

September 16, 1999

# ISO/IEC JTC 1/SC 32 N 0362

Date: 1999-09-30

REPLACES: 32N0262

<p><b>ISO/IEC JTC 1/SC 32</b></p> <p><b>Data Management and Interchange</b></p> <p><b>Secretariat: United States of America (ANSI)</b></p> <p><b>Administered by Pacific Northwest National Laboratory on behalf of ANSI</b></p>
--

<b>DOCUMENT TYPE</b>	Text for NP ballot (text submitted to JTC 1 for ballot)
<b>TITLE</b>	Text for NP – Technical Report - Procedures for Achieving Content Consistency in ISO/IEC 11179 Metadata Registries
<b>SOURCE</b>	JTC 1/SC 32/WG 2
<b>PROJECT NUMBER</b>	
<b>STATUS</b>	Text for NP ballot (text submitted to JTC 1 for ballot) as approved in SC 32 resolution 6 at the Matsue meeting
<b>REFERENCES</b>	
<b>ACTION ID.</b>	LB
<b>REQUESTED ACTION</b>	Approval from JTC 1 for NP
<b>DUE DATE</b>	
<b>Number of Pages</b>	105
<b>LANGUAGE USED</b>	English
<b>DISTRIBUTION</b>	P & L Members SC Chair WG Conveners and Secretaries

Douglas Mann, Secretariat, ISO/IEC JTC 1/SC 32

Pacific Northwest National Laboratory \*, 901 D Street, SW., Suite 900, Washington, DC, 20024-2115,  
United States of America

Telephone: +1 703 575 2114; Facsimile: +1 703 681 9180; E-mail: [MannD@battelle.org](mailto:MannD@battelle.org)

\*Pacific Northwest National Laboratory (PNL) administers the ISO/IEC JTC 1/SC 32 Secretariat on behalf of ANSI

September 16, 1999

ISO/IEC NWI xxx

Information Technology

Procedures for Achieving Content Consistency  
In  
ISO/IEC 11179 Metadata Registries

Working Paper  
Draft 3.0

September 1999

**Procedures for Achieving**

## Data Registry Content Consistency

### Contents

---

Foreword

Introduction

1 Scope

2 References

3 Definitions

4 Component framework

### **5.0 REGISTER A DATA ELEMENT**

- 5.1 General Procedures
  - 5.1.1 Understanding the Data Element
  - 5.1.2 Content Research
  - 5.1.3 Definition and Permissible Values
  - 5.1.4 Name and Identifiers
  - 5.1.5 Other Metadata Attributes
  - 5.1.6 Data Element Concept
  - 5.1.7 Classification Attributes
  - 5.1.8 Quality Control
- 5.2 International Standard with Enumerated Domain
  - 5.2.1 Understanding the Data Element
  - 5.2.2 Content Research
  - 5.2.3 Definition and Permissible Values
  - 5.2.4 Identify and Name the Data Element
  - 5.2.5 Other Metadata Attributes
  - 5.2.6 Data Element Concept
  - 5.2.7 Classification
  - 5.2.8 Quality Control
  - 5.2.9 Other Codes and Names from ISO 3166
  - 5.2.10 Summary of Attributes
- 5.3 International Standard with Non-Enumerated Domain
  - 5.3.1 Understanding the Data Element
  - 5.3.2 Content Research

- 5.3.3 Definition and Permissible Values
- 5.3.4 Identifying and Naming the Data Element
- 5.3.5 Other Metadata Attributes
- 5.3.6 Data Element Concept
- 5.3.7 Classification
- 5.3.8 Quality Control
- 5.3.9 Other Data Elements in ISO 6709
- 5.3.10 Summary of Metadata Attributes
- 5.4 Application Data Element
  - 5.4.1 Understanding the Data Element
  - 5.4.2 Content Research
  - 5.4.3 Definition and Permissible Values
  - 5.4.4 Identify and Name the Data Element
  - 5.4.5 Other Metadata Attributes
  - 5.4.6 Data Element Concept
  - 5.4.7 Classification
  - 5.4.8 Quality Control
  - 5.4.9 Related Data Elements
  - 5.4.10 Summary of Metadata Attributes
- 5.5 Register a Group of Data Elements
  - 5.5.1 Information System Entity Group
  - 5.5.2 Composite Data Element
  - 5.5.3 Use Group
- 5.6 Linking of Data Elements
- 5.7 Registration of Associated Sources/Documents

## 6. Complex data

### **Annexes**

#### **A Bibliography**

#### **B Definitions of representation class terms**

#### **C Principles of managing shared data**

#### **D Data registry uses and users**

#### **E Conceptual and logical data models**

**F Table of Data Elements Attributes for Examples**

**G Top Down Approach to Data Element Registration**

**G.1 Biological Organisms**

G.1.1 Data Element Concepts

G.1.2 Data Elements

G.1.3 Permissible Values

**G.2 Biological Organism Types**

G.2.1 Data Element Concepts

G.2.2 Data Elements

G.2.3 Permissible Values

**G.3 Top Down Registration**

**Y Business Rules for Populating a Metadata Registry**

**Y.1 Data Element Definition**

Y.1.1 Mandatory Rules

Y.1.1.1 Uniqueness

Y.1.1.2 Singular

Y.1.1.3 State the Concept; Not Only its Negative

Y.1.1.4 Descriptive Phrase or Sentence

Y.1.1.5 Contain Only Commonly Used Abbreviations

Y.1.1.6 No Embedded Definitions

Y.1.2 Guidelines for Definitions

Y.1.2.1 Essential Meaning of Concept

Y.1.2.2 Precise and Unambiguous

Y.1.2.3 Concise

Y.1.2.4 Stand Alone

Y.1.2.5 No Embedded Information

Y.1.2.6 Avoid Circular Reasoning

Y.1.2.7 Consistency for Related Definitions

Y.1.3 Data Element Definition Syntax

Y.1.4 Terms Commonly Used in Definitions

**Y.2 Representational Attributes**

Y.2.1 Permissible Values

Y.2.2 Value Domain

Y.2.3 Representational Terms

Y.2.4 Example

**Y.3 Identifying and Naming a Data Element**

Y.3.1 Name Context

Y.3.2 Establish a Naming Convention

Y.3.3 Example of a Naming Convention

Y.3.4 Formulating a Data Element Name

**Y.4 Identification**

Y.4.1 Data Element Identifier and Identifier

Y.4.2 Versioning

**Y.5 Conceptual Relationships**

Y.5.1 Data Element Concept

Y.5.2 Conceptual Domain

Y.5.3 Value Meanings

**Y.6 Classification**

**Y.7 Quality Review**

Y.7.1 Registration Status

Y.7.2 Administrative Status

**Y.8 Reference Documents**

## **Foreword**

ISO (the International Organization for Standardization) and the IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental or non-governmental, in liaison with ISO and IEC, also take part in the work.

This document was prepared by ISO/IEC JTC 1/SC 32, Data Management and Interchange.

## Introduction

The exchange of metadata between ISO/IEC 11179 metadata registries depends not only on registry software that conforms to the standard, but also on metadata contents that are compatible between registries. While the standard has provisions for data element specification and registration, there are pragmatic issues pertaining to populating the registries with content. Based on the experiences of organizations that are implementing the standard, a technical report to explore content issues will help current and future users.

Well-formed data elements and their domains can be recorded in a metadata registry as "models" for potential reuse. Additional attributes may be required to record essential facts about how a data element is used in an application, e.g., for data quality, collection method, collection purpose, etc.

The proposed revision of ISO/IEC 11179, Part 3, models a data element (DE) and its associated components. A data element consists of the data element concept plus its representation. Some questions raised in the process of implementing registries concern this structure. Creation of an application data element frequently requires additional qualification of the object class and/or property. Does this creation of an application element always cause the creation of an application data element concept? Does the qualified concept inherit meaning from the standard concept to which it is related, and is there an adequate place in the current scheme to store this relationship? How are application DEC's distinguished from other DEC's or is there a need to make such a distinction? These are examples of topics that might be explored in a document addressing content consistency among registry implementations.

Conceptualization and articulation of rules and relationships in the creation of object classes, properties, data element concepts and data elements are needed. Explication of the various possible levels of data elements and data element concepts and their relationships would greatly assist in the creation of shareable, well-formed data. Relationship and inheritance from the most abstract data element to the most concrete application data element needs to be specified. Reuse of data value domains should be enabled and regularized.

## 1 Scope

### 1.1 Background

A registry is a tool for the management of shareable data; a comprehensive, authoritative source of reference information about data. It does not contain data itself, but it provides information on the definition, origin, source, and location of data. It supports the standard-setting process by recording and disseminating data standards, which facilitates data sharing among organizations and users. It provides links to documents that refer to data elements and to information systems where data elements are used. When used in conjunction with an information database, the registry enables users to better understand the information obtained.

This Technical Report is based on the American National Standard Institute (ANSI) X3.285:1999 Standard, *Metamodel for the Management of Shareable Data*. The standard specifies the structure of a data registry in the form of a conceptual model. The conceptual model is more abstract than a logical data model in that it does not consider how the data is represented in any particular way. It is not intended to be a logical data model for a computer system, much less a physical model.

A data registry contains the metadata that is necessary to clearly describe, inventory, analyze, and classify data. It provides an understanding of the meaning, representation, and identification of a unit of data. The ANSI X3.285 standard "outlines the information elements associated with a data element concept that need to be available for determining the meaning of a data element to be shared between systems. The standard is a complement to the six-part International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 11179 standard that describes the organization of a data registry for managing the semantics of data elements in data systems."<sup>1</sup>

### 1.2 Purpose

The purpose of this Technical Report is to describe business rules for the registration of data elements and their attributes in a registry. This document is not a user's guide for data entry, but a guide for conceptualizing a data element and its components for the purpose of consistently establishing good quality data elements.

---

<sup>1</sup>American National Standard for Information Technology, *Metamodel for the Management of Shareable Data*, February 20, 1999, ANSI X3.285:1999, proposed as ISO/IEC 11179, Part 3 replacement.

### **1.3 Scope**

The scope of this document is limited to the essential components of a data element: the data element identifier, registry name, definition, and example; data concept; conceptual domain with its value meanings; and value domain with its permissible values. This document is not concerned with the entry of detailed metadata for documents, standards, systems, groups, partners, and message sets.

## 2 References

ISO/IEC DIS 11179-1, *Information technology - Specification and standardization of data elements - Part 1: Framework for the specification and standardization of data elements*

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

ISO/IEC 11179-3:1994, *Information technology - Specification and standardization of data elements - Part 3: Basic attributes of data elements*

ISO/IEC 11179-4:1995, *Information technology - Specification and standardization of data elements - Part 4: Rules and guidelines for the formulation of data definitions*

ISO/IEC 11179-5:1995, *Information technology - Specification and standardization of data elements - Part 5: Naming and identification principles for data elements*

ISO/IEC DIS 11179-6, *Information technology - Specification and standardization of data elements - Part 6: Registration of data elements*

ISO/IEC TR 15452, *Information Technology - Specification of Data Value Domains*

### 3 Definitions

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

**3.1 attribute:** A characteristic of an object or entity.

**3.2 conceptual domain:** A set of possible valid value meanings of a data element expressed without representation.

**3.3 context:** A designation or description of the application environment or discipline in which a name is applied or from which it originates.

**3.4 data element:** A unit of data for which the identification, meaning, representation and permissible values are specified by means of a set of attributes.

**3.5 data element concept (DEC):** A concept that can be represented in the form of a data element, described independently of any particular representation.

**3.6 data element registry:** An information resource that describes the meaning and representational form of data elements.

**3.7 data element representation:** A data element component consisting of a value domain and representation class.

**3.8 data identifier:** A language independent unique identifier of a data element within a registration authority. An unambiguous name for an object within a given context.

**3.9 data item:** An occurrence of a data element value.

**3.10 data value:** An element of a value domain.

**3.11 data value domain:** A set of possible valid values of a data element expressed in a certain representation, for a data element having a value domain.

**3.12 enumerated domain:** A value domain that is specified by a list of all permissible values.

**3.13 identifier:** See data identifier.

**3.14 international registration data identifier (IRDI):** The unique and registered identifier of a data element.

**3.15 metadata:** Data that defines and describes other data.

**3.16 name:** The primary means of identification of objects and concepts for humans.

**3.17 object class:** 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.

**3.18 permissible value (label):** An expression of a value meaning in a specific value domain.

**3.19 property:** A peculiarity common to all members of an object class.

**3.20 representation class:** A classification of types of representations.

**3.21 structure set:** A method of placing objects in context, revealing relationships to other objects. Examples include Entity-Relationship Models, taxonomies, and ontologies.

**3.22 value meaning:** A valid value in a conceptual domain.

**3.23 value meaning identifier (VMID):** A label that uniquely identifies a value meaning.

## 4 Component framework

This clause presents a conceptual framework for structuring data elements and data element components in a registry. Data elements are ideally the result of a process of development, involving several types of abstraction, producing a series of "layers" related to each other by the method of abstraction used to produce one from the other. Layers usually progress from the most general (conceptual) to the most specific (ultimately, the physical layer, although a metadata registry would not contain these).

One could use layers to structure development of a system using the Zachman Framework, for instance, with the highest levels of definition contained in the business view, and development progressing to the implemented system level. The number and granularity of layers are driven by user requirements. This clause will describe several (non-exhaustive) possible layers, none of which are intended to be mandatory for any particular implementation.

The members of each layer are called data element components. Components are envisioned as a set of building blocks that can be assembled into data elements. Some components may also be members of a registry in their own right.

### 4.1 Abstraction types

Abstraction is a tool which has been well-developed by the object-oriented community. It is used as a way of focussing on parts of the model of interest to a particular process or function. The term "abstraction" is used to refer both to the process and the results of the process. Abstraction can be applied to the registry environment as a way to articulate the development of components and their relationships to each other.

Several methods can be used to achieve the decomposition of layers from the most abstract to the more concrete. Starting with the most general conceptual notions and progressing to the data elements in applications, these layers can be labeled by the type or types of abstraction used to produce them from another level.

The three types of abstraction of most interest to data element development are: *decomposition/aggregation*, *instantiation/classification*, and *specialization/generalization*.

- **Decomposition/aggregation** relates an item to its parts. Decomposition may be described as "x is a part of y," or the *part-of* relationship. The reverse, aggregation, shows that y may be composed of x among other items.
- **Instantiation/classification** relates an item to a class of items. This is described as the *is-a* relationship, "x is a(n instance of) y." Classification reverses the relationship; y contains x as well as other items.

- The third type is **specialization/generalization**. This is a relationship between two classes, where all items in one (subclass) are also in the other (superclass).

## 4.2 Conformance

Layers of abstraction can be used to determine conformance of a registry implementation to a standard. Specification of the member classes and abstraction types used to determine the layer members can be used to define conformance. This will lead to improved chances for interoperability among registries.

## 4.3 Developing Layers of Abstraction

The process of deriving layers of abstraction for a registry can be described by a series of examples. Some or all of these layers may be useful for any given registry.

Abstraction relationship types define the boundaries between layers. Rules for conformance may be derived from both boundary abstraction and the relationships of the components of each layer.

A useful starting point is the set of real world things that the registry attempts to model. These can be described by the phrases "concepts (things, beings, ideas...)," "things about them," "how they look," and "what they mean." So, the first layer of abstraction is the translation of these phrases to model entities (figure 1). Applying the abstraction process of specialization, the result is that concepts become *object classes*, things about concepts become *properties*, how they look becomes *representation*, and what they mean becomes the *conceptual domain*. By this transformation, the amorphous content of superclasses of things in the real world become subclasses composed of entities of the model, subject to rules governing their behavior.

Of course, every model-based registry must include this layer. This is the basic assumption of model building.

Within the model, other layers of abstraction can be applied to produce model entities of use to the developers and users. For example, aggregation can be applied to the object class and property entities to produce the *data element concept*. These can be related to conceptual domains (which contain sets of value meanings) to produce a potentially useful entity, the *conceptual generic element* (figure 2).

Conceptual generic elements consist of the attributes associated with their constituent components. These serve to describe the object class property and its value meanings without any particular representation assigned. An example, using ISO 3166, would be to describe **country identifier** without specifying which one of the seven possible representations for names or codes for countries contained in ISO 3166 is preferred.

Consider representation. It was mentioned earlier as if it was a model entity, but it does not exist as such in the model. Representation is a combination of data value domain with its permissible values (if enumerated) or description (if not); representation class; and datatype, character set, and unit of quantity of the values in the value domain. Therefore, it must be abstracted by aggregation if it is to be considered as a unit.

Combining a property with the representation components can create a useful construct. A *logical generic element* such as "height measure in feet" can be used to record conformance criteria such as allowed range values. A narrower construct, limiting the components to property and representation class, can be created to record generalized conformance criteria such as that "height measure" must only be used with units of measure with values of "feet," "inches," "meters," "centimeters," etc. These would potentially be combined with object classes to produce data elements such as "tree height measure" with a conformance criterion of "height >0<500" (figure 3).

Another useful object-oriented concept can be applied to allow inheritance of attribute values between layers. This mechanism enables the process described in the last paragraph to be applied in many-to-one relationships: "height measure" can be applied to "telephone pole height measure" using the same conformance criterion as "tree height measure."

Other combinations of components can be created as the registry designer's discretion. Documentation by attributes and relationships must be complete if registry content consistency is to be maintained. Full use of generics promotes reuse of standardized data description.

Figure 1. Abstraction from the real world to the model.

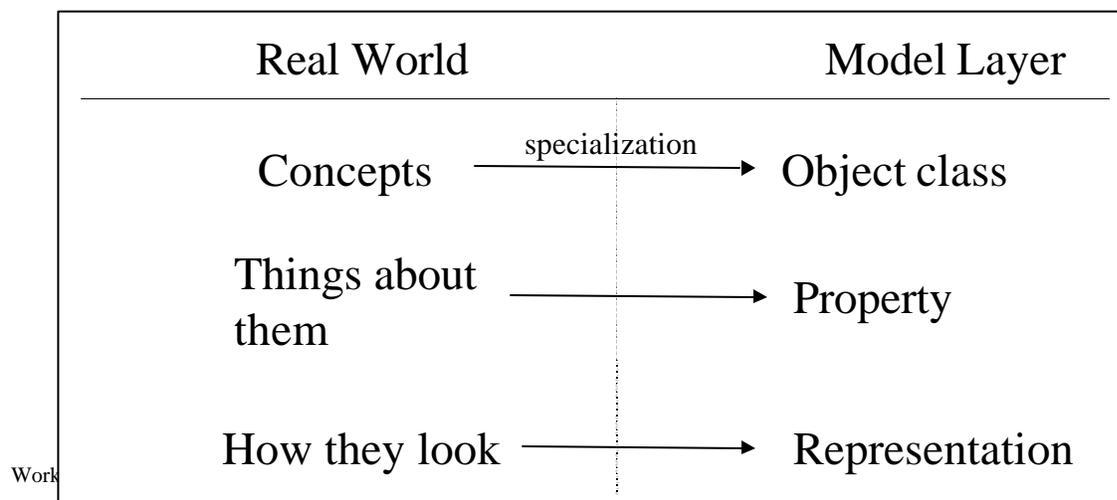


Figure 2. Abstraction of a conceptual generic element.

