 3), is composed of four related value domains, one for each level. They are related for two reasons: 1) each domain (level) is a categorization of industries; 2) each value meaning in the value domain at one level is a generalization of the value meanings in the level below it. In the ISIC, for instance, the value meaning associated with code 51 generalizes the value meaning associated with code 513 (and codes 511, 512, 514, 515, & 519). The following example shows the conceptual domains and value domains necessary to register ISIC for levels 2 and 3:

EXAMPLE 3 - A classification scheme as related value domains

Conceptual domain (general) name: Industrial Classification Systems

Conceptual domain definition: Nested levels of codes representing categories of industries.

Conceptual domain name: Industrial classification systems, Level 2

Conceptual domain definition: Second level codes representing categories of industries.

Conceptual domain relationship: specialization of {Industrial Classification Systems}

Value domain name: Division, ISIC

Value domain definition: Codes for the division (second) level of ISIC.

Permissible values:

<01, Agriculture, hunting and related service activities>

<02, Forestry, logging and related service activities>

<05, Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing>

…

<99, Extra-territorial organizations and bodies>

---

[2] ISIC is described on the Web at the URL http://esa.un.org/unsd/cr/registry/regcst.asp?Cl=2&Lg=1

Conceptual domain name: Industrial classification systems, Level 3

Conceptual domain definition: Third level codes representing categories of industries.

Conceptual domain relationship: specialization of {Industrial Classification Systems}

Value domain name: Group, ISIC

Value domain definition: Codes for the group (third) level of ISIC.

Permissible values:

<011, Growing of crops; market gardening; horticulture>

…

<020, Forestry, logging and related service activities>

…

<050, Fishing, operation of fish hatcheries and fish farms; service activities incidental to fishing>

…

<990, Extra-territorial organizations and bodies>

In this example, the relationships between the value domains are maintained through the relationship of the value domain to its conceptual domain and the fact that each conceptual domain is related to the same generalized conceptual domain. The relationships between the values of different domains are maintained through the codes (values) themselves. The number of characters in the code determines the level.

3.2 Data Element Concepts

Object classes are the things for which we wish to collect and store data. They are concepts. They are approximated by classes or entities in data models.

An object class may be a general concept. This happens when the set of objects corresponding to the object class has two or more members. As general concepts, they correspond to the notions of a universe, population, or domain. Examples are persons, households, or establishments; and the data associated with these object classes are microdata.

An object class may be an individual concept. This happens when the set of objects corresponding to the object class has one member. As individual concepts, they correspond to aggregates, such as all persons in the US or all establishments in the US. Here the data associated with the object class is macrodata.

Any object has an individual concept associated with it. When one thinks of a particular object, the conception of that object is the individual concept associated with it. Data associated with an object class that is its individual concept is descriptive of that object. This kind of data may be metadata, although this characterization is necessary but not sufficient. Data are only metadata when they are used to describe.

Properties are what humans use to distinguish or describe objects. They are characteristics of the object class and form its intension. They are also concepts. For general object classes, the properties correspond to characteristics of populations. Examples of properties of general object classes are sex, age, income, address, or industry. For individual object classes, the properties correspond to characteristics of aggregates. Examples of these properties are total income, average age, or total wages. Also, there are many properties of individual concepts that are used as metadata, such as the Dublin Core attributes.

In ISO/IEC 11179, properties are concepts that exist independently of object classes. First, characteristics of concepts may be concepts themselves. So, properties are managed as concepts in the standard. Secondly, a

concept may be both an object class and a property, depending on its use, for instance, "industry". Industry may be a property of establishments. On the other hand, it may be a set of things that we collect data about, such as the goods produced and processes used in particular industries.

A data element concept is generated from an object class and property. The property is both a characteristic of the object class and a concept in its own right. The combination of an object class and one of its properties is a data element concept. It is the conceptual part of data, that is, independent of its representation.

