

# ISO/IEC JTC 1/SC 32 N 0489

Date: 2000-06-16

REPLACES: --

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

<b>DOCUMENT TYPE</b>	Text for CD ballot or comment
<b>TITLE</b>	ISO/IEC PDTR 20943 Information technology - Procedures for achieving metadata registry (MDR) content consistency - Data elements
<b>SOURCE</b>	Judith Newton (Project Editor)
<b>PROJECT NUMBER</b>	1.32.16.01.00.00
<b>STATUS</b>	This PDTR has been prepared as specified in the New York, NY, USA meeting resolutions in May 2000
<b>REFERENCES</b>	
<b>ACTION ID.</b>	LB
<b>REQUESTED ACTION</b>	National Bodies are requested to use the comments template that is used by SC 32/WG 3 in returning their comments. This template is also on the SC 32/WG 2 ftp site.
<b>DUE DATE</b>	2000-09-29
<b>Number of Pages</b>	75
<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)

available from the JTC 1/SC 32 WebSite <http://www.jtc1sc32.org/>

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

# ISO/IEC JTC 1/SC 32 N 0489

Date: 2000-06-16

REPLACES: --

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

<b>DOCUMENT TYPE</b>	Text for CD ballot or comment
<b>TITLE</b>	ISO/IEC PDTR 20943 Information technology - Procedures for achieving metadata registry (MDR) content consistency - Data elements
<b>SOURCE</b>	Judith Newton (Project Editor)
<b>PROJECT NUMBER</b>	1.32.16.01.00.00
<b>STATUS</b>	This PDTR has been prepared as specified in the New York, NY, USA meeting resolutions in May 2000
<b>REFERENCES</b>	
<b>ACTION ID.</b>	LB
<b>REQUESTED ACTION</b>	National Bodies are requested to use the comments template that is used by SC 32/WG 3 in returning their comments. This template is also on the SC 32/WG 2 ftp site.
<b>DUE DATE</b>	2000-09-29
<b>Number of Pages</b>	74
<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)

available from the JTC 1/SC 32 WebSite <http://www.jtc1sc32.org/>

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

**ISO/IEC JTC1/SC 32 N**

Date: 2000-06-16

**ISO/IEC PDTR 20943**

ISO/IEC JTC1/SC 32/WG 2

Secretariat:

## **Information technology — Data management and interchange — Procedures for achieving metadata registry (MDR) content consistency - Data elements**

*Élément introductif — Élément central — Élément complémentaire*

### **Warning**

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this document are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

### Copyright notice

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:

*[Indicate :  
the full address  
telephone number  
fax number  
telex number  
and electronic mail address*

*as appropriate, of the Copyright Manager of the ISO member body responsible for the secretariat of the TC or SC within the framework of which the draft has been prepared]*

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

## Contents

Foreword .....	v
Introduction .....	v
<b>1 Scope .....</b>	<b>1</b>
1.1 Background .....	1
1.2 Purpose .....	1
1.3 Scope .....	1
<b>2 References .....</b>	<b>2</b>
<b>3 Terms and definitions .....</b>	<b>2</b>
<b>4 Data element abstraction .....</b>	<b>2</b>
4.1 Abstraction types .....	3
4.2 Example of specialization/generalization .....	3
4.3 Example of decomposition/aggregation .....	4
4.4 Layer relationships .....	5
<b>5 Bottom-up approach to data element registration .....</b>	<b>5</b>
5.1 General procedures for registering a data element .....	6
5.1.1 Understanding the data element .....	6
5.1.2 Content research .....	7
5.1.3 Definition and permissible values .....	7
5.1.4 Names and identifiers .....	8
5.1.5 Other metadata attributes .....	8
5.1.6 Data element concept .....	9
5.1.7 Classification schemes .....	10
5.1.8 Quality control .....	10
5.2 International standard with enumerated domain .....	10
5.2.1 Understanding the data element .....	11
5.2.2 Content research .....	11
5.2.3 Definition and permissible values .....	11
5.2.4 Identification and naming the data element .....	12
5.2.5 Other metadata attributes .....	13
5.2.6 Data element concept .....	14
5.2.7 Classification .....	14
5.2.8 Quality control .....	15
5.2.9 Other names and codes from ISO 3166 .....	15
5.2.10 Summary of Attributes .....	15
5.3 International standard with non-enumerated domain .....	17
5.3.1 Understanding the data element .....	18
5.3.2 Content research .....	18
5.3.3 Definition and permissible values .....	18
5.3.4 Identifying and naming the data element .....	20
5.3.5 Other metadata attributes .....	20
5.3.6 Data element concept .....	21
5.3.7 Classification .....	21
5.3.8 Quality control .....	22
5.3.9 Other data elements in ISO 6709 .....	22
5.3.10 Summary of metadata attributes .....	22
5.4 Application data element .....	24
5.4.1 Understanding the data element .....	25
5.4.2 Content research .....	25
5.4.3 Definition and permissible values .....	25