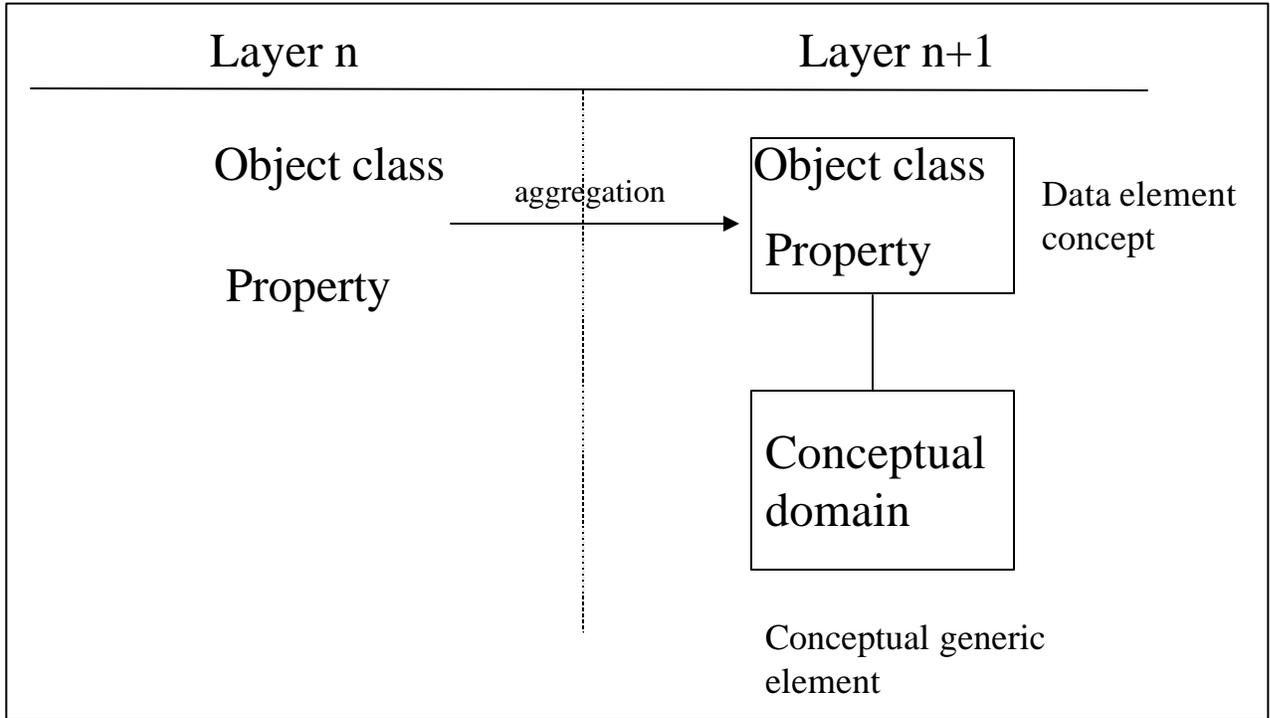
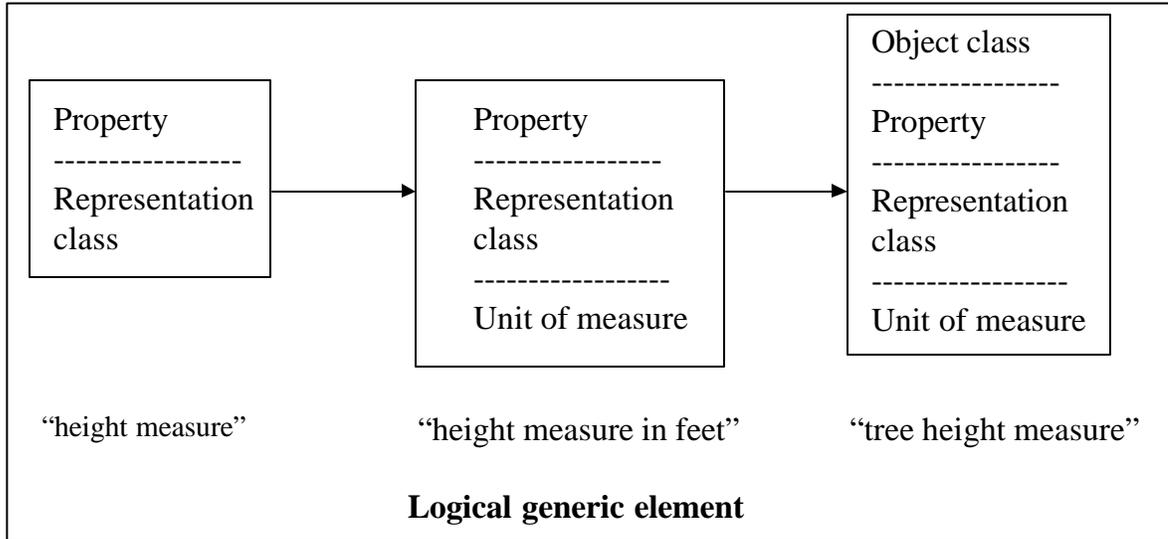


Figure 3. Inheritance of component values.



## REFERENCES (INFORMATIVE)

- X *Codes for the representation of names of countries and their subdivisions Part 1: Country codes*, International Organization for Standards (ISO), ISO 3166:1997.
- X *Standard representation of latitude, longitude and altitude for geographic point locations*, International Organization for Standards, ISO 6709, 1983-05-15.
- X *Information technology programming languages, their environments and system software interfaces language-independent datatypes*, International Organization for Standards, ISO/IEC 11404, First edition 1996 12-15.
- X *Information technology specification of data value domains*, ISO/IEC TR 15452, March 1999.

## 5.0 REGISTER A DATA ELEMENT

Registration of a data element in a data element registry requires that certain characteristics of the data element are recorded to clearly describe and define it. These characteristics are stored as attributes of the data element. A Registry can be used to record information about data elements ranging from carefully crafted data standards to those found in applications. The amount and quality of metadata information available can vary from good, complete information to poor, incomplete information. This document is intended to describe the population of a Registry with data elements for which good quality, consistent metadata can be created. Part 3 of the ISO/IEC 11179 specifies attributes for recording information about a data element in a data registry. This document gives examples that demonstrate the population of a data registry. It includes attributes that are mandatory and fully defined by the metamodel, as well as those where the registration authority must establish its own profile of required attributes.

Many metadata registry practitioners find that using a bottom-up approach to registering a data element is most appropriate. In many cases where a data element is submitted for registration, only limited information (e.g., a name, definition, and a set of permissible values) is provided. Other attributes must be determined based on an understanding of the underlying data values and concepts that are implied by those facts. These are most commonly registered by means of a bottom up registration procedure, where the basic metadata attributes about the data element (e.g., definition, name, and permissible values) are completed prior to defining the conceptual information about the data element. A bottom-up approach might also be used where the metadata registry is intended to serve as a distribution mechanism for metadata that describes the data in data products such as public data sets, query results, etc. The examples provided in this report describe how to formulate attributes about a data element, based on a bottom-up procedure. First a general procedure for registering data elements is described, followed by examples of registration of three types of data elements, including data elements from:

- X An international standard with an enumerated domain.
- X An international standard with a non-enumerated domain.
- X An information system, where the application data element uses an enumerated domain.

The registration procedures are presented in a logical order for analyzing and formulating attributes for a data element. Annex F contains a table that concisely summarizes the information registered for each data element in the examples that follow.

This report is intended to be used to help metadata registry practitioners to formulate the attributes that describe and define a data element. Section 5.1 presents an overall approach to data element registration. Sections 5.2, 5.3, and 5.4 should be consulted for more specific examples of registering the kinds of data elements described in international standards and in information systems. Annex Y, which is based on ISO/IEC 11179, contains more detailed information and examples to assist the practitioner who is registering data elements.

A top-down approach is useful in many circumstances. Although it requires more "up front" effort, top-down registration has the potential to produce more stable and uniform metadata. An example of a top-down registration, where registration begins with identification of conceptual domains, is provided in informative Annex G with an example of registration of data elements about biological organisms.

## **5.1 General Procedures**

Often only a limited amount of information is available about a data element that has been submitted for registration, e.g., the name and definition contained in a document or provided by the submitting organization and a set of permissible values, where appropriate. The general procedures that follow are intended to result in the registration of a complete, well-defined data element that meets the requirements of a particular registration authority.

It should be noted that the metadata for some data elements in a registry will never be complete. This is true of application data elements that are obtained from computer software, where very little information is known except the representational attributes (e.g., field length and datatype). For these data elements, only the most basic attributes will be entered, and the data element's registration status will remain incomplete.

### **5.1.1 Understanding the Data Element**

The first step in the registration procedure is to gain an understanding of the data element. What kind of data will be stored in this data element? Is there a definition or description of the data values? Were permissible values or examples of the data provided? Will the data values be determined by an arithmetic or statistical procedure? What will the data values look like; e.g., are they names or descriptions of things, numerals to be calculated, strings of characters and numbers that are identifiers? Where documentation is inadequate to fully understand the data element, the practitioner must consult those who represent the source of the data element to obtain the necessary information.

The result of this first step is an understanding of the semantic content of the data element.

### **5.1.2 Content Research**

Prior to formulating attributes towards registration of a new data element, the registrar should perform content research to determine whether a data element is described in an existing International or National standard, or whether a data element that has the potential for being reused exists in the registry or a federation of registries. It is necessary to recognize that the registration practitioner must make value decisions when recording metadata into the metadata registry. The practitioner will determine if a data element might be adapted to meet new requirements, or some attributes of an existing data element (e.g., value domain, data element

concept, or conceptual domain) might be reused with the new data element. Content research should include a search of conceptual domains, data element concepts, and value domains as well as data elements, to identify attributes that might be relevant to the data element to be registered. If a standard data element exists that can be used as a model to meet the particular specifications for a new purpose, some of its attributes may be reused for registration of the new data element.

The result of this step is confirmation that a new data element is needed, or a decision to modify or reuse an existing data element.

### **5.1.3 Definition and Permissible Values**

The essential semantic content of a data element must be captured in a data element definition. Part 4 of ISO/IEC 11179 describes rules and guidelines for formulating definitions. Part 3 identifies the attributes for describing the domain of potentially valid (i.e., permissible) values. The permissible values for a data element are defined as a value domain. Examples are provided in Annex Y for formulating definitions, based on the rules and guidelines set forth in ISO/IEC 11179-4. Annex Y also contains detailed information about the attributes in value domains and examples of how those attributes are used for both enumerated (i.e., established through a list) and non-enumerated domains (i.e., specified through a formula, rule, procedure, or reference).

Different attributes are used depending upon whether the potentially valid values are enumerated or non-enumerated. Each permissible value is associated with a valid value meaning that provides meaning to the permissible value, as described in Section 5.1.6. Each permissible value is also entered in the registry with its begin date (i.e., the date when that permissible value became valid for that value meaning). End dates will also be entered, when the permissible value for a value meaning becomes invalid.

Value domains for non-enumerated domains must include a definition/description of the values that are possible valid values for the data element. This report contains specific examples of registering data elements with enumerated domains (Sections 5.2 and 5.4) and with non-enumerated domains (Section 5.3).

### **5.1.4 Name and Identifiers**

Part 5 of ISO/IEC 11179 gives principles for naming and identification of data elements. Each data element registered within a Registration Authority (RA), i.e., an organization authorized to register metadata, is unambiguously identified with a unique identifier. Although the standard does not specify the format or content of the data element identifier (DI), the DI should carry no useful information about the data element, e.g., it might be a number assigned sequentially by an automated system. If the attributes of a data element change, a new version of the data element is created and registered with a version identifier (VI).

Since each RA establishes its own identification scheme, the same DI might be used to identify a different data element in another metadata registry. Therefore, a Registration Authority Identifier (RAI) must be established for unique identification of a data element. Data elements registered under the provisions of ISO/IEC 11179 are assigned an international registration data identifier (IRDI), which is a composite of the RAI, the DI, and the VI. Part 6 of ISO/IEC 11179 describes the requirements for a RA and the construction of a RAI. The IRDI is discussed further in Part 6.

Most people prefer to use names when talking about a data element, rather than a non-intelligible identifier. Therefore, one or more names can be assigned to a data element, each associated with the context in which the name is used. A name can be developed for a scientific discipline, an organization, a particular computer language, a database management system, or other purpose. Each name is developed according to the naming convention for the particular name context. The naming convention can vary from "whatever you want to call it" to a highly structured name. ISO/IEC 11179, part 5 does not specify a mandatory naming convention, but does explain how to document one. For this report, the data element names are based on a naming convention described in Annex Y. Annex Y also expands on Part 5 of the standard by providing examples of the use of names and name contexts.

### **5.1.5 Other Metadata Attributes**

Other mandatory and optional data element attributes are described in Part 3 of ISO/IEC 11179. In addition to the definitional attributes described in Section 5.1.3 and the identifying attributes described in Section 5.1.4, there are administrative, relational, classifying, and other miscellaneous attributes that serve to define and describe a data element.

In addition to the mandatory attributes specified by Part 3 of the standard, a RA might establish a profile for a particular metadata registry, where some of the attributes described as optional in the standard are mandatory for that registry, some optional attributes are not included, and additional attributes might be identified to extend the registry.

The attributes that relate data elements through data element concepts (Section 5.1.5), and those that classify data elements (Section 5.1.6) are described in subsequent sections of this report. Many information sources do not provide information about the data element for these categories. Some administrative information is related to quality control, and is described in Section 5.1.8. Annex Y includes detailed information about these metadata attributes.

For the registration procedure described in this report, some administrative and miscellaneous attributes are recorded at this time, including:

- X      Submitting organization: The submitting organization is the Office or organization that has submitted the data element for registration.

- X Data Steward: The data steward is the individual who has been assigned by a submitting organization to be responsible for authorizing and maintaining one or more data elements.
- X Note: A data element may have a "Note" or "Comment" that can be used to capture additional descriptive information about a data element, including usage, procedure, and other explanatory information that is not appropriate to include in the data element definition attribute.
- X Example: A data element shall be registered with an example, which must be one of the permissible values for enumerated value domains or must conform to the value domain description/definition and other value attributes for non-enumerated domains.
- X Origin. A data element can be associated with any kind of source, including a document, standard, system, group, partner, or message set. One source, as a minimum, must be associated with a data element to indicate the origin of information about the data element.

### **5.1.6 Data Element Concept**

At this stage in registering a data element, it is possible to specify conceptual information about the data element through the data element concept. The data element concept can be thought of as an idea or perception about something, identified and described independently of any representation. The data element concept may relate several data elements that record data about that concept with different representations, e.g., names and codes that represent provinces of Canada and share the same concept, which is "Canadian Province Identifier."

The data element concept is singular (only one concept is represented). It can be associated with many data elements, including other names and codes, and it does not include a representation class term in its name or definition. The data element concept is associated with only one Conceptual Domain, as described in the following paragraph.

Data element concepts are specified through a definition, an identifier, a name, and a conceptual domain, i.e., the meanings of the possible set of valid values for a data element, expressed without representation. The conceptual domain, "Canadian Provinces", would include valid value meanings such as "The Canadian province of (Alberta,....., Yukon Territory)," where each value meaning would identify one Canadian province. Each value meaning is entered in the registry, associated with its conceptual domain, with its begin date (i.e., the date when that value meaning became valid) and end (i.e., when the value meaning became invalid). Permissible values are associated with value meanings, according to the representation defined by the value domain.

Derivation of data element concepts and conceptual domains, including value meanings are described in detail in Annex Y.6.

### 5.1.7 Classification Attributes

The classification attributes are recorded, where appropriate, at this time. Classification helps to add information not easily included in definitions, helps to organize the contents of a metadata registry, and helps to provide access by supporting more meaningful queries. Part 2 of ISO/IEC 11179 describes general categories of classification; Part 5 describes three classified components: object class, property, and representation class.

A metadata registry might choose to classify data elements as groups, e.g., the group of data elements used in a mailing address, the group of data elements used to identify chemical substances, or the group of data elements that locate a point on the surface of the earth. Keywords might also be used to classify data elements, e.g., altitude, date, facility, industrial, and organization.

### 5.1.8 Quality Control

Initially, only some of the attributes will be recorded for a newly registered data element. Such a data element will be assigned the registration status of "incomplete." When all of the mandatory data elements have been completed, but the quality of the metadata has not been verified, the registration status will be "recorded." Through the quality review process, some data elements will be determined to be "certified," and some might become "standard." The "standard" data element is the preferred data element to be used for data sharing, to ensure consistent representation and understanding of the data being communicated.

Part 6 of ISO/IEC 11179 describes the registration process and the registration status assigned to a data element as the metadata are reviewed and quality is improved. Many data elements might be entered into a data registry, but only a relatively small number of them might be assigned a "standard" registration status. Annex Y describes the assignment of Registration and Administrative Status throughout the life cycle of a registered data element. ISO/IEC 11179 Part 6 specifies the levels of registration status; the administrative statuses, however, are established for each registry by the RA.

## 5.2 International Standard with Enumerated Domain

This section provides a specific example of the registration of a data element from an international standard, where the possible valid values are itemized. The International Organization for Standardization (ISO) 3166-1:1997(E/F), *Codes for the representation of names of countries and their subdivisions B Part 1: Country codes*, is used as the source for this example. ISO 3166:1997 is a complete revision of ISO 3166, which was first published in 1974. The names of countries in the standard correspond to those given, in English and French, in the

current "Terminology Bulletin B Country Names," issued by the United Nations Department of Conference Services, entitled "States Members of the United Nations, Members of the Specialized Agencies or Parties for the Statute of the International Court of Justice" and to those published in the "Standard Country or Area Codes for Statistical Use" established by the United Nations Statistics Division. The full name is the formal title as notified by the country concerned to the UN Secretary General.

(ISO) 3166-1:1997(E/F) cancels and replaces the fourth edition (ISO 3166:1993) and comprises a consolidation of all changes to the lists of the fourth edition agreed to by the ISO 3166 Maintenance Agency: ISO 3166 Maintenance Agency Secretariat, c/o DIN Deutsches Institut für Normung e.V., Burggrafenstrasse 6, D-10787 Berlin, Germany.

ISO 3166 includes the following domains: short country name in English, full (official) country name in English (not provided for all countries), 2-character alphabetic code, 3-character alphabetic code, 3-character numeric code, short country name in French, and full country name in French.

The following paragraphs are presented in the logical order for formulating attributes for a standard, enumerated data element, using the short English-language country name as the example. The table in Section 5.2.10 contains all of the metadata attributes recorded for the enumerated data element from an international standard.

### **5.2.1 Understanding the Data Element**

The data element to be registered is taken from an international standard, and it includes an authoritative conceptual domain of country identifiers for all of the countries of the world. The short English-language name was selected for standardization because it has the most utility for information systems used by United States (U.S.) federal agencies as well as the private sector. The short form of the English-language name is used by the U.S. Postal Service (USPS) for all outgoing international mail, in preference to any of the codes or full names that are included in the standard. The name is also preferred by the USPS to any names that are used locally by a country to identify itself, e.g., Japan is recognized by the USPS in preference to Nihon, which is the country name commonly used by that country itself. The short form of the name in English has been used in the development of ISO 3166 as the basis for assigning codes to avoid, wherever possible, any reflection of a country's political status.

The English-language short name in the standard varies in length from four alphabetic characters (e.g., Peru) to 44 alphabetic characters (i.e., South Georgia and the South Sandwich Islands). The names use the English language alphabet for their character set.