Object classes and properties often are specialized to account for data associated with domains (of populations) or narrower properties. The generic relation is used, and value meanings from enumerated conceptual domains (i.e., classifications) provide the additional characteristics to narrow the concept. Here is an example illustrating the idea for an object class and for a property:

EXAMPLE 4 – Specializing Object Class and Property

Object Class:	Persons of the US
Classification:	Sex {Male, Female}
Specialized Object Class:	Male Persons of the US
Property:	Income
Classification:	Income {Wages, Retirement, Dividends, Interest, Other}
Specialized Property:	Income derived by wages only

3.3 Data Elements

A data element is formed by associating a data element concept with a value domain. A data element concept may form many data elements, and a value domain may form many data elements. Reuse and sharing of these constructs makes the ISO/IEC 11179 metadata registry specification powerful.

The term data element is synonymous with the term variable as it is understood by programmers. Thus, the datatype associated with a data element is important. The data element concept and the value domain provide a semantic model for a data element. The datatype provides a computational model.

As they are defined in ISO/IEC 11404 [6], datatypes have three components:

- Value space, which is equivalent to a value domain without the value meanings
- Properties, which are provisions (See [7]) that are assumed to be true about the values in the value space
- Characterizing operations, which are functions on the values of the value space, some of which implement the properties

Surprisingly, datatype is not completely determined by value domains and is a characteristic of data elements instead. Similarly, accuracy, precision, and units of measure are characteristics of data elements. The common link among them is that all refer to the values that a data element takes. This should imply that they are characteristics of a value domain. This is not the case, as the following shows:

- Accuracy refers to how nearly the recorded data match the truth, which has nothing to do with a set of allowed values
- Precision refers to how narrowly the recorded data can be distinguished, which, again, has nothing to do with a set of allowed values

- Units of measure refers to the quantity that recorded data represent, but the same set of values (i.e., the same value domain) can be used to represent many different quantities

- Datatype is the computational model for data, which requires the context (the data element concept) to determine; E.g., countries represented by country codes (See [4]) can refer to political boundaries or the boundaries defining air space.

4.0 Metadata Registries

4.1 Introduction

Metadata is data, so metadata may be stored in a database. A database of metadata that supports the functionality of registration is a metadata registry. A metadata registry contains metadata describing data constructs (i.e., data elements, data element concepts, value domains, conceptual domains, object classes, properties, and value meanings).

When metadata describing a given data construct is captured and stored in a metadata registry, then that data construct is said to be registered. Of course, the data construct is not stored in the registry, 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.

4.2 ISO/IEC 11179 Metadata Registry Model

The ISO/IEC 11179 metadata registry model contains two main parts: the conceptual level and the syntactical level. The conceptual level contains the data element concept and conceptual domain. Both are concepts. The syntactical level contains the data element and value domain. Both are containers for data values.