5.4.4	Identify and name the data element .....	26
5.4.5	Other metadata attributes .....	27
5.4.6	Data element concept .....	27
5.4.7	Classification .....	28
5.4.8	Quality control .....	28
5.4.9	Related data elements .....	28
5.4.10	Summary of metadata attributes .....	28
5.5	Classification of data elements .....	30
5.5.1	General procedures for registering a classification of data elements .....	31
5.5.2	Data elements in a document .....	32
5.5.3	Data elements in a standard .....	33
5.5.4	Composite data element.....	34
5.6	Linking of data elements .....	35
6	Top-down approach to data element registration .....	36
6.1	Biological organisms .....	36
6.1.1	Data element concepts .....	36
6.1.2	Data elements .....	36
6.1.3	Permissible values .....	37
6.2	Biological organism types .....	37
6.2.1	Data element concepts .....	37
6.2.2	Data elements .....	37
6.2.3	Permissible values .....	38
6.3	Top down registration .....	39
7	Complex data .....	40
	Annexe A Table of data element attributes for examples .....	42
	Annexe B Application of ISO/IEC 11179, Parts 4-6 .....	46
	Annexe C Crosswalk of data element component names in technical report to ISO/IEC 11179-3 metamodel, revision 2000-05-34 .....	64
	Bibliography .....	67

## Foreword

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

In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The main task of technical committees is to prepare International Standards, but in exceptional circumstances a technical committee may propose the publication of a Technical Report of one of the following types:

- type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;
- type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;
- type 3, when a technical committee has collected data of different kind from that which is normally published as an International Standard ("state of the art", for example).

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

Technical Reports are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3. ISO/IEC TR 20943, which is a Technical Report of type 3, was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 32, *Data management and interchange*.

## 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 comparable 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.

Metadata registries can be used to register data elements, value domains, and associated attributes for many kinds of organizational data resource collections. Metadata registries can store information on data elements used on forms, represented in enterprise data models, contained in EDI message sets, described in documents and standards, as well as those data elements that are part of computer system applications. Some organizations use the registry to record essential facts about how data elements are used in existing applications, while other organizations use the registry as a repository of standard data elements to be used as models for data elements in application development. Part 6 of ISO/IEC 11179 specifically addresses the development and population of metadata registries.

The proposed revision of ISO/IEC 11179, Part 3, models a data element (DE) and its associated data element concept (DEC). Conceptualization and articulation of rules and relationships in the creation of data element concepts, data elements, and value domains 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 generalized data element to the most specialized application data element needs to be specified. Reuse of data value domains should be enabled and regularized.

While metadata registries can be used for storing information about a variety of metadata entities, this report addresses only data elements and their associated components. The goal of this paper is to ensure that there is a common understanding of the content of the data element attributes so that metadata can be shared between registries, despite their differences.