### **5.2.2 Content Research**

Other standards that contain conceptual domains for country identification include U.S. Federal Information Processing Standards (FIPS), published by the U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology (NIST). FIPS 10-4 is maintained by the Office of the Geographer and Global Issues, U.S. Department of State. It is intended for use in activities by the Department of State and national defense programs, and can also be used for Federal interchanges of information with the non-Federal sector of the U.S. FIPS 10-4, published in April 1995, reflects changes through May 6, 1993. FIPS 104-1 implements an American National Standards Institute standard ANSI Z39.27-1984, and adopts, with qualifications, entities, names, and codes prescribed by ISO 3166. FIPS 104-1 was last updated on May 12, 1986. The maintenance organization is the National Bureau of Standards (now NIST) in coordination with the U.S. Department of State, the U.S. Board of Geographic Names, and the maintenance organization for ISO 3166. There are no known plans to update either of the FIPS standards, and neither of these standards is recognized internationally.

An authoritative international source of value domains which has ongoing maintenance is a necessity for maintaining data values for the data elements identifying countries of the world. Therefore, the ISO 3166:1997 is used as the origin of the data element for country name.

### **5.2.3 Definition and Permissible Values**

The definition and permissible values are the most important metadata attributes in uniquely describing a data element.

#### **5.2.3.1 Definition**

Understanding that the essential meaning of this data element is to identify countries using a short name in the English-language, the data element definition can be formulated as "The short name of a country, represented in the English language." This definition is formulated using the mandatory rules and guidelines established in ISO/IEC 11179-4. The rules and guidelines from Part 4 are described with examples in Annex Y.2. The definition is singular, since any instance of the data element contains only one value.

#### **5.2.3.2 Permissible Values**

The permissible values for the data element are the short names in English, listed in ISO 3166 (e.g., Afghanistan, Albania, ....., Zimbabwe). Each permissible value is entered into the registry with the date when that permissible value was valid for that value domain (in this case the date is January 10, 1997, the same as the begin date for the value meaning). There is no end date to enter at this time.

The scope of the permissible values for this data element includes the short English-language name for all countries. A value domain is defined as the permissible values for a data element.

For this example, the value domain is described as "All short, English-language names of all countries." Note that Part 3 of ISO/IEC 11179 does not require a description or definition for enumerated domains. Some RA, however, prefer that all value domains be registered with a description/definition. Record the other value domain attributes for this example at this time, including:

- X Character Set: The character set for Short English-Language Country Name is "English language."
- X Domain Type: Country names are a fixed list of countries, maintained by international standards; therefore, the domain type is "enumerated."
- X Datatype: The datatype for country name is "alphanumeric."
- X Maximum and minimum field lengths: Based on prior research (Section 5.2.1), the minimum length for values for the data element is known to be four. The known maximum length for names in the current standard is 44. The maximum field length, however, is set to 60, to accommodate any changes or additions to the domain of values.
- X Format: The format selected by the registration authority for this example is A(60) to accommodate the longest of the English-language short names.

#### **5.2.4 Identify and Name the Data Element**

Name do not identify a data element. Identification requires a unique identifier, preferably one that does not contain information about the data element. The name provides a designator so that users of the registry have terms by which they refer to the data element.

##### **5.2.4.1 Identification**

Assign a unique identifier to the data element for short English-language country name, as described in Annex Y for the identification of data elements. In the metadata registry for this example, a unique DI and VI (20903:1) are assigned by the computer at the time of registry.

##### **5.2.4.2 Name Context and Naming Convention**

ISO/IEC 11179 Part 5 describes the naming of data elements. Annex Y gives examples of name contexts and naming conventions. For this international standard data element, the name is assigned the context of "Registry," and it is derived based on the example naming convention provided in Annex Y and summarized as follows:

Scope: The scope of this example naming convention is Registry Name.

- § Authority: The authority for this example is the U.S. Environmental Protection Agency for its Environmental Data Registry.
- § Semantic Rules: Names shall include an object and a property, where appropriate. Qualifiers shall be used to differentiate between names that would otherwise be the same. The representation class term shall always be included as the last term in the name.
- § Lexical Rules: A data element name shall have a maximum of 100 alphanumeric characters. The language of the registry shall be English, and the character set ASCII. There are no controlled word lists.
- § Name Uniqueness: Names shall be unique within a registration authority.

### 5.2.4.3 Name the Data Element

Using the above naming convention, the name is entered with the context of "Registry." The convention specifies that the name should include the object "Country", to indicate the data values to be stored in the data element. The name should also include the representation for the concept, in this example "Name." For this particular example, it is necessary to qualify the name, since there are four value domains of country names in the ISO 3166 standard. The qualifiers: "short" and "English-language" are appropriate to this example. The name that has been formulated for this data element, therefore, is "Short English-Language Country Name."

### 5.2.5 Other Metadata Attributes

Other metadata attributes that can be recorded at this time are:

- X Select the example for this data element; it must be one of the permissible values in the value domain.  
  
Example: China
- X Identify the origin for this data element as the standard from which the permissible values are obtained.  
  
Origin: ISO 3166-1:1997, *Codes for the representation of names of countries and their subdivisions - Part 1: Country codes* (Document)
- X Record any notes or comments that might provide additional information about the data element that is not included in the definition.  
  
Note: This data element is included in the EPA revised interim Facility Identification Standard.

- X Enter the name of the submitting organization, which is the Office that submitted the data element for registration.

Submitting Organization: Office of Information Resources Management

- X Record the name of the individual or organization assigned the responsibility for monitoring and maintaining the data element as the data steward.

Data Steward: Marian Cody

- X Administrative metadata, such as Create Date and User Name are recorded or captured automatically by the system where applicable.

### **5.2.6 Data Element Concept**

Identification of the data element concept, as described in Section 5.1.6 is based on the data element name and definition, without the representation. The concept represented by the data element "Short English-Language Country Name" is "Country Identifier," defined as "An identifier for a primary geopolitical entity of the world." This concept can be represented by all seven of the names and codes included in ISO 3166.

The conceptual domain is a collection of value meanings that provide meaning to the permissible values for a data element. The conceptual domain that contains value meanings related to the identity of countries of the world is named "Countries of the World." It is defined as "The primary geopolitical entities of the world." The value meanings associated with this conceptual domain are defined as "The primary geopolitical entity of the world known as <country name>," where country name is one of the country names listed in ISO 3166. Each value meaning is identified by its own value meaning identifier (VMID) and each is entered into the registry with the date when that value meaning was entered into the conceptual domain (in this case the date is January 10, 1997). End dates will also be entered, when the value meaning becomes invalid (e.g., when a country name changes or the territory of a country changes to be combined with another country or to be subdivided into two or more other countries).

### **5.2.7 Classification**

This data element might be classified according to the following classification schemes:

- Identify one or more keywords, where the keyword is a name or subject matter descriptor that will facilitate grouping like data elements for retrieval.

Keyword: Country.

- Group Short, English-Language Country Name with similar data elements according to concept for translation or by general subject matter.

Conceptual group: Country Identifiers  
Subject group: Geopolitical Entities.

- Identify the class by which this data element is represented.

Representation Class: Name

§ One or more real world objects that identify this data element can be identified at this time.

Object: Country

### **5.2.8 Quality Control**

When all of the mandatory metadata attributes have been entered for this data element, it is assigned the Registration Status of "Recorded" and the administrative status of "In Quality Review." Because the data element was identified by an international standard, and it is expected to be the preferred data element for representing country name within the example metadata registry, the registration status will be updated to AStandard@ with administrative status AFinal@ -after the necessary quality review has been completed.

### **5.2.9 Other Codes and Names from ISO 3166**

Other codes, official English names, and French names (both official and short) from ISO 3166 are registered with their individual value domains, representation, data element definitions, and data element names. All of the data elements associated with ISO 3166 will share the same data element concept (i.e., Country Identifier, defined as "An identifier for a primary geopolitical entity of the world.") and the same conceptual domain (i.e., Countries of the World, defined as "The primary geopolitical entities of the world."). All of the ISO 3166 data elements will share the same value meanings. They will, however, have different sets of permissible values associated with the value meanings, depending upon the data element, its representation, and its value domain.

### **5.2.10 Summary of Attributes**

The metadata attributes that have been assigned to this data element, the short, English-language country name identified by the ISO 3166:1997 standard, are summarized in the following table, and in the first column of the table in Annex F.

<b>Data Element Meta-- model Attribute Name</b>	<b>Example</b>	<b>ISO 3166 Enumerated, Name</b>
<b>1. Data Element Definition and Permissible Values</b>		
Data Element Definition Context	Registry	
Data Element Definition	The English-language short name of a country.	
Permissible Values	All English-Language Short Country Names from ISO 3166, matched with value meanings. (Afghanistan, Albania,....., Zimbabwe)	
PV Begin Date	19971001	
PV End Date	(Not Applicable)	
Value Domain Definition	All English-language short names of all countries.	
Character Set	English language	
Domain type	Enumerated	
Determinant Type	(Not Applicable)	
Range Limits	(Not Applicable)	
Datatype	Alphanumeric	
Minimum	4	
Maximum	44	
Format	A(60)	
Unit of Measure	(Not Applicable)	
Precision	(Not Applicable)	
<b>2. Data Element Name and Identifier</b>		
Data Element Name Context	Registry	
Data Element Name	Short English-Language Country Name	
DE Identifier/	20903:1	

<b>Data Element Meta-- model Attribute Name</b>	<b>Example</b>	<b>ISO 3166 Enumerated, Name</b>
Version Number (DI:VI)		
<b>3. Other Metadata Attributes</b>		
Example		China
Origin		ISO 3166-1:1997, Codes for the representation of names of countries and their subdivisions B Part 1: Country codes (Document)
Note/Description		This data element is included in the EPA revised interim Facility Identification Standard.
Submitting organization		Office of Information Resources Management
Data Steward		Marion Cody
<b>4. Data Element Concept (DEC)</b>		
Data Element Concept Name		Country Identifier
Data Element Concept Definition		An identifier for a primary geopolitical entity of the world.
Conceptual Domain Name		Countries of the World
Conceptual Domain Definition		The primary geopolitical entities of the world.
Enumerated Value Meaning Text		The primary geopolitical entity known as <China>.
VM Begin Date		19971001
VM End Date		(Not Applicable)
<b>Classification</b>		
Keyword		Country
Group		Country Identifiers, Geopolitical Entities
Representation Class		Name
Object		Country
<b>Quality Control</b>		
Registration Status		Standard

Data Element Meta-- model Attribute Name	ISO 3166 Enumerated, Name
Administrative Status	Final

### 5.3 International Standard with Non-Enumerated Domain

This section provides a specific example of the registration of a data element from an international standard, where the possible valid values are not enumerated, but must be determined by a procedure. The International Organization for Standardization (ISO) 6709-1983 (E), *Standard representation of latitude, longitude and altitude for geographic point locations*, is used as the source for this example. ISO 6709 was developed by ISO Technical Committee ISO/TC 97, Information processing systems, and was circulated to member bodies in November 1981. Eighteen countries approved the standard, no member body expressed disapproval. There is no known schedule for review and update of the standard. ISO/TC 32 has been assigned as the maintenance authority for the standard; ISO/TC 211 has expressed an interest in assuming responsibility for its maintenance.

The table in Section 5.3.10 contains all of the metadata attributes recorded for the non-enumerated data element from an international standard.

#### 5.3.1 Understanding the Data Element

Latitude is a measure of the angular distance on a meridian north or south of the equator. The standard provides for a variable format and more than one representation for recording the latitude measure (i.e., degrees and decimal degrees and sexagesimal [i.e., degrees, minutes, and seconds]. The standard also includes more than one representation and format for longitude, and a flexible format for altitude. In addition, a standard format for data transfer is included in the standard.

Although new technology and new tools (e.g. Global Positioning System [GPS]) and analytical and mapping software have caused some geographic information specialists to prefer the measurement of locational coordinates in degrees and decimal degrees, many organizations continue to measure latitude and longitude in degrees, minutes, and seconds. Therefore, the RA of the metadata registry in this example, has determined a need to register a data element for latitude measured in degrees, minutes, and seconds. According to the standard, the placement of the decimal point indicates the transition from degrees to sexagesimal measures. Examples of data in the standard include sexagesimal latitudes that are measured to a range of one or two decimal places for seconds. The standard, however, does not limit the precision, but requires only that the number of decimal places indicate the precision of the measurement. The RA for

this example requires that latitude be recorded up to 5 decimal positions, where it can be measured to that level of precision.

Latitude values are measured in a range of 0 (on the equator) to 90 degrees. Minutes and seconds each are measured in a range of 0 to 60. Latitude values on or North of the equator are recorded as positive numbers; those South of the equator are negative. Where latitude degrees are measured in single digit, they must be recorded with a preceding zero. For data transfer, latitude measures must be preceded by the directional symbol (+ or -), and they must include decimal point, when the measurement includes decimal seconds. Latitude always precedes longitude, which precedes altitude. The latitude and longitude must be expressed in the same format style and to the same precision (indicated by the number of decimal positions). There are no separators between the latitude, longitude, and altitude; the directional symbol serves as a separator for the data element values.

### 5.3.2 Content Research

Part 11 of ISO 15046, Spatial referencing by coordinates, describes the minimum data required to define 1-, 2-, and 3-dimensional coordinate reference systems. The coordinate reference system must be fully defined for a position to be unambiguous. Knowledge of the reference system is necessary to determine if coordinate points are comparable. The standard does not, however, provide information about representation of the coordinates. ISO/TC 211/ WG 3, the workgroup that is currently revising ISO 15046, has expressed an interest in revising (ISO) 6709-1983 (E), *Standard representation of latitude, longitude and altitude for geographic point locations*. Because of TC211's interest in ISO 6709, and their current work on the closely related standard, ISO 15046, it seems likely that ISO 6709 will soon be reviewed and updated if needed. Therefore, ISO 6709 seems appropriate to be identified as a standard data element for latitude measure where latitude is measured as sexagesimal (i.e., in degrees, minutes, and seconds).

A search of the metadata registry in our example reveals about 40 data elements related to latitude measure. One, an EPA interim standard for latitude, measured in degrees and decimal degrees, is compliant with the ISO 6709 data element for degrees. None of the other data elements has the potential for compliance with ISO 6709 for sexagesimal measure of latitude. The other latitude data elements in the registry have been assigned the registration status of incomplete, and many data elements are qualified (e.g., latitude where a facility is located, latitude of a smoke stack). For the purpose of this example, none have the potential for being modified to meet the requirements of the ISO 6709 standard for latitude, measured in degrees, minutes, and seconds.

Therefore, in this example, the ISO 6709 latitude, sexagesimal measure, is selected for registration as a new data element.

### 5.3.3 Definition and Permissible Values

### 5.3.3.1 Definition

The data element definition is formulated according to the rules and guidelines described in Annex Y, based on ISO/IEC 11179-4. The rules require that a data element definition be unique within the registry, so the unit of measure has been included in the definition as "The sexagesimal measure of the angular distance on a meridian north or south of the equator." Including the unit of measure in the definition distinguishes the data element from the EPA interim standard, defined simply as "The measure of the angular distance on a meridian north or south of the equator." The definition is singular, because it refers to only one instance of the data value. Note that ISO 6709 does not include a definition for latitude.

### 5.3.3.2 Permissible Values

ISO 6709 is an international standard that does not list specific values that are valid for the data element; the measure of latitude is a non-enumerated domain. There are no stored permissible values in a registry for non-enumerated domains. The values that are permissible for the ISO 6709 sexagesimal latitude data element are those values that conform to the definition of the value domain and the attributes for datatype, format, unit of measure, and precision. The value domain for sexagesimal latitude can be described as "All sexagesimal measures of the distance of an angle north or south of the equator." By including the unit of measure in the definition, the value domain is distinguished from the value domain description for latitude measured in degrees. The definition is plural, because it includes all possible measurements of latitude determined by this type of measurement.

Latitude values that are measured as degrees, minutes, and seconds must conform to the format +/-DDMMSS to +/-DDMMSS.SSSSS. The precision of the value is indicated by the number of decimal places recorded.

Other value domain attributes for this example include:

- § Character Set. The character set for latitude measure is "English language."
- § Domain Type. Non-enumerated.
- § Description/definition. All sexagesimal measures of the distance of an angle north or south of the equator.
- § Datatype. The datatype for latitude measure is "alphanumeric" to explicitly include the directional symbol and decimal point, where appropriate.

- § Maximum and minimum field lengths. The known minimum field length at this time is seven (+/- DDMMS) where no decimal seconds are recorded. The maximum field length is 13 (+/- DDMMS.sssss), to accommodate up to five decimal places for seconds.
- § Format. The format selected by the registration authority for this example is A(13) to accommodate the maximum number of decimal positions.
- § Range for degrees is 0-90; for minutes is 0-60; for seconds is 0-60.

### **5.3.4 Identifying and Naming the Data Element**

#### **5.3.4.1 Identifiers**

A unique identifier is required for the latitude data element. For the RA in this example, the DI and VI (312345:1) are assigned automatically by the metadata registry software.

#### **5.3.4.2 Name Context and Naming Convention**

For this ISO standard data element, the name is assigned with the context of Registry, using the naming convention described in the example in Annex Y, summarized as follows:

- § Scope: The scope of this example naming convention is Registry Name.
- § Authority: The authority for this example is the U.S. Environmental Protection Agency for its Environmental Data Registry.
- § Semantic Rules: Names shall include an object and a property, where appropriate. Qualifiers shall be used to differentiate between names that would otherwise be the same. The representation class term shall always be included as the last term in the name.
- § Lexical Rules: A data element name shall have a maximum of 100 alphanumeric characters. The language of the registry shall be English, and the character set ASCII. There are no controlled word lists.
- § Name Uniqueness: Names shall be unique within a registration authority.

### 5.3.4.3 Name the Data Element

Using the above naming convention, the name is entered with the context of "Registry." The convention specifies that the name should include the object "Latitude", to indicate the data values to be stored in the data element. Include the representation for the concept in the name; in this example "Measure." There is no requirement in ISO/IEC 11179 Part 5 that data element names be unique in a registry. However, the naming convention used in this example specifies that names must be unique within a registry. It is advisable to use a qualifier in the data element name to differentiate between data elements that might otherwise have the same name. The name includes the object (latitude) and the representation (measure). For this example, the name of the latitude data element will carry the qualifier "sexagesimal" as a discriminator. The name that has been derived for the latitude data element is "Latitude Sexagesimal Measure."

### 5.3.5 Other Metadata Attributes

Other metadata attributes that can be recorded at this time are:

- § Provide an example of the data value that conforms to the description in the value domain, and to the datatype, format, and other value domain attributes for this data element.

Example: +674532 and +674531.85435

- § Record the origin of this data element as the standard where the data element was identified.

Origin: ISO 6709-1983 (E), *Standard representation of latitude, longitude and altitude for geographic point locations.*

- § Record notes and comments that contain additional information about the data element that is not appropriate for the definition.

Note: Latitude sexagesimal converts to latitude degrees by the following formula:  
seconds x 60 = decimal minutes, total minutes x 60 = decimal degrees.

- § List the Office that submitted the data element for registration as the submitting organization.

Submitting Organization: Office of Information Resources Management

- § The organization or individual that has responsibility for maintaining and updating the data element is recorded as the data steward for that data element.

Data Steward: Larry Fitzwater

- § Administrative metadata, such as Create Date and User Name are recorded or captured automatically by the system where applicable.