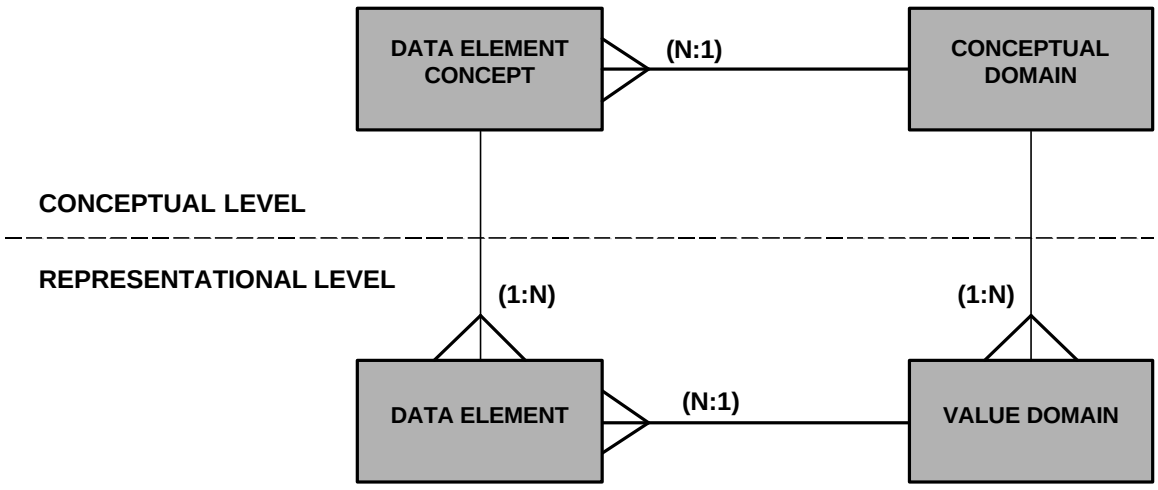

Figure 1: Overview Model for ISO/IEC 11179 Metadata Registry

Figure 1 pictorially represents several fundamental facts about the four classes:

- A data element is the association of a data element concept and a value domain

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

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

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

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

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

- Many value domains can share the same conceptual domain

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

In addition to the facts illustrated in the Figure 1, there are two other facts that ISO/IEC 11179 provides:

- Relationships among data element concepts may be maintained in a metadata registry, which implies that a concept system of data element concepts may be maintained

- Relationships among conceptual domains may be maintained in a metadata registry, which implies that a concept system of conceptual domains may be maintained

4.3 Registration

Registration functions are what separate an ISO/IEC 11179 metadata registry from a database of metadata. Registration is the set of rules, operations, and procedures that apply to a metadata registry. The three most important outcomes of registration are the ability to monitor the quality of metadata, provenance (the source of the metadata), and assigning an identifier to each object described.

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 the standard.

Each description of a data construct is maintained in a uniform and prescribed manner. Identifiers, quality measures, responsible organizations, names, and definitions are recorded for every data construct. Registration is the process of creating or maintaining common and other detailed metadata.

There are several purposes to monitoring metadata quality. The main purposes are

- Monitoring adherence to rules for providing metadata

- Monitoring adherence to rules for forming definitions and following naming conventions

- Determining whether a description still has relevance

- Determining the similarity of related data constructs and harmonizing their differences

- Determining whether it is possible to ever get higher quality metadata for some data constructs

Every data construct registered in an ISO/IEC 11179 metadata registry is assigned a unique identifier. The identifier is used as a means to keep track of descriptions for administration purposes, to refer to descriptions by remote users of the registry, and aid in metadata transfer between similar registries.

The registration authority is the organization responsible for setting the procedures, administering, and maintaining a registry. The submitting organization is responsible for requesting that a new description 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 – *Registration*.

5.0 Conclusion

This paper was written to summarize the main ideas, requirements, and benefits of an ISO/IEC 11179 metadata registry. The main benefits are as follows:

- Ability to describe data with extensive semantics
- Ability to describe data from any subject matter area
- Manage the descriptions of data constructs in a uniform manner
- Link data constructs with similar semantics
- Semantic and syntactic principles follow statistical theory about data
- Build extensive concept systems from semantic descriptions of data constructs

An increasing number of NSO's are implementing ISO/IEC 11179. When more NSO's follow the same framework for describing their data, the ability to share and understand each other's data increases.

6.0 References

[1] ISO/IEC 11179 – *Metadata registries* (All Parts)

[2] ISO 704 – *Principles and methods of terminology*

[3] ISO 1087-1 – *Terminology* – Part 1: *Vocabulary*

[4] ISO 3166-1:1997 - *Codes for the representation of names of countries and their subdivisions* - Part 1: *Country codes*

[5] ISO/IEC TR 20943-3 *Procedures for achieving metadata registry content consistency* – Part 3: *Value domains*

[6] ISO/IEC 11404 – *Language independent datatypes*

[7] ISO/IEC *Guide 2*