# Information technology — Data management and interchange — Procedures for achieving metadata registry (MDR) content consistency - Data elements

## 1 Scope

### 1.1 Background

An ISO/IEC 11179 metadata registry (MDR) (hereafter referred to as a "registry") is a tool for the management of shareable data; a comprehensive, authoritative source of reference information about 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 revision of Part 3 of the six-part International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 11179 standard that describes the organization of a registry for managing the semantics of data elements in data systems, and incorporates content from it. The standard specifies the structure of a registry in the form of a conceptual model. The conceptual model is more abstract than a logical data model in that it describes how the human mind thinks about information. It is not intended to be a logical data model for a computer system, much less a physical model.

A registry does not contain data itself. It 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 units of data. The standard identifies 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.

### 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 technical report is not a data entry manual, but a user's guide for conceptualizing a data element and its components for the purpose of consistently establishing good quality data elements.

### 1.3 Scope

The scope of this technical report is limited to the components of a data element: the data element identifier, names and definitions in particular contexts, and examples; data element concept; conceptual domain with its value meanings; and value domain with its permissible values. This technical report also addresses registration of classification schemes and their use in classifying data elements.

Two approaches to data element registration are included in this report: a bottom-up approach, where limited information about the data element is available prior to registration, and a top-down approach, where conceptual information has previously been identified.

The bottom-up registration procedure provides for the basic metadata attributes about the data element (e.g., definition, name, and permissible values) to be completed prior to defining the conceptual information about the data element. This approach is described in Clause 5 of the technical report.

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 Clause 6 with an example of registration of data elements about biological organisms.

## 2 References

Standards from which examples have been drawn to be used in this document are listed in the Bibliography.

ISO/IEC 11179-1:1999, *Information technology - Specification and standardization of data elements - Part 1: Framework for the specification and standardization of data elements.*

ISO/IEC 11179-2:1999, *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 11179-6:1996, *Information technology - Specification and standardization of data elements - Part 6: Registration of data elements.*

ISO/IEC TR 15452:2000, *Information Technology - Specification of Data Value Domains.*

## 3 Terms and definitions

For the purposes of this technical report, the terms and definitions given in ISO/IEC 11179 and ISO/IEC TR 15452 apply.

## 4 Data element abstraction

This clause presents a conceptual framework for structuring data elements and data element contents 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 may progress from the most general (conceptual) to the most specific (ultimately, the physical layer, although a registry might or might not contain these).

Depending on the type of abstraction, relationships among the members of each layer and between layers are meaningful in terms of defining the structure of the registry contents. This provides a means of comparison of the contents of different registries. In addition to the data element definition and other attributes, comparing the type and level of abstraction by which the data element was derived can ensure that content can be shared among registries.

One could use layers to structure development of a system, for instance, with the highest layers of definition contained in a business view, and development progressing to the implemented system layer. The number and granularity of layers are driven by user requirements. This clause will describe two ways to derive layers, neither of which are intended to be mandatory for any particular implementation, and will present examples of the types of abstraction most useful to registry implementations.

## 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 focusing 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 data elements 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 layer. Although this description of the process of abstraction development implies that it always proceeds in one direction (from the general to the concrete), there may certainly be cases where the reverse is true. Also, after the structures are in place, there is no implication that the relationships are all in single directions. Indeed, the names of the abstraction types deliberately include two levels in a effort to avoid that assumption.

The two types of abstraction of most interest to data element development are *specialization/generalization* and *decomposition/aggregation*.

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

The first type of abstraction is most useful for structuring registry members, as most things described in a registry will be classes by definition.

## 4.2 Example of specialization/generalization

The mechanism of specialization/generalization can be used to express a hierarchical structural relationship among data elements.

The United States Postal Service (USPS) establishes postal codes for the United States. The data element State USPS Code forms part of an address group (figure 1). It might be used to map to application data elements directly, but it could also be the highest layer of a hierarchy of data elements, each displaying a greater degree of specialization according to its position in the structure.

The next layer down contains the elements Geographic Address State Code and Mailing Address State Code. The two State Code elements are now members of two differing address groups. The conceptual domains of the two elements are still identical, however, as the set of potentially valid values of geographic addresses and mailing addresses are the same.