### 5.3.6 Data Element Concept

The methodology to be used for deriving a data element concept is described in Section 5.1.6 and Annex Y of this document. A data element concept is the data element without representation. We have indicated previously that latitude is a distance measure, where measure is its representation. The data element concept for latitude measure is "Latitude Distance" with the definition "A measure of the angular distance of a point on the surface of the earth north or south of the equator." Note that this concept definition incorporates the term "measure," which is a representation term. The concept of latitude, however, is the measure of a distance. Therefore, it is appropriate in this instance to use the term measure when defining the concept.

A conceptual domain is a collection of value meanings. The collection must be identified with a name and a definition. The latitude is one of the horizontal coordinates that fix a position on the surface of the earth either north or south of the equator. For this example, the name of the conceptual domain for latitude measure is "Latitude Coordinates" with the definition "The coordinates that indicate the distance north or south of the equator for locations."

For non-enumerated domains, such as latitude measure, the value meanings are not explicitly identified. The conceptual domain for the Latitude Distance data element concept is the perceived repository of all latitudes that mark positions on the earth with relation to the equator. The value meanings could be defined as "The distance measure of a point north or south of the equator that is <value>." No value meanings are stored in the registry.

### 5.3.7 Classification

This data element might be classified according to the following classification schemes:

- § Identify one or more keywords, where the keyword is a name or subject matter descriptor that will facilitate grouping like data elements for retrieval.

Keyword: Latitude, Horizontal Coordinate, Spatial

- § Group Short, English-Language Country Name with similar data elements according to concept for translation or by general subject matter.

Subject group: Geographic Point Location.

- § Identify the class by which this data element is represented.

Representation Class: Measure

§ One or more real world objects that identify this data element can be identified at this time.

Object: Latitude

### 5.3.8 Quality Control

When all of the mandatory metadata attributes have been entered for this data element, it is assigned the registration status of "Recorded" and the administrative status of "In Quality Review." This data element was identified in an international standard, and so would soon be updated to reflect higher status of the data element. The data element, however, would not be expected to be assigned the status of AStandard.@ The data element is not expected to be come the preferred representation for latitude measure, since geographic information specialists prefer that latitude and longitude be recorded in degrees and decimal degrees. Therefore, after quality review has been completed, the data element will be assigned the registration status of ACertified@ with an administrative status of ANo further action.@"

### 5.3.9 Other Data Elements in ISO 6709

ISO 6709 identifies five data elements: sexagesimal latitude, degrees latitude, sexagesimal longitude, degrees longitude, and altitude. The different formats represented by the units of measure for latitude (i.e., degrees and sexagesimal) express representation (i.e., unit of measure). The two latitude data elements from ANSI 6709 are translatable at the concept level, based on their unit of measure representations. They share the same conceptual domain, because their implied value meanings are the same. Likewise, the longitude data elements share a data element concept and a conceptual domain, and longitude data can be translated based on unit of measure conversions. .

Whereas the multiple data elements identified in ISO 3166 share the same data element concept and the same conceptual domain, the data elements identified in ISO do not share data element concepts and conceptual domains. All three concepts: latitude, longitude, and altitude, are distance measures. Latitude, however, is a north/south measure with respect to the equator; longitude is an east/west measure with respect to the prime meridian; and altitude is a vertical measure with respect to a point of reference such as sea level. Each has its own data element concept and its own conceptual domain.

These data elements do share classification. All can be classified as the group "Geographic Point Location" and as the representation class "Measure."

### 5.3.10 Summary of Metadata Attributes

The following table summarizes the metadata attributes assigned to latitude sexagesimal measure in the preceding paragraphs in Section 5.3. The table in Annex F also contains this data in the second metadata column.

Data Element Meta-- model Attribute Name	ISO 6709 Non-enumerated, Latitude
1. Data Element Definition and Permissible Values	
Data Element Definition	The measure in degrees of the angular distance of a position on earth on a meridian north or south of the equator.
Permissible Values	Measures of Latitude in Degrees, Minutes, and Seconds
PV Begin Date	(Not Applicable)
PV End Date	(Not Applicable)
Value Domain Definition	All measures of the distance of an angle north or south of the equator measured in degrees, minutes, and seconds.
Character Set	English language
Domain type	Non-enumerated
Determinant Type	Range
Range Limits	00-90
Datatype	Alphanumeric
Minimum	7
Maximum	13
Format	A(13) +/-DDMMSS.SSSSS
Unit of Measure	Sexagesimal
Precision	Number of decimal places recorded.
2. Data Element Name and Identifier	
Data Element Name Context	Registry
Data Element Name	Latitude Sexagesimal Measure
DE Identifier/	312345:1

<b>Data Element Meta-- model Attribute Name</b>	<b>Example</b>	<b>ISO 6709 Non-enumerated, Latitude</b>
	Version Number (DI:VI)	
3. Other Metadata Attributes		
	Example	+674532 and +674531.85435
	Origin	ISO 6709-1983 (E), Standard representation of latitude, longitude and altitude for geographic point locations.
	Note/Description	Latitude on or north of the equator is preceded by a plus sign; south of the equator by a minus sign.
	Submitting organization	Office of Information Resources Management
	Data Steward	Larry Fitzwater
4. Data Element Concept (DEC)		
	Data Element Concept Name	Latitude Distance
	Data Element Concept Definition	A measure of the angular distance of a point on the surface of the earth north or south of the equator
	Conceptual Domain Name	Latitude Coordinates
	Conceptual Domain Definition	The coordinates that indicate the distance north or south of the equator for locations.
	Enumerated Value Meaning Text	(Not Applicable)
	VM Begin Date	(Not Applicable)
	VM End Date	(Not Applicable)
5. Classification		
	Keyword	Horizontal Coordinate, Latitude
	Group	Geographic Point Locations
	Representation Class	Measure
	Object	Latitude

Data Element Meta-- model Attribute Name	Example ISO 6709 Non-enumerated, Latitude
6. Quality Control	
	Registration Status Recorded
	Administrative Status In quality review

## 5.4 Application Data Element

Application data elements are data elements that are used for a particular application. For this report, an application data element, such as is found in a computer system application has been identified as an example for data registration. Data elements used in computer systems are associated with an entity (e.g., table) and might be identified with a qualifier. The country name attribute in the mailing address entity has been selected from an information management system that contains data about facilities (i.e., the Facility Data System). This data element was selected to illustrate the relationship between an application data element and a standard data element with the same data values. It also illustrates how a well defined data element might differ from one that is identified from a computer application system. The methodology is the same as that described in Sections 5.1. It should be noted that many computer application systems contain metadata that is incomplete. Often, only the data element name, the data type and the field length are known about a data element. Where data elements can be attributed, as in this example, where the data element can reuse domain and conceptual information, based on a standard data element, the data element can be registered as Recorded. Many data elements, however, must be registered as Incomplete, and all metadata attributes identified as Mandatory, might never be complete.

The table in Section 5.4.10 contains a summary of all the metadata for the application data element described in this report .

### 5.4.1 Understanding the Data Element

The application data element for country code, used in a mailing address, must be capable of being used on a mail piece for delivery of mail to any country throughout the world. The country must be represented in such a way that it is easily read and conforms to a known identifier for that country. Therefore, authoritative names of all countries must be included in the value domain. The name must be of a length that will fit on one line of the address block.

## 5.4.2 Content Research

The United States Postal Service mailing address standard requires that the country name be included as the last line of a mail piece. Before registering a data element for the country name used in a mailing address the metadata registry for the RA is examined to determine if there is a data element, value domain or permissible values, or data element concept and conceptual domain that might be reused in attributing this data element.

A search of the registry will find that a standard data element has been registered, based on the international standard ISO 3166. The standard data element is not specific enough to describe the application of the data element in a mailing address entity. The appropriate value domain for country name to be used in a mailing address, however, should be the short name from the ISO 3166 standard. All value domain information for this application data element (i.e., country name used in a mailing address) is the same as for the ISO standard Short English-Language Country Name, described in Section 5.2, and the conceptual domain for this data element is the same. Therefore, the data element is registered, reusing domain information from the standard data element.

## 5.4.3 Definition and Permissible Values

### 5.4.3.1 Definition

The definition for the country name attribute in the mailing address entity is formulated according to the rules and guidelines listed in ISO/IEC 11179-4. The rules and guidelines are provided in Annex Y of this document, with additional examples that will provide assistance in formulation the definition. Because this data element has been submitted through a computer application system (i.e., the Facility Data System), the definition provided by the application system is retained, identified by the context for the system. Name Context for this applicaiton data element is described in Section 5.4.4.2. Definitions may be entered into the registry in conjunction with the context used for the data element name. The definition with the context for the Facility Data System is "The name of a country where the addressee is located." The Registry name context definition includes the concepts for country identifier, mailing address, and representation. The rules and guidelines specified in ISO/IEC 11179-4 are used to formulate the data element definition as "The name of the country where a mail piece is delivered."

### 5.4.3.2 Permissible Values

The permissible values for a data element are determined by the value domain. The application data element for mailing address country name uses the same permissible values as the standard data element for English-language short country names listed in the ISO 3166 standard (e.g., Afghanistan, Albania, ....., Zimbabwe). Each permissible value is entered into the registry with the date when that permissible value was valid for that value domain (in this case the date is January 10, 1997, the same as the begin date for the value meaning). There is no end date to enter at this time.

The scope of the permissible values for this data element includes the short English-language name for all countries. The value domain is described as "All short, English-language names of all countries." Note that Part 3 of ISO/IEC 11179 does not require a description or definition for enumerated domains. Some RA, however, prefer that all value domains be registered with a description/definition. Record the other value domain attributes for this example at this time, including:

- § Character Set: The character set for Short English-Language Country Names is "English language."
- § Domain Type: Country names are a fixed list of countries, maintained by international standards; therefore, the domain type is "enumerated."
- § Datatype: The datatype for country name is "alphanumeric."
- § Maximum and minimum field lengths: Based on prior research (Section 5.2.1), the minimum length for values for the data element is known to be four. The known maximum length for names in the current standard is 44. The maximum field length, however, is set to 60, to accommodate any changes or additions to the domain of values.
- § Format: The format selected by the registration authority for this example is A(60) to accommodate the longest of the English-language short names.

## 5.4.4 Identify and Name the Data Element

### 5.4.4.1 Identification

For this example, the data element for the country name used in a mailing address is assigned a unique data identifier (DI) and version identifier (VI) (5394:1) by the computer application software when it is entered into the metadata registry.

### 5.4.4.2 Name Context and Naming Convention

This data element is assigned two names, each with its own context. First is the system name context, since this data element was identified as contained in an application system, and retaining the name used by the system is valuable for documenting the system. The naming convention that has been established for this application system is as follows:

- § Scope: The scope of this example naming convention is application data elements in the Facility Data System.
- § Authority: The authority for this example is the U.S. Environmental Protection Agency for its Environmental Data Registry
- § Semantic Rules: Names shall be the same as those used by the application software, using the convention of Entity Name.Attribute Name.
- § Lexical Rules: A data element name shall have a maximum of 200 alphanumeric characters. The language of the registry shall be English, and the character set ASCII. There are no controlled word lists.
- § Name Uniqueness: Names shall be unique within a registration authority for the entity.attribute relationship.

The second name to be assigned to this data element is the registry name. It follows the naming convention for registry name context, as described in Annex Y.

- § Scope: The scope of this example naming convention is Registry Name.

- § Authority: The authority for this example is the U.S. Environmental Protection Agency for its Environmental Data Registry.
- § Semantic Rules: Names shall include an object and a property, where appropriate. Qualifiers shall be used to differentiate between names that would otherwise be the same. The representation class term shall always be included as the last term in the name.
- § Lexical Rules: A data element name shall have a maximum of 100 alphanumeric characters. The language of the registry shall be English, and the character set ASCII. There are no controlled word lists.
- § Name Uniqueness: Names shall be unique within a registration authority.

#### **5.4.4.3 Name the Data Element**

When documenting an application system, it is important to know the name of the system and the entity in which the data element exists as an attribute. This data element is assigned a name for the context "Facility Data System." It is also valuable to know the name of the attribute in that system. For this example, the system name is Facility Data System, which is documented in the registry as a system. The name of the attribute in the system is Country\_Name, and the entity name is Mailing\_Address. Therefore, the data element name for the context Facility Data System is Mailing\_Address.Country\_Name.

The data element name with Registry as context should identify the data values to be contained in the value domain (i.e., country) and the entity (i.e., address) associated with the data element. It should also include the name of the representation class. For the application data element (e.g., country name in a mailing address entity) the entity is "address" qualified by "mailing." The data values and representation are the same as for the ISO standard data element.

The qualifier is appropriate, since the registry might also have an application data element that designates the country name in a geographic (i.e., physical location) address entity. The qualifier is needed to discriminate between the country name in mailing and geographic addresses. The guidelines described in Section 5.1.3 should be followed. The registry name of this data element, based on ISO/IEC 111779-5 guidelines is "Mailing Address Country Name."

### 5.4.5 Other Metadata Attributes

Other metadata attributes that can be recorded at this time are:

- § Select the example for this data element; it must be one of the permissible values in the value domain.

Example: China

- § Identify the origin for this data element as the standard from which the permissible values are obtained.

Origin: Facility Data System, Environmental Protection Agency, Office of Enforcement and Compliance Assessment.

- § Record any notes or comments that might provide additional information about the data element that is not included in the definition.

Note: The country name is always located as the last line of a mail piece for international mailings.

- § Enter the name of the submitting organization, which is the Office that submitted the data element for registration.

Submitting Organization: Office of Enforcement and Compliance Assessment

- § Record the name of the individual or organization assigned the responsibility for monitoring and maintaining the data element as the data steward.

Data Steward: James Jones

- § Administrative metadata, such as Create Date and User Name are recorded or captured automatically by the system where applicable.

### 5.4.6 Data Element Concept

The data element concept for this data element includes the object class (entity) of address, as well as the property of being a country identifier. It does not include the qualifier for "mailing." This data element concept is not the same as the concept for the standard Country Short Name data element, which is limited to the concept of country identifier. The name of this data element concept, following the guidelines described in Section 5.1.6, is "Address Country Identifier" and the data element concept definition is "An identifier for an address of a primary geopolitical entity of the world." This data element concept could be reused for other address country identifiers, such as a geographic address country name, a geographic country code, or other representations and data element qualifiers.

The conceptual domain for this application data element is the conceptual domain for all the countries of the world. It uses the same value meanings and the same permissible values as the standard data element for country name. Therefore it reuses the conceptual domain and the value domain that were established for the ISO standard, Short English-Language Name.

### 5.4.7 Classification

This data element might be classified according to the following classification schemes:

- § Identify one or more keywords, where the keyword is a name or subject matter descriptor that will facilitate grouping like data elements for retrieval.

Keyword: Country, Mailing Address

- § Group the mailing address country name with similar data elements according to concept for translation or by general subject matter.

Subject group: Mailing Address

- § Identify the class by which this data element is represented.

Representation Class: Name

- § One or more real world objects that identify this data element can be identified at this time.

Object: Country, Mailing Address

### 5.4.8 Quality Control

When all of the mandatory metadata attributes have been entered for this data element, it is assigned the registration status of "Recorded" and the administrative status of "In Quality Review." This data element was identified by an application, and so would often not be completely attributed. This application data element, however, has been completed by reusing the value domain, permissible values, and conceptual domain of a standard data element, and so can be entered with a registration status of Recorded.

### 5.4.9 Related Data Elements

Data elements related to this application data element for Country Name are other data elements that are used in the mailing address entity, including such data elements as street name or other delivery point, city or other jurisdictional name, state or province name or code, and ZIP+4 code or other international postal code. None of these share the same value domains, conceptual domains, or permissible values. The data elements, however, can be classified as a group that make up the Mailing Address entity.

### 5.4.10 Summary of Metadata Attributes

The following table contains a summary of the values assigned to the metadata attributes in the preceding paragraphs of Section 5.4. The table in Annex F also contains this metadata.

Data Element Meta-- model Attribute Name	Example	Application Enumerated, (System Reference)	
1. Data Element Definition and Permissible Values			
Data Element Definition Context	Registry	Facility Data System	
Data Element Definition	The name of the country where a mail piece is delivered.	The name of a country where the addressee is located.	
Permissible Values	All English-Language Short Country Names from ISO 3166, matched with value meanings. (Afghanistan, Albania,....., Zimbabwe)		
PV Begin Date	19971001		

<b>Data Element Meta-- model Attribute Name</b>	<b>Example</b>	<b>Application Enumerated, (System Reference)</b>	
PV End Date		(Not Applicable)	
Value Domain Definition		All English-language short names of all countries.	
Character Set		English language	
Domain type		Enumerated	
Determinant Type		(Not Applicable)	
Range Limits		(Not Applicable)	
Datatype		Alphanumeric	
Minimum		4	
Maximum		44	
Format		A(60)	
Unit of Measure		(Not Applicable)	
Precision		(Not Applicable)	
<b>2. Data Element Name and Identifier</b>			
Data Element Name Context	Registry	Facility Data System	
Data Element Name	Mailing Address Country Name	Mailing_Address.Country_Name	
DE Identifier/ Version Number (DI:VI)	5394:1		
<b>3. Other Metadata Attributes</b>			
Example	China		
Origin	Facility Data System, Environmental Protection Agency, Office of Enforcement and Compliance Assessment		
Note/Description	This data element is required when mail is intended to be delivered outside the country of origin.		
Submitting organization	Office of Enforcement and Compliance Assessment		
Data Steward	James Jones		

Data Element Meta-- Example model Attribute Name	Application Enumerated, (System Reference)
4. Data Element Concept (DEC)	
	Data Element Concept Name Address Country Identifier
	Data Element Concept Definition An identifier for a primary geopolitical entity of the world which indicates an address.
	Conceptual Domain Name Countries of the World
	Conceptual Domain Definition The primary geopolitical entities of the world.
	Enumerated Value Meaning Text The primary geopolitical entity known as <Denmark>.
	VM Begin Date 19971001
	VM End Date (Not Applicable)
Classification	
	Keyword Country
	Group Mailing Address
	Representation Class Name
	Object Address
Quality Control	
	Registration Status Recorded
	Administrative Status In quality review

## 5.5 Register a Group of Data Elements

For some data elements, the registration authority may determine that is appropriate to group them, out of some observed relationship among the data elements or a perceived value in identifying those data elements together. After the data elements that are to be associated have been identified, the group itself is registered with the metadata that provides certain information about the group. The metadata answers the following questions: How is the group identified? Why has the group been established? What is the authority for the data elements in the group? What is the potential use for the group of data elements?

Registering a group of data elements in a metadata registry requires that certain characteristics of the group are recorded to clearly describe and define it. The data elements are then associated with the group. The characteristics are stored as attributes of the group. Attributes specific to a group, as defined by one RA are:

- § Group Name: The name of a group of data elements.
- § Group Definition: Text that describes the features of, specifies relationships of, or establishes the context for a group of data elements.
- § Authoritative Source: The originating point of information that provides an authoritative reference for a group of data elements.
- § Source Rationale: The text that explains the reasons for using the selected source materials in development of a group of data elements.
- § Potential Usage Comments: The text that describes how a group of data elements can be used.
- § Group Identifier: The system generated identifier for a group of data elements.

Groups of data elements can be registered in a registry, where a common relationship has been identified among data elements including the following:

- § Information system architecture, where the data elements make up a logical entity (e.g., mailing address).
- § Data element components, where individual data elements are grouped to make another data element (e.g., urban style street address).
- § Usage, where the elements have a common usage (e.g., data elements in a data standard).

Each of these types of groups are described in the paragraphs that follow, with a list of the data elements that have been grouped together. The table which is Exhibit 5.2 illustrates the information necessary to register the group characteristics for each of the three groups.

### **5.5.1 Information System Entity Group**

Chemical Substance Identity is an example of an entity in a system architecture where it is appropriate to group together data elements that are attributes in that entity. The list of data elements for this entity are identified as:

- X Urban-style Street Address Text. The text that describes the urban-style street name and number where the mail is delivered
- ? Post Office Box Number. The number of the post office box where mail for the addressee is delivered.
- ? Mailing Address City Name. The name of the city, town, or village where the mail is delivered.
- ? Mailing Address State Code. The alphabetic code assigned by the U.S. Postal Service that represents the state where the mail is delivered.
- ? Mailing Address Postal Code. The code that represents the code assigned by a postal service that provides information about the location of a place where mail is delivered.
- ? Mailing Address Country Name. The name of the country where a mail piece is delivered. Note: Required only for international mailings.

An example of metadata for the Mailing Address group is provided in Exhibit 5.1.

### **5.5.2 Composite Data Element**

Composite data elements are made up of more than one distinct data element that cannot be subdivided further, and that are maintained in a registry as separate data elements. Urban-style street address is an example of a composite data element. It contains the following data elements:

- ? Building Number. The number assigned to a building or a land parcel along the street to identify location and to ensure accurate mail delivery.
- ? Pre-Directional Code. The code that represents the direction the street has taken from some arbitrary starting point, and that precedes the street name.
- ? Street Name. The name assigned to a street or road, not including other urban-style street address components.
- ? Street Suffix Code. The code that represents the qualifier that follows the street name in a street address.
- ? Post-Directional Code. The code that represents the direction the street has taken from some arbitrary starting point, and that follows the street suffix.

- ? Secondary Unit Code. A code that represents the type of secondary unit where mail is delivered, e.g., the code for room, suite, or apartment.
- ? Suite Number. The number that represents the specific room, apartment, or other secondary component of an address.

Each of the data elements in the composite data element group is a distinct data element that cannot be further subdivided. The directional codes, street suffix codes, and secondary unit codes all have enumerated domains that are used to validate portions of the street address. The street address, however, is used as one item of data on a mail piece, and is, therefore, appropriately registered as an individual data element.

### 5.5.3 Use Group

An example of a group of data elements that are used together, perhaps for purposes of data translation (e.g., the ISO 3166 group of data elements that can be used to translate names and coded values that identify a country) or for data transfer (e.g., ISO 6709 that specifies formats for transfer of latitude, longitude, and altitude values that distinguish a geographic point). Data elements for a Geographic Point Location group, based on ISO 6709, include the following data elements:

- ? Latitude Degrees Measure. The measure in degrees of the angular distance of a position on earth on a meridian north or south of the equator.
- ? Longitude Degrees Measure. The measure in degrees of the angular distance of a position on earth on a meridian east or west of the prime meridian.
- ? Altitude. The measure of the distance in meters of a position above or below the surface of a reference datum.
- ? Latitude Sexagesimal Measure. The sexagesimal measure of the angular distance of a position on earth on a meridian north or south of the equator.
- ? Longitude Sexagesimal Measure. The sexagesimal measure of the angular distance of a position on earth on a meridian east or west of the prime meridian.

The latitude and longitude data elements provide information about the formats and units of measure that enable translation of the data for data sharing. The rules associated with the standard provide instructions for grouping the data elements for data sharing (e.g., latitude and longitude must be measured by the same unit when grouped together for data transfer). The following table contains examples of the metadata that should be captured about a group of data elements when the group is registered.

	<b>Information System Entity Group</b>	<b>Composite Data Element Group</b>	<b>Usage Group</b>
<b>Group Name</b>	Mailing Address	Urban-style Street Address	Geographic Point Locations
<b>Definition</b>	A set of data elements that can be used to create a mailpiece.	A set of precise and complete data elements that cannot be subdivided and that can be combined into an urban-style street address.	The horizontal and vertical coordinates and associated metadata that define a point on earth.
<b>Group Source</b>	U.S. Postal Service, Publication 28: Postal Address Standards	U.S. Postal Service, Publication 28: Postal Address Standards	International Standard ISO 6709
<b>Source Rationale</b>	The U.S. Postal Service is the nationally recognized authority for defining the requirements for creating a mailpiece, in addition to being responsible for most mail delivery within the U.S.	The U.S. Postal Service is the nationally recognized authority for defining the requirements for creating a mailpiece, and for maintaining standards and domains for formatting street address information.	ISO data standards are used internationally for consistent representation of data that enables data sharing. The standard also provides rules for formatting spatial data transfer files.
<b>Usage Comments</b>	System developers will use the Mailing Address group when creating an entity for mailing address.	The Street Address group is used to parse the components of an urban-style street address into individual segments for validation and to facilitate searching.	The geographic point locations group is used by system developers to develop a system entity for spatial data, to develop translation software, and data transfer files.

**Exhibit 5.1.** Metadata for Groups

## 5.6 Linking of Data Elements

Data elements can be linked based on their levels of abstraction. Linkages can occur in both vertical relationships and horizontal relationships, defined as follows:

§ Vertical relationships are those where a data element that has been qualified for a particular purpose is related to a more generic data element that is not qualified and is intended for a more general purpose. For example, the following data elements can be linked vertically in parent/child relationships, where 1 is the highest, and the vertical linkages increment by 1:

- 1 State USPS Code: The U.S. Postal Service abbreviation that represents a state or state equivalent for the U.S. (DI:VI 48:1)

- 2 Mailing Address State Code: The alphabetic code assigned by the U.S. Postal Service that represents the state where the mail is delivered. (DI:VI 5408:1)
- 3 Facility Mailing Address State Code: The code that represents a state of the United States in the mailing address for a facility. (DI:VI 5680:1)

§ Horizontal relationships are those where data elements with different names have equivalent definitions that represent equivalent/equal data domains. For example, the following data elements can be linked horizontally as equivalents in Envirofacts, a data warehouse of EPA environmental systems.

The third level: Facility Mailing Address State Code (DI:VI 5680:1) is linked horizontally to:

- 3a PCS\_PERMIT\_FACILITY.MAILING\_STATE The state in the primary facility mailing address. (DI:VI 24684:1)
- 3b BRS\_SITE\_INFORMATION.MAIL\_STATE The two-character state postal code for the site's mailing address. (DI:VI 23984:1)
- 3c RCR\_MAILING\_LOCATION.STATE The two-letter postal code for the state in the address associated with the facility mailing address. (DI:VI 24528:1)

## 5.7 Registration of Associated Sources/Documents

Talk about documents, citations, classified components. There are at least 5 or 6 accepted standards for citation.

Need to register forms that have about 30 to 40 data elements (items). Should include a graphic (picture) of forms.

## 6 Complex Data

Many organizations produce data for internal or external use. As a result, information that describes that data (metadata) must be readily available. With the advent of electronic access to data through the Internet and other media, the metadata must be accessible electronically, too. Registries are deployed to manage and organize the metadata, and standards such as ISO/IEC 11179 address the content and basic functions of those registries.

Organizations around the world are implementing registries based on the framework described in ISO/IEC 11179 and the metamodel defined in ANS X3.285. However, the framework has limitations that constrain the usefulness of the registries. The proposed modifications to ANS X3.285 will remedy some of these limitations.

ISO/IEC 11179 addresses the specification and standardization of data elements. The metadata that is specified in the standard describes data elements at the fundamental level. Organizations that produce and use data generate new data elements from existing ones, and the standard does not address this issue. Also, object oriented technology, multimedia applications, and advanced scientific applications produce very complex data types that are not described very well by the standard.

Some data elements are generated from other existing ones in many ways. Mathematical calculations (e.g. variance estimations), aggregation (e.g. multivariate cross tabulation), concatenation (e.g. formation of telephone number from its constituent parts), or grouping (e.g. address) are typical examples. Metadata registries that contain the descriptions of how data elements are generated from others will help users to understand the data more fully.

Even the fundamental data elements of an organization, ones that are not generated from others in the sense described above, can be generated. The functions of the business themselves can generate data elements. Identifying these functions, especially within the context of the organization, will help users increase their understanding of data.

At this point in time, the only identified types of complex data are derived data and data groups. These are defined as

- **Derived Data Element** - A data element whose values are derived through a transformation of the values of one or more other data elements. This transformation may be mathematical, logical, linkages, or some other type (including a combination of these basic types).
- **Data Group** - A set of data elements considered as a logical unit.

An important point about data groups is that they are equivalent to abstract derived data elements, where an abstract data element is a data element that is not part of a particular application. This view means that data groups don't need to be treated separately.

These minor changes to ANS X3.285 will improve the handling of complex data items:

- Have the Rule entity account for the transformation formula.
- Put an attribute on the relationship between Data Element and Rule, called role, to distinguish data elements which are input to a transformation and the data elements which are output from a transformation (derived).
- A lookup table entity, such as Derivation Type, is needed to keep track of the type of transformation used.
- A recursive or hierarchical relationship on Rule is necessary to account for combinations of transformations.

## Annex A

### Bibliography

- [1] ISO 1087:1990 Terminology - Vocabulary.
- [2] ISO/DIS 1087-1 Terminology - Vocabulary - Part 1: Theory and application (Partial revision of ISO 1087:1990).
- [3] ISO/DIS 1087-2 Terminology - Vocabulary - Part 2: Computer applications (Partial revision of ISO 1087:1990).
- [4] ISO/IEC 2382:1979-1998 Parts 1-32 Information technology - vocabulary.
- [5] ISO 2788:1986 Documentation - Guidelines for the establishment of monolingual thesauri.
- [6] ISO 3166-1:1997 Codes for the representation of names of countries and their subdivisions.
- [7] ISO 5964:1985 Documentation - Guidelines for the establishment of multilingual thesauri.
- [8] ISO 6709, 1983-05-15 Standard representation of latitude, longitude and altitude for geographic point locations.
- [9] ISO/IEC 7826-1:1994 Information technology - General structure for the interchange of code values - Part 1: Identification of coding schemes.
- [10] ISO/IEC 7826-2:1994 Information technology - General structure for the interchange of code values - Part 2: Registration of coding schemes.
- [11] ISO/IEC 11404:1996 Information technology ? Programming languages, their environments and system software interfaces ? Language-independent datatypes.
- [12] SC32 N0147 Horizontal Issues and Encodable Value Domains in Electronic Commerce.
- [13] ANSI X3.61? 1986 Representation of Geographic Point Locations for Information Interchange.
- [14] Firesmith, Donald G., *Object-Oriented Requirements Analysis and Logical Design*, John Wiley and Sons, New York, 1993.
- [15] Senehi, M.K., and Thomas R. Kramer, "A Framework for Control Architectures," *International Journal of Computer Integrated Manufacturing*, Vol. 11, No. 4, July-August, 1998, pp. 347-363.
- [16] Zachman, John A., *The Framework for Enterprise Architecture: Background, Description and Utility*, 1997, <http://www.ozemail.com.au/~visible/papers/zachman3.htm>

## **Annex B**

### **Definitions of representation class terms**

- X Amount - the sum total of two or more quantities; an aggregate.
- X Code - a symbol used to represent something.
- § Discriminator - A distinction that differentiates one from another.
- X Graphic - diagrams, graphs, mathematical curves, or the like.
- Identifier - Something that represents to be, regards, or treats as the same or identical.
- § Indicator - Anything that serves to point out or direct attention to, as of a measuring device.
- § Label - A short word or phrase descriptive of a person, group, or intellectual movement, or indicating that what follows belongs in a particular category or classification.
- X Measure - the extent, dimensions, quantity, etc. of something ascertained by comparison with a standard.
- X Name - a word or combination of words by which a person, place, object, or thought is known.
- X Number - a numeral or group of numerals.
- X Picture - a visual representation of a person, object, or scene.
- X Quantity - the property of magnitude of something.
- X Text - a unit of connected speech or writing often composed of one or more sentences that form a cohesive whole.
- § Tag - A descriptive word or phrase applied to a person, group, organization, etc. as a label or means of identification or epithet.

## Annex C

### Principles of managing shared data

These principles were used while developing the metamodel. Each principle is directly or indirectly supported by the metamodel. Conversely, this is an itemized description of much of the conceptual data structure depicted in the data model. It includes many of the more significant:

- Fundamental principles and “business rules” for data registration.
- Definitions that are applicable within the scope of this standard.
- Constraints and integrity rules for the data used for data registration.
- Structural relationships and cardinalities among data element components.
- References to terminology used elsewhere.
- Objectives for good information management.

These principles are fundamental to the use of a conformant implementation this metamodel. If the user deviates from any principle, to resulting data registry may not realize expectations.

#### C.1 Data

C.1.1 Data is a representation of a fact, idea, or instruction in a formalized manner suitable for communication, interpretation, or processing by humans or by machines. (This definition refers to a group taken as a unit thus it is used with a singular verb.)

C.1.2 Data must be able to be created, collected, organized, recorded, processed, and stored in a medium in a retrievable form.

C.1.3 Data represents data element concepts (i.e., the properties of object classes) by using a set of symbols that are perceived. These may be words made up of characters, icons, sounds, Braille, etc.

C.1.4 Data allows us to consider an object that exists in the real world without having the actual object present. In other words, data provides an abstraction of the real world object.

C.1.5 Data that is derived should be registered the same as any other data if it is stored.

C.1.6 An instantiation of a single element of data is called a “data item” (a.k.a. datum).

C.1.7 A single type or class of structured data treated as a cohesive whole is called a “data unit”.

C.1.8 A single unit of data that is considered indivisible within a universe of discourse is called a "data element". It is identical to what some others call a "simple data element".

C.1.9 Data used to describe the meaning or characteristics of data is called "metadata".

## C.2 Concept

C.2.1 A concept is a unit of thought (an idea) constituted by the abstraction of the common characteristics of a set of objects.

C.2.2 An object may be any person, place, event, or other thing that has separate and distinct existence in the real world.

C.2.3 Each concept can be shown as a more specialized type, or a component part, of one or more higher-ordered concepts.

C.2.4 A concept inherits characteristics from one or more generalized supertype(s).

## C.3 Object class

C.3.1 Humans tend to group objects when they have similar traits. When we group a set of similar things, we refer to it as a "type" or "class". A single category of "things" or "objects" is called an "object class".

C.3.2 An object class is a set of concepts, abstractions, or things in the natural world that can be identified with explicit boundaries and meaning and whose properties and behavior all follow the same rules.

C.3.3 Object classes may be a single concept or a set of concepts in a relationship with each other to form a more complex concept. Concepts in relationship with other concepts are sometimes called "concept systems".

C.3.4 Data is a representation of properties of object classes.

C.3.5 An object class is the same as an entity (entity type) or relationship in the relational paradigm.

C.3.6 It is desirable to describe object classes without redundancy within the universe of discourse. The same object class, but with different names and/or wording of definitions, should eventually be normalized.

## C.4 Property

C.4.1 A property is a classification of any feature that humans naturally use to distinguish one individual object from another.

C.4.2 When we describe an object, we describe its properties. If we know nothing about the kind of properties an object has, we are not aware of the object.

C.4.3 A property class refers to the conceptual part of an attribute, i.e., without representation.

C.4.4 A property class has no particular associated means of representation by which it can be communicated.

C.4.5 A property class may be associated with more than one object class where it describes a conceptual attribute (one without representation).

C.4.6 A property class is a concept playing the role of a property class in a data element concept. Only certain concepts have the ability to behave as a property class. Whether one of these concepts is acting as a property class cannot be determined until it is associated with an object class in a data element concept.

C.4.7 It is desirable to describe properties without redundancy within the universe of discourse. The same property class, but with different names and/or wording of definitions, should eventually be normalized (a.k.a. harmonized or rationalized).

C.4.8 Properties are sometimes called "characteristics".

#### C.5 Data element concept

C.5.1 A data element concept is the union of two or more concepts with one concept playing the role of a property.

C.5.2 A data element concept is the human perception of a single property of an object class, identified and described independently of any particular representation.

C.5.3 A data element concept has a definition different from its object class or property.

C.5.4 While any specifically defined data element concept may have several representations in a universe of discourse, each such data element concept should have a preferred data element representation in a data registry.

C.5.5 If an object class and a property are normalized across the universe of discourse, the data element concept will also be normalized.

C.5.6 Since the object class and the property have no representation, the data element concept will have no representation.

C.5.7 A data element concept may be represented as a data element.

C.5.8 Data element concepts are sometimes called "Basic Semantic Units".

## C.6 Attribute

C.6.1 An attribute is a characteristic of an object class that the business chooses to record as data.

C.6.2 An attribute is always associated with only one object class.

C.6.3 An attribute is complex. It is composed of both a property and a representation. The concept of an attribute is separate from how it is represented.

C.6.4 When a characteristic of a data unit is being described, the attribute is called a "meta-attribute".

C.6.5 The metadata used to describe data units requires many meta-attributes. A set of meta-attributes of data units bundled together as a module for reusability is called a "metadata set".

## C.7 Representation

C.7.1 Before a data element concept can become a data unit it must be expressed as a term, character, symbol, et cetera that represents a meaning of the property class. Such a notation is called "representation".

C.7.2 Representation describes how a data element concept appears in a persistent store, on a screen, on paper, et cetera. Representations are human-interpretable (sound, tactile, visual).

## C.8 Data element representation

C.8.1 A data element representation is the part of a data element having a value domain, datatype, and, if a quantity, a unit of quantity.

C.8.2 A set of similar data element representations (i.e., a "type" or "class") are grouped as a representation class for classification purposes.

C.8.3 A data element representation may be associated with one or more data element concepts.

C.8.4 The permissible values of a value domain may be expressed by specifying the range from its lower to upper limit, by a rule, by a procedure or scheme, or by enumeration in a finite list.

C.8.5 A data element representation may have a "compound datatype" that separates the representation into constituent parts. A compound datatype would only be plausible where the data element representation could be used as the representation of a single data element concept.

C.8.6 A value domain may be an aggregation of a set of smaller value domains.

C.8.7 It is desirable to describe data element representations without redundancy within the universe of discourse. Data element representations with the identical value domain, datatype, and, if a quantity, a unit of quantity, should eventually be normalized.

## C.9 Data element

C.9.1 A data element is a single unit of data that is considered indivisible in its shareable universe.

C.9.2 A data element cannot be decomposed into more fundamental constituent parts of information that have useful meanings within its shareable universe.

C.9.3 A data element is an electronic or written representation of a data element concept.

C.9.4 Data elements are the basic building blocks of data.

C.9.5 A data element is the association of a data element concept with a data element representation.

C.9.6 There may be more than one alternate way a data element concept is represented as a data element by associating it with different data element representations.

C.9.7 A data element concept associated with two or more data element representations are different data elements.

C.9.8 The term "data element" refers to a type or class (i.e., the complete set of instances) and not any particular instantiation of a value for a data element. Where a specific data element specimen occurs, it is called a "data element instance".

C.9.9 Each data element will represent no more than a single data element concept.

C.9.10 A data element is identical to an attribute in many data modeling paradigms. In a logical data model, a data element is often considered an attribute.

C.9.11 Data elements are individual, discontinuous or discrete pieces of information. They are not defined in analog or digital flows as used in electronically transmitted audio or video.

C.9.12 Data elements can be "persistent data" or "transient data" — data that is created and consumed without ever being stored in a database.

C.9.13 A data element is described independent of the physical space in which it is stored or transmitted. A single physical space (e.g., a field or column in a database) may be reused for more than one data element.

C.9.14 If a data element concept and a data element representation are normalized across the universe of discourse, the data element will also be normalized.

C.9.15 Each data element should have one identifier, one definition, one representation, one data steward, and one common set of business rules governing that element throughout the enterprise.

C.9.16 A data element is associated with a specific set of values. Any value can be expressed by a set of symbols.

C.9.17 A data element always takes on a value from a set of allowed data values. If it cannot be associated with a set of distinct values, it is not a data element. These values can include written characters, sounds, or images.

## C.10 Enumerated domain

C.10.1 Each value in an enumerated domain represents an abstraction of an object in the real world.

C.10.2 The collection of the object concepts in an enumerated domain is called a "conceptual domain". It is composed of a set of all permissible value meanings without a specified representation.

C.10.3 Once a data element concept is associated with a data element representation with an enumerated domain, a value meaning must be associated with each permissible value in the set.

C.10.4 Each value meaning in a conceptual domain may be associated with a permissible value member of more than one enumerated domain representations.

## C.11 Identifier

C.11.1 Each data element, object class, object class, property, data element concept, conceptual domain, value domain, and representation class will be uniquely identified by its identifier within a Registration Authority.

C.11.2 Identifiers will carry no intelligence.

## C.12 Name

C.12.1 A name will not be used as an identifier.

C.12.2 Various names for various contexts where the names are used and have meaning are important metadata.

C.12.3 Classification names can be constructed from the various name meta-attributes associated with object classes and representation classes.

## C.13 Quality

C.13.1 Data elements have several levels of quality.

C.13.2 All data used in the enterprise should be recognized, regardless of quality.

## C.14 Registration Authority

C.14.1 A Registration Authority is self-nominated.

C.14.2 A Registration Authority obtains a registration authority identifier.

C.14.3 A Registration Authority manages a data registry.

C.14.4 Each Registration Authority establishes the datatype categories used in its data registry.

C.14.5 Each Registration Authority establishes the procedures used to register data.

C.14.6 A Registration Authority may have an organization or individual within acting as a registrar.

## C.15 Data Registry

C.15.1 A data registry is a structure to store data about data that may be shared among Information Systems and/or organizations.

C.15.2 A data registry does not include data about Information Systems.

C.15.3 A data registry does not include data about the (conceptual, logical, or physical) structure of databases.

C.15.4 A data registry will be administered by a Registration Authority who acts as a resource to the registry's clients for establishing metadata about registered data and their applications.

C.15.5 A data registry is a place to keep characteristics of classes of objects that exist in the real world that the business chooses to record as data.

C.15.6 A data registry provides a centralized directory to describe the meaning, representation, and identification of units of data and their values.

C.15.7 A data registry enables data to be well described so that users know exactly what facts are represented.

C.15.8 A data registry supports data sharing with cross-system and cross-organization descriptions of common data.

C.15.9 A data registry is a database with appropriate analysis and user interface software.

C.15.10 A data registry may be a stand-alone system, or may be part of an Information Resource Dictionary System (IRDS) or any other information repository.

C.15.11 A data registry assists in preventing redundancy of registering the same data (described by a metadata set) multiple times within the same registry.

C.15.12 A data registry assists in preventing unplanned redundancy of the same business fact in different data elements.

C.15.13 A data registry promotes reusability of data descriptions. Metadata in a data registry should be structured as modules to maximize the reusability of these metadata sets.

C.15.14 The structure of the data registry is purposely contrived to avoid the common confusion between multiple-element units of data and single elements of data.

C.15.15 Descriptions of shareable data must be conveniently and immediately accessible to all users.

C.15.16 Registered data will be organized for easy accessibility.

C.15.17 Each data element will be classified by the object class for which it represents a property.

C.15.18 A data registry that is available to all interested parties facilitates harmonization and interchange among the parties.

C.15.19 A data registry incorporates all of the fundamental principles itemized above.

C.15.20 A data registry is sometimes called a "register".

## **Annex D**

### **Data registry uses and users**

Data users can share data if they use a common database. However, users often wish to exchange data across organizations and systems without incurring the delay and cost of creating a communal database. A more practical way of sharing data is to create a catalog of descriptions of shareable data. The catalog contains descriptions of the type of data we have reason to share with others. It does not contain any information about instances of data. It describes types of data including their allowed values. This data describing shareable data is what we call metadata.

With this approach, the key to sharing data is thus to share and reuse metadata. We can put this metadata in a catalog that is organized in a way that all stakeholders can use it. Users can have direct access to items in the catalog with convenient retrieval procedures.

When we catalog all the data used in an enterprise, we are confronted with several ways to represent the same “fact”. Information in the catalog can be organized to assist data administrators to identify redundancy. Data administrators can use the metadata catalog to standardize preferable data descriptions. By labeling well-described and sanctioned units of data in the catalog, other users will know which form of data representation to use.

Software engineers can view descriptions of data that others have already documented in the catalog. If software engineers find it easy to copy from others, they promote shareable data. The efficient software engineer can simply use what other analysts had already done. Not only will they make data shareable, their task will be easier. Also, ultimately their clients will likely be happier since this will reduce software development time. It also increases the quality of the information system product.

Electronic data interchange (EDI) data element designers' needs are similar to those of software engineers. They know what types of information trading partners need to share, but they need to describe it as data elements. If it exists in a catalog, they can use it. If it does not exist, they describe a new data element and put it into the catalog.

End users have trouble finding the data that interests them. They often do not know its definition, what it is called, the possible values, what the values mean, et cetera. The catalog can give them the information they need. Of course, the structure of the metadata must allow them to find what they are looking for. That is also true for the other users.

Originally, in its most rudimentary form, we called this catalog a data dictionary. More recently it has expanded to become the data encyclopedia. The even more comprehensive data repository or information repository came next. In the form described in this document, the directory is a data registry. The data registry is only a sub-set of the complete metadata that can be included in

a data or information repository. However, that metadata sub-set is structured in a way that supports administration and retrieval of registered data. A data registry is definitely more than just another data dictionary.

A data registry facilitates sharing data without requiring that all users obtain this data from a single communal database. Data can be shared among disparate databases and users.

## Annex E

### Conceptual and logical data models

A conceptual data model describes how relevant information is structured in the natural world. This has (somewhat inaccurately or cryptically) been called the "model of the business" (it is not always a business) or "enterprise model" (the term enterprise has several common uses). The conceptual data model provides an excellent place to start modeling data within universe of discourse. It is also the most viable level at which to integrate different data models.

A conceptual data model can be used to develop a more specific logical data model of the identical universe of discourse.

A logical data model describes the same data as structured in an information system. It is often and accurately referred to as a "model of the information system". A logical data model can be directly used for database design. This is the level where most software engineers start. This often hinders the identification of the basic concepts to be represented by the data. It also makes correct integration of data models significantly more difficult.

A conceptual data model is converted to a logical data model with several translations, additions, and decisions. Generally these:

- Add any control and interface objects or entities.
- Eliminate or resolve many-to-many relationships.
- Combine entities with one-to-one relationships.
- Identify key attributes.
- Decide which entities from related entities can become attributes based upon the intended use and importance of the data.
- Specify which entities will inherit foreign keys.
- Specify representation class, datatype, character count, and other value domain metadata attributes that describe the data elements used to represent the data element concepts described in the conceptual model.
- Convert all special relationships such as subtypes, components, and dependencies into conventional relationships.
- Specify whether each attribute is mandatory, conditional, or optional.

## Annex F

### Table of Data Element Attributes for Examples (Informative)

Annex F contains a table that includes the data element attributes for the examples provided earlier in this document. The table provides examples of the metadata associated with three data elements from the ISO 3166 standard (i.e., Country Short Name, Country Long Name, and Country Numeric Code), an illustrative application data element, and two data elements from the ISO 6709 standard. The data element attributes are given in the first column and the illustrative data that could be registered for each of the example data elements is given in subsequent columns.

**Table of Data Element Attributes for Examples (Informative)**

Data Element Meta-- model Attribute Name	Example ISO 3166 Enumerated, Name	ISO 6709 Non-enumerated, Latitude	Application Enumerated, (System Reference)	
1. Data Element Definition and Permissible Values				
Data Element Definition Context	Registry	Registry	Registry	Facility Data System
Data Element Definition	The English-language short name of a country.	The measure in degrees of the angular distance of a position on earth on a meridian north or south of the equator.	The name of the country where a mail piece is delivered.	The name of a country where the addressee is located.
Permissible Values	All English-Language Short Country Names from ISO 3166, matched with value meanings. (Afghanistan, Albania,....., Zimbabwe)	Measures of Latitude in Degrees, Minutes, and Seconds	All English-Language Short Country Names from ISO 3166, matched with value meanings. (Afghanistan, Albania,....., Zimbabwe)	
PV Begin Date	19971001	(Not Applicable)	19971001	
PV End Date	(Not Applicable)	(Not Applicable)	(Not Applicable)	
Value Domain Definition	All English-language short names of all countries.	All measures of the distance of an angle north or south of the equator measured in degrees, minutes, and seconds.	All English-language short names of all countries.	
Character Set	English language	English language	English language	

<b>Data Element Meta-- model Attribute Name</b>	<b>ISO 3166 Enumerated, Name</b>	<b>ISO 6709 Non-enumerated, Latitude</b>	<b>Application Enumerated, (System Reference)</b>	
Domain type	Enumerated	Non-enumerated	Enumerated	
Determinant Type	(Not Applicable)	Range	(Not Applicable)	
Range Limits	(Not Applicable)	00-90 for degrees	(Not Applicable)	
Datatype	Alphanumeric	Alphanumeric	Alphanumeric	
Minimum	4	7	4	
Maximum	44	13	44	
Format	A(60)	A(13) +/-DDMMSS.SSSS	A(60)	
Unit of Measure	(Not Applicable)	Sexagesimal	(Not Applicable)	
Precision	(Not Applicable)	Number of decimal places recorded.	(Not Applicable)	
<b>2. Data Element Name and Identifier</b>				
Data Element Name Context	Registry	Registry	Registry	Facility Data System
Data Element Name	Short English-Language Country Name	Latitude Sexagesimal Measure	Mailing Address Country Name	Mailing_Address. Country_Name
DE Identifier/ Version Number (DI:VI)	20903:1	312345:1	5394:1	
<b>3. Other Metadata Attributes</b>				
Example	China	+674532 and +674531.85435	China	
Origin	ISO 3166-1:1997, Codes for the representation of names of countries and their subdivisions B Part 1: Country codes (Document)	ISO 6709-1983 (E), Standard representation of latitude, longitude and altitude for geographic point locations.	Facility Data System, Environmental Protection Agency, Office of Enforcement and Compliance Assessment	
Note/Description	This data element is included in the EPA revised interim Facility Identification Standard.	Latitude sexagesimal converts to latitude degrees by the following formula: seconds x 60 = decimal minutes, total minutes x 60 = decimal degrees.	This data element is required when mail is intended to be delivered outside the country of origin.	
Submitting organization	Office of Information Resources Management	Office of Information Resources Management	Office of Enforcement and Compliance Assessment	
Data Steward	Marian Cody	Larry Fitzwater	James Jones	

<b>Data Element Meta-- model Attribute Name</b>	<b>Data Element Example</b>	<b>ISO 3166 Enumerated, Name</b>	<b>ISO 6709 Non-enumerated, Latitude</b>	<b>Application Enumerated, (System Reference)</b>
<b>4. Data Element Concept (DEC)</b>				
Data Element Concept Name	Country Identifier	Country Identifier	Latitude Distance	Address Country Identifier
Data Element Concept Definition	An identifier for a primary geopolitical entity of the world.	An identifier for a primary geopolitical entity of the world.	A measure of the angular distance of a point on the surface of the earth north or south of the equator	An identifier for an address of a primary geopolitical entity of the world.
Conceptual Domain Name	Countries of the World	Countries of the World	Latitude Coordinates	Countries of the World
Conceptual Domain Definition	The primary geopolitical entities of the world.	The primary geopolitical entities of the world.	The coordinates that indicate the distance north or south of the equator for locations.	The primary geopolitical entities of the world.
Enumerated Value Meaning Text	The primary geopolitical entity known as <Denmark>.	The primary geopolitical entity known as <Denmark>.	(Not Applicable)	The primary geopolitical entity known as <Denmark>.
VM Begin Date	19971001	19971001	(Not Applicable)	19971001
VM End Date	(Not Applicable)	(Not Applicable)	(Not Applicable)	(Not Applicable)
<b>Classification</b>				
Keyword	Country	Country	Horizontal Coordinate, Latitude	Country
Group	Country Identifiers, Geopolitical Entities	Country Identifiers, Geopolitical Entities	Geographic Point Locations	Mailing Address
Representation Class	Name	Name	Measure	Name
Object	Country	Country	Latitude	Address
<b>Quality Control</b>				
Registration Status	Standard	Standard	Certified	Recorded
Administrative Status	Final	Final	No Further Action	In quality review

## Annex G

### Top down approach to data element registration

A small amount of data that are added to a registry comes in groups or classifications (e.g., Chemical Substances or Biological Taxonomy), rather than as individual data elements. When a classified group of data elements is to be added to the registry, the analyst might choose to identify the conceptual domains that are relevant to the group, consider their value meanings, and work down to data elements. For the purpose of this informative annex, the group Biological Taxonomy will be used as the example.

More than one conceptual domain might be identified at the start. Names and definitions for these might include:

- 1) Biological Organisms ?All life forms considered as entities.
- 2) Biological Organism Types ?All ways of typing biological organisms.

#### G.1 Biological Organisms

Starting with the first conceptual domain, Biological Organisms, we must envision the value meanings that would be appropriate for Biological Organisms. Just as the value meanings for Countries of the World are "The principal geopolitical entity of the world known as ...." where the entity might be France, Germany, Canada, or any of the countries of the world, the value meaning of Biological Organisms would be "The biological organism known as ...."

An essential difference between the two conceptual domains is that we know the names of the "Countries of the World." We do not, however, intend to enumerate all of the life forms that are known. The value meanings for Biological Organisms will not be identified and listed, but will be determined from references. Therefore, only non-enumerated domains will be associated with this conceptual domain.

##### G.1.1 Data Element Concepts

One data element concept that would be associated with Biological Organisms would be "Biological Organism Label," where "Biological Organism" would be the object, and "Label" the property. **Note:** Label is defined as a short word indicating that what follows belongs in a particular category or classification (see 5.1.6). The definition of this data element concept would be "A label that identifies a biological organism."

##### G.1.2 Data Elements

Data elements to be associated with the "Biological Organism Label" would be all of the names, codes, and identification numbers associated with biological organisms, including:

- X Biological Organism Taxonomic Name? The systematic name that provides a definitive classification for a biological organism.
- X Biological Organism Vernacular Name? The common name that is associated with a biological organism.

- X ITIS Taxonomic Serial Number? The unique number assigned to a biological organism by the Integrated Taxonomic Information System (ITIS)<sup>2</sup>.
- X Biological Identification Number? The unique number assigned to a biological organism by the Biological Registry System.

### G.1.3 Permissible Values

Permissible values for these data elements would not be enumerated, as described above in Section G.1. The permissible values, however, will all be names, numbers, and codes that represent an implied value meaning of "The biological organism known as...".

## G.2 Biological Organism Types

Biological information can be separated into several categories or types of related entities. Types of biological organisms can be limited for a particular application, and can be expected to have value meanings associated with them. The selection of the types to be included and the definition of each grouping could be based on widely accepted criteria or useful only for a specific application. For example, the types of biological organisms in this sample scheme could include:

- X Biota? An animal, plant, fungus, or other biological organism of a region or period.
- X Virus? An ultramicroscopic agent that replicates only within the cells of living hosts, which are mainly bacteria, plants, and animals.
- X Group? A collection of biological organisms that are related in some way.

**Note:** The selection of these types, for this example, is based on the fact that ITIS currently does not contain information on viruses and groups. ITIS Taxonomic Serial Numbers would be available only for each biota. Virus identification would come from The Universal Virus Database (<http://life.anu.edu.au/viruses/welcome.htm>). Groups would include such things as macro-invertebrates, minnows, and coliform that are counted and recorded as aggregates in environmental studies. Although ITIS currently does not contain identification for groups of organisms, it might store information about the individual organisms that are members of a group.

### G.2.1 Data Element Concepts

A data element concept associated with the conceptual domain "Biological Organism Types," might be "Biological Organism Type," where Biological Organism is the Object, and Type is the property. **Note:** It is not always necessary to include the word Label in a Data element concept name. The definition of the data element concept might be "A type of a biological organism."

---

<sup>2</sup>The ITIS is a partnership of U.S., Canadian, and Mexican agencies, other organizations, and taxonomic specialists cooperating on the development of an on-line, scientifically credible, list of biological names focusing on the biota of North America. ITIS uses the five kingdom system for identification and assigns taxonomic serial numbers to each taxonomic level in an identification. ITIS is meant to serve as a standard to enable the comparison of biodiversity datasets, and therefore aims to incorporate classifications that have gained broad acceptance in the taxonomic literature and by professionals who work with the taxa concerned.

### **G.2.2 Data Elements**

Data elements associated with this data element concept might be:

- X Biological Organism Type Name? The name of the type of a biological organism.
- X Biological Organism Type Code? The code that represents a type of biological organism.

### **G.2.3 Permissible Values**

Permissible values for the "Name" representation would be the same names as the value meaning names, and the "Code" representation would be some kind of number or character used to represent the Type.

### **G.3 Top Down Population of a Registry**

The information that is included in a registry would be the same as that shown in Annex F, but the order of population would be different. The following is a reordering of the first column of Annex F to illustrate the top down approach to registry population.

Conceptual Domain (CD) Name  
Conceptual Domain Definition  
CD ID  
Value Meanings  
    VM Begin Date  
    VM End Date  
    VM ID  
Data Element Concept (DEC) Name  
Data Element Concept Definition  
    DEC ID  
Representation  
Value Domain (VD)  
    VD ID  
    Domain Type  
        Determinant type  
        Range limits  
    Datatype  
    Format  
    Minimum  
    Maximum  
    Unit of Measure  
    Precision  
Data Element Name Context  
Data Element Definition  
Data Element Name  
DI:VI  
Permissible Values  
    PV Begin Date  
    PV End Date  
Example  
Origin

September 16, 1999

Note/Description

## ANNEX Y

### Business Rules for Populating a Metadata Registry ANNEX Y

This annex includes information on how to record particular metadata attributes, in more detail than described in Section 5.1 of this technical report.

#### Y.1 Data Element Definition

The purpose of a data element definition is to define a data element with words or phrases that describe, explain, or make definite and clear its meaning. Precise and unambiguous data element definitions are one of the most critical aspects of ensuring data shareability. The value domain, described in Section Y.2, identifies the complete set of values that can be contained in a data element. Each data value in a domain must conform to the definition for that data element.

ISO/IEC 11179-4 provides the standard for formulating data element definitions. There are mandatory rules, to which all data element definitions must comply, and there are guidelines which should be followed in formulating a definition. The standard does not specify syntactical requirements (i.e., word order and structure), which may be established by the registration authority. A registration authority might choose to allow multiple definitions, in context, for a data element in the same manner that multiple names, in context, are allowed. In the case of multiple definitions, each definition must convey the same, exact meaning so that there is no ambiguity to the values for that data element. See Section 5.4.1.2 for examples of names and definitions in context.

The rules and guidelines applicable to the Registry Definition (i.e., the unique definition that has been assigned to the data element for registration in a metadata registry) follow. A syntax that has been adopted by one registration authority is also included in this section.

##### Y.1.1 Mandatory Rules

Rules for formulating a data element definition are mandatory and testable for compliance. The following rules must be followed when formulating a data element definition:

- § Unique (within any data dictionary in which it appears).
- § Singular.
- § State what the concept is, not only what it is not (i.e., never exclusively in the negative).
- § Descriptive phrase or sentence.
- § Contain only commonly used abbreviations.
- § Does not contain embedded definitions of other data elements or concepts.

Examples of definitions that meet the above requirements are described in the following paragraphs.

#### **Y.1.1.1 Uniqueness**

According to the standard rules for formulating data definitions, a data definition shall be unique within any data registry and registration authority in which it appears. Each definition shall be distinguishable from every other definition within a registration authority to ensure that specificity is maintained. One or more characteristics expressed in the definition must differentiate its concept from other concepts.

Note that a registration authority that registers incomplete application data elements might contain several data elements with the same definition, each within the context of the source of that data element. These data elements should be linked to the appropriate well-formulated data elements that contain the same data values. See Section 5.6.5 for linking of data elements.

**Good:** Regulation Effective Date: The calendar date when a regulation became effective.  
Sample Collection Start Date: The calendar date when collection of the sample began.

**Poor:** Regulation Effective Date: The date when the event started.  
Sample Collection Start Date: The date when the event started.

#### **Y.1.1.2 Singular**

The concept expressed by the data definition shall be expressed in the singular.

**Good:** The commonly known, short name of a country.

**Poor:** The commonly known, short name of countries.

**Note:** The poor definition implies that a name might identify more than one country.

#### **Y.1.1.3 State the Concept; Not Only its Negative**

A definition cannot be constructed exclusively by saying what the concept is not. The following are definitions of "Country Name" demonstrate good and bad definitions.

**Good:** The commonly known, short name of a country.

**Poor:** The name that is not the long name of a country.

**Note:** In some instances, a good definition that specifies what the concept is, might also specify what the concept is not, as in the following example:

**Good:** The commonly known, short name of a country *that is not its long name.*

#### **Y.1.1.4 Descriptive Phrase or Sentence**

A phrase or sentence is necessary to describe the essential characteristics of the concept. Stating the name as a synonym, or restating it with the same words is insufficient.

**Good:** The commonly known, short name that identifies a country.

**Poor:** Name of a country.

**Note:** The poor definition does not describe the concept that this is the short name, not an expanded or long name.

#### **Y.1.1.5 Contain Only Commonly Used Abbreviations**

Understanding the meaning of an abbreviation, including acronyms and initials, is usually confined to a certain environment. In other environments the same abbreviation can cause misinterpretation or confusion. An exception to this rule can be made if an abbreviation is more readily understood than the full form and has been adopted as a term in its own right, such as *email* (i.e., electronic mail), *radar* (i.e., radio detecting and ranging) and *fax* (i.e., facsimile). When an abbreviation or an acronym is included in a definition, it should follow the full term and be enclosed in parentheses.

Example 1:

**Good:** The code that represents the economic activity of a company as specified by the Standard Industrial Classification (SIC) of Establishments.

**Poor:** The SIC code for a company.

Example 2:

**Good:** The code that represents the unit for measuring the mass per unit (m.p.u.) volume.

**Poor:** The code that represents the unit for measuring the m.p.u. volume.

#### **Y.1.1.6 No Embedded Definitions**

The definition of a second data element or related concept should not appear in the definition proper of the primary data element.

**Good:** The text that describes the method used to calibrate the analysis equipment.

**Poor:** The text that describes the method used to calibrate the analysis equipment. Calibration is the process of rectifying the graduation of an instrument that gives quantitative measurements.

**Note:** The term calibration should be defined in an associated glossary or dictionary.

## Y.1.2 Guidelines for Definitions

Highly recommended guidelines, although not mandatory, are principles that should be followed when formulating a data element definition. A definition should:

- § State the essential meaning of the concept.
- § Be precise and unambiguous.
- § Be concise.
- § Be able to stand alone.
- § Be expressed without embedding rationale, functional usage, domain information, or procedural information.
- § Avoid circular reasoning.
- § Use the same terminology and consistent logical structure for related definitions.

Examples of these guidelines are provided in the following paragraphs.

### Y.1.2.1 Essential Meaning of Concept

Include all primary aspects of the concept, but avoid non-essential characteristics.

**Good:** The name of a country where mail is delivered.

**Poor:** The last line of a mail piece that names the country where mail is being sent.

**Note:** The poor definition contains extraneous information (i.e., the line where the country name is placed on a mail piece). This information is valuable to those who are preparing mail pieces (e.g., letters and packages), but does not serve to define the data element. This information might be included in a comment about the data element, or in business rules applicable to mailing address.

### Y.1.2.2 Precise and Unambiguous

The exact meaning of a data element should be apparent from the definition. Codes that are derived from different standards or identifiers assigned by different sources must be distinguished.

Example 1:

**Good:** The 2-character alphabetic code assigned by the International Standard Organization (ISO) 3166-1 to represent a country.

**Poor:** The code that represents a country.

**Note:** Country Codes are assigned by ISO 3166-1:1997, FIPS PUB 10-4, FIPS PUB 104-1, and ANSI Z39.27-1984. Some are alphabetic (both 2- and 3-character), and at least one is numeric. The poor definition is imprecise, making it difficult to clarify the source of the code and its decode.

**Note:** The source of standard data values in a domain are documented by association with the source of those values. The source is sometimes reflected in the definition, however, so that there is no misunderstanding as to the source of the data content for the data element.

Example 2:

Other examples of **good** definitions that clearly distinguish between similar data elements are:

§ The commonly recognized, short name that identifies a country.

§ The official name that identifies a country.

### Y.1.2.3 Concise

The definition should be brief and comprehensive. Extraneous terms are to be avoided.

**Good:** The surname of a person.

**Poor:** The part of a person's name that describes the surname of a person.

**Note:** The person=s surname does not describe the surname - it is the surname of a person. It is extraneous to say that the surname is "part of a person's name."

### Y.1.2.4 Stand Alone

A good definition must be able to stand alone, without further definition to understand its meaning.

**Good:** The Hydrologic Unit Code (HUC) that represents a geographic area that includes part or all of a surface drainage basin, a combination of drainage basins, or a distinct hydrologic feature.

**Poor:** The Hydrologic Unit Code (HUC) code that represents a cataloging unit.

**Note:** The term "cataloging unit" does not provide the understanding that the code represents a drainage basin. For data registries that include a dictionary or thesaurus, the term cataloging unit should be defined in the thesaurus.

### Y.1.2.5 No Embedded Information

A good definition does not include embedded rationale, functional usage, domain information, or procedural information.

Example: The rationale for using meters instead of feet should not be embedded in the definition.

**Good:** The distance in meters either above or below a reference surface.

**Poor:** The distance either above or below a reference surface, measured in meters instead of feet because meters is an international standard for measuring distance.

Example: Functional usage should not be included in the definition (i.e., this data element is [or is not] used for..).

**Good:** The code assigned by a state to uniquely identify a facility.

**Poor:** The code assigned by a state to uniquely identify a facility and to be used by the state in all data transfer for that facility.

Example: Procedural remarks (e.g., optionality) should not be part of a data element definition.

**Good:** The name of the capacity that an organization serves for a facility.

**Poor:** The name of the capacity that a company serves for a facility. The role name is used in conjunction with an organization name in association with a facility.

**Note:** A data element may have a "Note" or "Comment" attribute that can be used to capture usage, procedure, and other explanatory information that is not appropriate to include in the definition attribute.

#### **Y.1.2.6 Avoid Circular Reasoning**

Two definitions should not be defined in terms of each other. A definition should not use another concept's definition as its definition. Examples of poor definitions with circular reasoning are:

**Poor:** A code number assigned to an object.

**Poor:** An object identified by a code number.

#### **Y.1.2.7 Consistency for Related Definitions**

A common terminology and syntax (i.e., consistent logical structure) should be used for similar or related definitions to facilitate understanding. Where the terminology and syntax is not the same, a user might assume that there is an implied difference between related definitions.

**Good Consistency.** The following three definitions represent good consistency for the code and the name of the method for determining the vertical coordinate, and also with the name of the method for determining vertical and horizontal coordinates:

The code that represents the method used to determine the vertical coordinate.

The name of the method used to determine the vertical coordinate.

The name of the method used to determine the horizontal coordinates.

**Poor Consistency.** The following two definitions represent poor consistency for code and name of the method for determining horizontal coordinates:

The name of the method used to determine the horizontal coordinates.

The code that represents the method used to determine the latitude and longitude.

**Note:** Because the terminology is different (horizontal coordinates vs. latitude and longitude), the registry user might assume that the different terms have a somewhat different meaning, even though they are simply different representations of the same concept.

### Y.1.3 Data Element Definition Syntax

Only semantic structures of data element definitions are addressed in ISO/IEC 11179-4. For consistency, a registration authority might choose to establish syntax rules for the registry, as in the following example:

C Use a phrase, not a sentence.

**Phrase:** The name of the country where a mail piece is delivered.

**Sentence:** The mailing address country name is the name of the country where a mail piece is delivered.

**Note:** The sentence above is not as concise as the phrase, it repeats the data element name, and adds nothing that clarifies or further defines the data element.

C Since a data element always includes representation, begin the phrase that defines the data element by stating the representation class for the data element and its value domain. The definite article "the" is used, because the definition refers to a specific data value.

**Name:** The name of ....

**Code:** The code that represents ....

**Text:** The text that describes (or defines)....

**Number:** The number assigned by (Dun & Bradstreet; Chemical Abstracts Service; the state) to identify a (business establishment, chemical substance, legislative district)....

OR The number that represents ....

**Measure:** The measure of the (distance, area, mass)....

**Picture:** The picture of ....

**Graphic:** The graph that depicts ....

**Quantity:** The (sum, dimension, capacity, amount) of ....

**Note:** For quantity, instead of repeating the term "quantity" in the definition, more specific terms are used to describe the type of quantity for which the data element is applicable. This avoids the wordiness of a phrase such as "The quantity that indicates the sum of ...."

#### **Y.1.4 Terms Commonly Used in Definitions**

Although not part of the standard, there are action terms commonly used in definitions that are frequently misused or mistakenly interchanged. The terms have similar, but different, meanings that make subtle changes to the interpretation of the definitions. These terms might be included in a user manual, to provide guidance for formulating definitions. The following are examples of terms that a registration authority might designate to be used in definitions, according to the meanings provided:

- § Define. To set forth the meaning of a word or phrase.
- § Depict. To represent by, or as if by painting, or to characterize by words with vividness of detail.
- § Describe. To convey in words the appearance, nature, or attributes of something.
- § Designate. To select or nominate for a purpose.
- § Identify. To recognize or establish as being a particular person or thing; to verify the identity of something.
- § Indicate. To show (as by measuring or recording), point to, draw attention to, or make known briefly in a general way.

For definitions to be precise and unambiguous, the above terms should be used carefully so that the exact meaning of the concepts reflected by the definitions is well understood.

#### **Y.2 Representational Attributes**

One of the first things to consider when registering a data element is how the data element is to be represented in an implementation. The relational aspects of a data element include the permissible values (i.e., code sets), value domain, representation class, and examples of data values. The value domain is the set of permissible values that will be stored in the data element as well as other representational attributes.

### Y.2.1 Permissible Values

Permissible values are the exact names, codes, and text that can be stored in a data field in an information management system. For value domains that are enumerated, permissible values must be entered into the registry. The permissible values for country identification in "Short, English-Language Country Name" will be those names that are listed in the ISO 3166 standard for that category.

The permissible values for an enumerated value domain are associated with the value meanings (i.e., the names and definitions that are included in the conceptual domain of possible values). The entry of value meanings and their association with permissible values is described later in this Annex as Y.5.3.

For non-enumerated domains, the permissible values are those defined by the value domain description/definition and the rule description, as described in Section Y.2.2.

### Y.2.2 Value Domain

The value domain is formulated, based on an understanding of the data content. A data element is associated with only one value domain, and the name of the value domain describes all of the data values that are included in that domain. Value domains can have the attributes identified in the following list, not all of which are in the standard. Data elements referenced in ISO/IEC TR 15452, *Information technology, Specification of data value domains*, are indicated with an asterisk (\*), and those additional attributes also referenced in the ISO/IEC 11179-3, *Information technology - Specification and standardization of data elements Part 3: Basic attributes of data elements*, are indicated with a double asterisk (\*\*).

(Note: Part 3 defines value domain as "A set of permissible values. It provides representation, but has no implication as to what data element concept the values are associated with nor what the values mean.")

- §     \*\*Label. The record identifier that represents the value domain. Each value domain must have an identifier, which can be generated by computer software to ensure uniqueness.
- §     \*Name. The name by which a value domain is known. The name should be plural, since a value domain encompasses all values that are included in the domain (e.g., Short English-Language Country Names). Note that a definition can also be used to describe the value domain.
- §     \*Character Set. The collective symbols of a formalized writing system for a language used to intelligibly communicate data. The descriptor >character set= of a data element attribute is valid at the data element dictionary level and shall be explicitly stated in case

of interchange among dictionaries. If one or more of the data element attributes uses a character set that differs from the set generally used for the complete data element dictionary, then the descriptor >character set= shall be specified. Examples of character sets are "ASCII" (i.e., consisting of 128, 7-bit characters) and "EBCDIC " (i.e., consisting of 256, 8-bit characters).

For the examples described in this technical report, the character set does not need to be specified.

(Note: There is a discrepancy between TR 15452 and Part 3 regarding character set. TR 15452 indicates that character set can be "alphabetic character" or "numeric character," both of which are described as "datatype" in Part 3. Part 3 defines character set as in the above paragraph.)

§ **\*\*Datatype.** The format used for the collection of letters, digits, and/or symbols, to depict values of a data element, determined by the operations that may be performed on the data element. Datatypes are characterized as language independent. They do not follow any particular Database Management System (DBMS) or software language. The standard does not specify the datatypes to be used for the value domains. They must be established by the registration authority. The registration authority might choose to record datatypes in context (e.g., ORACLE or COBOL), in which case the context for the datatype should also be recorded.

An alphanumeric datatype is composed of either alphabetic characters, numerals, or both. A numeric datatype is composed of numerals. In general, values that are intended to be sorted, whether numerals or alphabetic characters, are described as "alphanumeric." Only numbers that are used in calculations are given the datatype of "numeric." The character set for date (i.e., day of a calendar year) has been identified as "date," and whole numbers as "integers." When creating metadata for more complex datatypes (e.g., arrays and bit strings), ISO 11404 provides guidance on datatypes.

§ **Domain Type.** Value domains are either enumerated or non-enumerated:

(Note that TR 15452 addresses enumerated domains only. Part 3 describes enumerated and non-enumerated domains, but does not provide for an attribute to distinguish between them.)

Enumerated domains are those for which all values can be explicitly expressed in a structured or unstructured set. Structured sets (e.g., taxonomies or thesauri) are not addressed in this document. Country names are a fixed list of countries, maintained by international standards; therefore, the domain type is enumerated.

Non-enumerated domains have an unspecified set of values. The values, however, must fall within the scope of the definition. Latitude measures are not restricted to a fixed list. Therefore,

the domain type is non-enumerated. A non-enumerated domain must be described by exactly one "non-enumerated domain description."

§ \*\*Value Domain Description/Definition. Non-enumerated domains must include a textual description of the potentially valid values to be stored in the data element.

§ \*\*Non-enumerated domain description. A designation of procedure or rule for a set of all permissible values for the value domain or the upper and lower limit to a value domain range. The non-enumerated domain must be described as one of the following:

- Procedure. Measurements and quantities are determined by procedure (e.g., they are calculated, measured, or generated).
- Reference. Telephone numbers and facility names are determined by reference (e.g., they can be validated in some type of directory).
- Range. Percentages and temperatures are examples of range determinations. Maximum and minimum values are always required for range determinations. Examples: 1-100% and 32-212°F.

§ \*\*Rule description. The rule is the logical, mathematical, or other operation that specifies the derivation for a data element. The rule description specifies the derivation of the data element values. For non-enumerated value domains, the rule description describes the procedure, the reference, or the maximum and minimum values for the range that limits the permissible values for a data element.

§ \*Maximum and minimum field lengths.

For non-enumerated domains, the minimum length can be as small as one; the maximum length must be adequate to accommodate the largest, reasonable amount of data for that value domain (e.g., the maximum length for a text field might be 240 characters).

For enumerated domains, the actual permissible values determine the minimum and maximum field lengths. For a 3-digit code, both the minimum and maximum field lengths are three. For short, English-language country names, the minimum length is 4 (e.g., Peru or Oman) and the maximum length is 44 (e.g., South Georgia and the South Sandwich Islands).

§ \*Format. The format is a template for the structure of the elements of a value domain. A registry might adopt its own format for displaying data element format, independent of the DBMS or software language. For example, alphanumeric might be depicted as  $A(n)$ , where "A" represents alphanumeric and "n" is the maximum field length for the data element value. Numerics might be depicted as  $N(n.d)$  where the data value has  $n$ -digits

to the left and *d*-digits to the right of the decimal point. Integer format might be depicted as I and date as D. The format must distinguish between integers, decimal marks, and floating point notations. It must also reflect any embedded punctuation in the stored data element. Note that ISO 6093 provides guidance on formats.

- § **\*\*Unit of Measure.** Some value domains require that values for a data element be measured in only one unit (e.g., a requirement that altitude be measured in meters). This attribute contains the name of the unit of measure for all data values for the value domain.
- § **\*\*Precision.** Where the value for a data element must be measured or recorded according to a specific level of precision, that information is recorded in the precision attribute (e.g., a requirement that the molecular weight for a chemical substance be recorded to two decimal places). Examples of value domain identifiers (i.e., labels) have been assigned to the examples provided in Annex F to demonstrate uniqueness and reusability of the value domain.

### Y.2.3 Representational Terms

Representation is the form of expression of the data element. Representation and value domain together provide the data element representation. Representation terms are used to describe the form of representation of a data element. An informational list of representation terms is provided in ISO/IEC 11179-5. The list has been expanded in this document to provide a more comprehensive list of examples that might be used to describe representation classes, including the following:

- § **Amount.** The sum total of two or more quantities; an aggregate.
- § **Code.** A symbol used to represent something.
- § **Graphic.** Diagrams, graphs, mathematical curves, or the like.
- § **Icon.** A sign or representation that stands for its object by virtue of a resemblance or analogy to it.
- § **Measure.** The extent, dimensions, quantity, etc. of something ascertained by comparison with a standard.
- § **Name.** A word or combination of words by which a person, place, object, or thought is known.
- § **Number.** A numeral or group of numerals.
- § **Picture.** A visual representation of a person, object, or scene.

- § Quantity. The property of magnitude of something.
- § Text. A unit of connected speech or writing often composed of one or more sentences that form a cohesive whole.

#### **Y.2.4 Example**

Each set of metadata attributes for a data element includes an example of the kind of data value that can be stored in that data element. Data element names and definitions are always defined as singular; therefore, examples are always singular. More than one example can be used, however, where necessary to illustrate the value domain. The example can be a name, text, code, number, or any of the data representations described in the value domain. The following rules apply:

- ? For enumerated domains, the data element example must be one of the permitted values for that value domain.

Example for "Country Name": Australia

When the representation for the data element is a coded value, a registration authority might choose to use one of the permitted values for the code as the example, followed by the value meaning name, enclosed in parentheses.

Example for "Country Numeric Code": 036 (Australia)

- ? For non-enumerated domains, the data element example must be representative of the data that complies with the definition of the value domain.

Example for "Latitude Degrees Measure": 87.123456

Example for "Location Comments Text": The coordinates reference the flag pole in the North parking lot of the installation. This location is near the center of the facility.

### **Y.3 Identifying and Naming a Data Element**

The data element name can be constructed, based on the value domain values and the data element definitions.

Names are not used as identifiers for data elements, but as designators that enable humans to refer to a data element. The definition is the attribute that provides a full understanding of the data element, and the data identifier, version identifier, and registration authority identifier together uniquely identify a data element, as described in ISO/IEC 11179-5.

Every data element must have at least one name, and each name must be identified with a context. Each context (e.g., source of a data element name) can have its own naming convention. Rules for formulating a data element name are dependent upon the registry in which the data element is registered. An example follows in Section Y.3.3.

Multiple names may be appropriate for a data element based on the intended use for the data element.. Contexts for names are described in Section Y.3.1. Each data registry establishes its own naming convention. Suggestions for establishing a naming convention are provided in Section Y.3.2.

### Y.3.1 Name Context

Context names are not listed in the standard. Examples of name contexts that might be used for a registration authority include:

- § Legacy - a name that has been used in the past.
- § Standard - a name that has been used in a standard (e.g., ANSI, ISO, or other standard).
- § Short Abbreviation - a name that is used in a computer system.
- § *<source system name>* - the name that is used by the source that submitted the data element for registration.
- § Registry - the unique name that has been assigned to the data element for registration by a registration authority.

The multiple names for a single data element might be the same or different names, depending upon their contexts. The names in context are often associated with definitions for that context. The definitions must state the exact same concept for the data element as the registry definition, even if they are defined in different terms. Examples of non-unique names and definitions, associated with the same data element but stating the same concept, are listed as follows:

- Registry:** Vertical Measure. The vertical measure, in meters, of the measured point, above or below a reference point.
- Legacy:** Vertical Measure. The measure of elevation (i.e., the altitude), in meters, above or below a reference datum.
- Standard:** Altitude. The vertical distance in meters either above or below a reference surface.

It is clear when reading these three definitions, that the concept is the same for all (i.e., the measure of the height (or depth) of an object above or below some point of reference). The following definition would not be appropriate, because it would convey a different concept:

Facility Altitude. The height or depth of a facility relative to sea level.

This definition includes the concept of "facility," which limits the objects where measurements are appropriate; "sea level," which limits the point of reference for the measurement; and it does not restrict the unit of measure to meters. The last data element described (i.e., Facility Altitude) is not the same data element as was the previous example of Vertical Measure/Altitude.

**Note:** Part 3 of ISO/IEC 11179 includes an attribute for "Unit of Measure" in the value domain of the metadata registry. This is the appropriate attribute to indicate the unit by which the data value is to be recorded. In a standard developed by the American National Standard Institute (ANSI), however, unit of measure was included in the definition, so it has been replicated in this example. The metadata registry model also includes an attribute for the precision required for recording the data value.

### Y.3.2 Establish a Naming Convention

The Registration Authority (RA) should establish a naming convention for each name context in the registry. Where data element names are provided from other sources, the naming convention may not be fully known (e.g., the names assigned to data elements in an application software system). The naming convention shall be constructed according to ISO/IEC 11179-5 naming conventions, as explained in the following paragraphs.

§ **The Scope of the Naming Convention.** The scope of the naming convention determines how broadly the naming convention is applied. For the example registry described in this document, the scope is limited to the Registry name context. For example, a data element might have the name ARegulation Abstract Text@ with the context ARegistry@ and the name AAbstract@ in another context. The conventions used for names in contexts other than for the Registry name context may not be known to the registration authority and the naming convention would be documented as Aunknown.@

§ **The Authority That Establishes Names.** The RA establishes the Registry Names for a registry. The Environmental Data Registry (EDR) has as its RA the Environmental Protection Agency (EPA). The data steward appointed by that agency is the final authority for the assignment of names. Other registries will establish their own RA's.

§ **Semantic Rules for Source and Content of Terms.** Semantic rules enable meaning to be conveyed. Each registry shall specify the guidelines used, if any, that govern the source and content of words used in a name. Name components may come from object class terms, property terms, representation terms, and qualifier terms. These terms may

be part of a thesauri or terminology system. The logical group or entity where a data element might be modeled and the conceptual domain where the data values are defined and maintained can be used as source terms in a data element name. The naming convention for some name contexts might specify that the data element name is simply what the data element is commonly called in the organization, and that no semantic rules are enforced.

- § **Syntactic Rules for Word Order.** Syntactic principles specify the arrangement of components within a name. The specific syntactic rules for a registry, if any, should be specified in the naming convention. In the examples in this document, the convention for syntax for the Registry name context is to include the representation class term as the last term in the name, as in Regulation Abstract *Text*. Representation class terms are defined in Section Y.2.3 of this Annex.
- § **Lexical Rules.** These principles concern preferred and non-preferred terms, synonyms, abbreviations, component length, spelling, permissible character set, case sensitivity, and similar rules. Rules for these subjects, if any, are part of the specifications of the naming convention. A RA might choose to establish controlled, well defined word lists for formulating a name.
- § **Name Uniqueness.** Each registration authority determines whether a name within a context must be unique. Because users often rely on names as an indication of data values, *qualifiers* may be used to distinguish similar data elements within a registry (e.g., *Horizontal* Collection Method Code and *Vertical* Collection Method Code; *Mailing* Address Country Name and *Geographic* Address Country Name).

### Y.3.3 Example of a Naming Convention

An example of a naming convention for the context "Registry Name," and its adaptation for a specific RA is provided in this section. For this example, registry name is considered to be the official name by which a data element is registered in a specific registry.

- § **Scope.** The scope of this example naming convention is for use in the example registry. Each data element must be assigned a "Registry Name". It is not intended to be the official or preferred name for the organization or industry.
- § **Authority.** The authority for this example is the U.S. Environmental Protection Agency for its Environmental Data Registry.
- § **Semantic Rules.** Names shall include a term that indicates the type of values that will be stored in that data element. For example, a data element that represents a domain of Country Identifiers, should have the term ACountry@ in its name. Qualifiers shall be

used to differentiate between names that would otherwise be the same. The representation class term shall always be included as the last term in the name.

§ Lexical Rules. A data element name in the example registry shall have a maximum of 100 alphanumeric characters. The language of the registry shall be English, and the character set ASCII. There are no controlled word lists.

§ Name Uniqueness. Names shall be unique within a registration authority for the context Registry.

### Y.3.4 Formulating a Data Element Name

The examples used in this document are based on a naming convention for name context Aregistry,@ established by one registration authority. The example requires that the data element name be constructed to reflect both the logical entity which includes the data element (i.e., the object) and the attribute which identifies the type of data value to be contained in the data element (i.e., the property). Although the entity is not always required to be a term in the name, the attribute (i.e., type of data value) is a requirement. For the registration authority used in this example, data element name would always include the representation class term, such as name, measure, amount, number, code, quantity, text, or others, as defined in Section Y.2.3.

The data element names in the following Exhibit 5.1 are provided as examples of names to be found in one registry, with the context Registry Name. The table columns identify the name components. Syntactic rules for name are relative. The only rule in this example is for syntax; the representation should be the last component in a name.

Object	Property (Data Values)	Representation	Qualifier	Resultant Data Element Name
Primary Geopolitical Entity	Country Name	Name <sup>1</sup>		Country Name
Address	Country Name	Name <sup>1</sup>	Mailing	Mailing Address Country Name
Address	Country Name	Code	Geographic	Geographic Address Country Code
Address	Person Name	Name <sup>1</sup>	Mailing	Mailing Address Person Name
Facility	Legal Name	Name <sup>1</sup>		Facility Legal Name
Geographic Coordinates <sup>2</sup>	Latitude	Measure		Latitude Measure

Location	Latitude	Measure	Facility	Facility Location Latitude Measure
Location	Latitude	Measure	Stack	Stack Location Latitude Measure
Geographic Coordinates <sup>2</sup>	Collection Method	Code	Horizontal	Horizontal Collection Method Code
Geographic Coordinates <sup>2</sup>	Collection Method	Code	Vertical	Vertical Collection Method Code

<sup>1</sup> "Name Name" is redundant, so only one "Name" is used in the data element name.

<sup>2</sup> "Geographic Coordinates" is an implied entity not included in the data element name.

### **Exhibit Y.1. Data Element Names**

## **Y.4 Identification**

### **Y.4.1 Data Element Identifier and Identifier**

Part 5 of ISO/IEC 11179 gives principles for naming and identification of data elements. Each data element registered within a Registration Authority (RA), i.e., an organization authorized to register metadata, is unambiguously identified with a unique identifier. At the time a data element is registered into a metadata registry, a Data Element Identifier (DI) is assigned to the data element. When a data element is first registered, it is assigned a Version Identifier (VI) of "1". The version number is incremented by "1" for each subsequent change to the data element. The DI and VI can be assigned by the system software when a data element is registered in the registry (i.e., a new data element record is created in the system). Each registration authority should develop business rules for versioning data elements and their attributes.

The combination of RAI, DI, and VI shall constitute the International Registration Data Identifier (IRDI). This identifier provides unique identification to a data element internationally. For the examples listed in Annex F, DI and VI have been recorded to demonstrate uniqueness.

A registration authority might require certain associated administrative information for a data element. Some attributes are specified in the standard (e.g., registration status). Others are determined by the registration authority. Examples of administrative attributes that might be established by a registration authority are described in this section. No administrative data attributes have been assigned to the examples described in the text of this document or in the table provided in Annex F.

### **Y.4.2 Versioning**

Data elements in a metadata registry are generally entered in sets associated with a document, a standard, or an application system. In many cases, a data element may be changed or a data element source like a document, a standard, or an application system may change and a new version may be required.

One approach to tracking changes is to enable database transaction logging which automatically captures the date and time of all changes. The drawbacks to that solution are that all changes, whether significant or insignificant, are logged, and additional processing and space resources are required to retain all versions of a data element. In addition, just logging the date and time of a change doesn't effect a version change. The alternative is to manage version information in database fields that are updated by data analysts who make judgments according to a set of business rules. It is generally understood that the business rules will initially be implemented by analysts.

Following are draft business rules to guide versioning of various objects in a metadata registry.

1. The following objects need to be versioned: group, standard, document, system, data element, value domain.
2. Value meanings and permissible values would not be versioned as part of a value domain, but begin and end dates will document changes to these values.
3. Any version change to a permissible value would result in a new version of a value domain. Begin and end dates would be stored.
4. Any value domain changes would result in the need to review related data elements to determine whether or not they should be versioned. In some cases, a decision by a steward or working group would be required to affirm that a data element would adopt the new version of the value domain.
5. In order to ensure that versioning is effectively applied, it cannot be decided by software, but requires interpretation of the business rules by a data analyst. Versions would be incremented only for non-trivial changes (not typos). In some cases, the data steward and the registrar would need to agree on changes.
6. Data elements would be versioned based on changes to definition or representation or format.
7. Changes to data elements within a group would result in incrementing the version of the group.
8. All changes made to data standards require some documentation of authorization. This could be indicated within a text field for each standard.

9. Typographical changes (errata) would require a notification process. More substantive changes may require balloting or a consensus process to approve the changes. This approval could be recorded as a new document in the registry, and could be cited as the source for the new versions of the data elements.

New data element versions would be indicated by incrementing the version number associated with the identifier. This is a new physical record for the data element, and the registry would continue to store the earlier versions (i.e., both 6125:1 and 6125:2).

## **Y.5 Conceptual Relationships**

Data element concepts, conceptual domains, and value meanings are described in this section.

### **Y.5.1 Data Element Concept**

The data element concept is readily derived, based on the name and definition of the data element. It is a concept that can be represented in the form of a data element, described independently of any particular representation. The data element "Country Name" is a representation of the data element concept "Country Identifier."

The following list is provided as guidance for terms that might be used in names and definitions of data element concepts. Terms that do not denote representation include the following:

- § Identifier. Something that represents to be, regards, or treats as the same or identical.
- § Label. A short word or phrase descriptive of a person, group, or intellectual movement, or indicating that what follows belongs in a particular category or classification.
- § Tag. A descriptive word or phrase applied to a person, group, organization, etc., as a label or means of identification or epithet.
- § Indicator. Anything that serves to point out or direct attention to, as of a measuring device.
- § Discriminator. A distinction that differentiates one from another.

The data element concept is the concept for which the conceptual domain contains representative values. The following list of characteristics is provided as guidance to ensure consistency in formulating the names and definitions of data element concepts:

- § Singular. Each data element concept represents only one concept.

- § Does not include representation. It does not incorporate the representation terms such as name, code, text, number, or other terms that denote how the concept can be represented in either the name or the definition of the concept.
- ? Indefinite article. The definition is stated with the indefinite articles "a" or "an" since the concept does not specify a particular data value or representation.
- ? Can be associated with multiple data elements, each with its own representation and value domain.

ISO 3166, for example, represents the data element concept "Country Identifier," which can be represented as names, or it can be represented by codes (e.g., "Country Name" or "Country Code"). There are more than one name and more than one code associated with the concept for "Country Identifier." Each name and each code requires its own data element and value domain.

- ? Can be associated with only one conceptual domain.

The appropriate level for exchanging data values is the conceptual level, through data element concept and conceptual domain. The value domains of country codes and country names are translatable, where the value meanings associated with the conceptual domain reference the same data element concept for countries of the world.

A data element concept identifier can be created by the system software, to provide unique identification and versioning for data element concepts, and an identifier that can be used to indicate the domain for translation of data values.

## **Y.5.2 Conceptual Domain**

A conceptual domain is a perception template of understanding that might be an enumerated set of meanings. A data element concept uses a conceptual domain to constrain its perception meaning. An enumerated conceptual domain is a set of all possible, valid value meanings of a data element concept expressed without representation. The conceptual domain for the "Country Identifier" data element concept is the collection of all the value meanings that can be used to identify all of the countries of the world.

Characteristics of conceptual domains include:

- ? Plural. Whether enumerated or non-enumerated, a conceptual domain includes the entire body of information that might be included as meanings of the data values in a particular

data element for a particular concept. Therefore, the name and definition are always described as plural.

- ? Object oriented. The name is used to identify the component contained in the conceptual domain. It does not require a property identifier or an object class. For example, "Countries of the World" includes the identification of all countries.

- ? Lacking representation. The definition identifies the type of information that a conceptual domain encompasses, without using representation class terms such as code, name, text, number, picture, measure, quantity, and identifier. For example: "Countries of the World" is defined as "The primary geopolitical entities of the world," not as "The *names of* the primary geopolitical entities of the world."
- ? Conceptual domains can be, and often are, associated with more than one data element concept. Data element concepts that "Countries of the World" could be associated with include, but are not limited to:
- Address Country Identifier.
  - North American Country Identifier.
  - NATO Country Identifier.
  - Geographic Country Identifier.

A conceptual domain can be associated with any data element concept that uses the same value meanings (e.g., United States, Canada, and Mexico are value meaning names for both the Address Country Identifier and the North American Country Identifier concepts). Different value meanings require a different conceptual domain. For example, in a database about countries, a data element that contains information about a country other than country identification (e.g., size, type of government, economic activities) would have its own conceptual domain.

A rule for determining if a data element concept can be associated with a conceptual domain is to consider the value meanings associated with the conceptual domain. Names such as Frigid, Tropical, or Temperate could be permissible values for a conceptual domain about geographic zones where countries are located, but they cannot be defined as "The principal geopolitical division of the world known as <country name>." They would not be associated with the conceptual domain "Countries of the World."

Where the content of the value meanings is the same for more than one data element/data element concept/value domain, the conceptual domain can be reused for multiple data element concepts as described previously in this section. Conceptual domain identifiers have been recorded for the examples provided in Annex F to demonstrate uniqueness and reusability.

### **Y.5.3 Value Meanings**

Every enumerated conceptual domain is associated with more than one value meaning. A value meaning is the meaning (description) of a permissible value that will be stored in a data element. Value meanings can have both name and definition. Often the "name" of a value meaning becomes the permissible value of that value meaning in a data element with "name" representation. Characteristics of value meaning names and definitions are:

- § Cannot be a representation. The name and definition do not contain representation class terms such as name, number, text, code, or other representation terms.
- § Must be associated with at least one conceptual domain.
- § Can be associated with more than one conceptual domain.

Example 1: Value meaning names associated with the conceptual domain "States of the United States" is also associated with the conceptual domain "Data Collection Sources" in one data registry.

Example 2: The value meaning name "Unknown," indicating that the data value for a particular data element is not known, can be associated with many conceptual domains.

- § Begin and End Dates. The dates when a value meaning was entered into a conceptual domain and when a value meaning was no longer valid for a conceptual domain are required in a data registry.
- § Unique Identifier. Each value meaning has a unique identifier (VMID) in a registry. The VMID and the data element unique identifier (IRDI) provide unique identification of a particular data element item occurrence. This combination of identifiers is valuable for data transfer.

In addition, the value meaning should be singular. Each value meaning represents one instance of the meaning of a value to be found in a data element.

## Y.6 Classification

Classification helps to add information not easily included in definitions, helps to organize the contents of a metadata registry, and helps to provide access by supporting more meaningful queries. Part 2 of ISO/IEC 11179 describes general categories of classification; Part 5 describes three classified components: object class, property, and representation class. An object class term represents an activity or object in a context. Property terms are terms that modify an object term. Representation class terms describe the form of representation. Representation terms are described in Annex Section Y.2.3.

A metadata registry might choose to classify data elements as groups, e.g., the group of data elements used in a mailing address, the group of data elements used to identify chemical substances, or the group of data elements that locate a point on the surface of the earth.

Keywords might also be used to classify data elements, e.g., altitude, date, facility, industrial, and organization.

## **Y.7 Quality Review**

As metadata for data elements are completed, the data element progresses through a review process to standardization, where appropriate. The Registration and Administrative Statuses indicate the status of a data element in the registration/standardization process.

### **Y.7.1 Registration Status**

The standard values for registration status include the following:

- § Incomplete. The data element does NOT have all the necessary metadata.
- § Recorded. The data element has all the necessary metadata, but has NOT met all the quality requirements.
- § Certified. The data element has all the necessary metadata and has met all quality requirements.
- § Standard. The data element has all necessary metadata, has met all quality requirements, and has been approved by the Registration Authority.
- § Retired. The data element is no longer used in the registry.

The registration authority might also choose to use Legacy as a registration status:

- § Legacy. The data element was obtained from a Legacy System and may be missing some metadata. It has not been considered for standardization.

The registration status for a new data element is always listed as "Incomplete" until such time as all attributes associated with that data element are completed. After all of the data element attributes have been verified to be complete, the registration status is changed to "Recorded." Other status changes are determined by the registration authority.