Further specialization takes place at the next layer. Mailing Address State Code is partitioned into Facility Mailing Address State Code and Customer Mailing Address State Code. As before, the address groups of these elements differ, and in addition the data value domains do not contain the same sets of valid values. The domain of Customer Mailing Address State Code remains the same as Mailing Address State Code, but the domain of Facility Mailing Address State Code is now restricted to the subset of states in which facilities are located.

In this example, application data elements are mapped to the registry elements at this layer. Clause 5.5 and 5.6 describe linking data elements that occur in different databases, under different names, to one registry data element. These relationships encourage data interchange by assuring semantic equivalency among disparately appearing data elements.

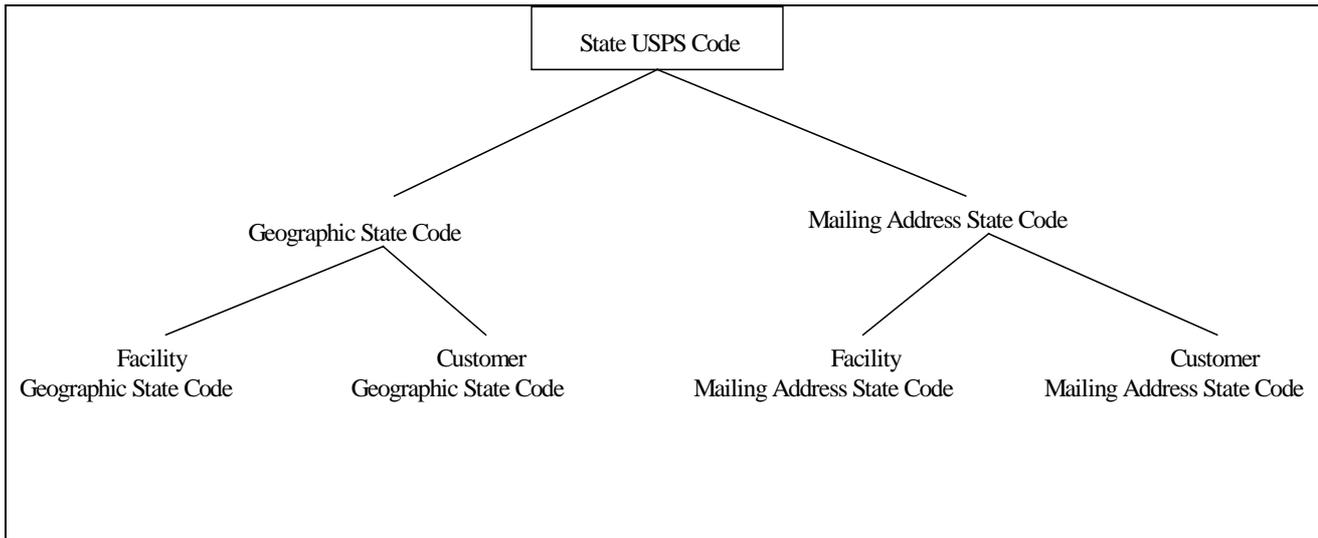


Figure 1 — Specialization/generalization hierarchy.

Clause 5 and Annexe A show a way to track the layers of specialization by assigning an attribute called Layer of Abstraction Type to the data element description. This attribute shows the type of abstraction of the data element in the abstraction structure. In the example above, the type of the data element State USPS Code is given the Type: Generalization. This designation indicates that this data element is at the top of the specialization/generalization structure for data elements in this registry. The data elements in the next layer, Geographic State Code and Mailing Address State Code, are assigned the designation Specialization. The data elements in the third layer, namely Facility Geographic State Code, Customer Geographic State Code, Facility Mailing Address State Code, and Customer Mailing Address State Code, are also assigned Specialization.

There may be other structures in the registry in which data elements are placed relative to other registry components, such as data element concepts. In that case, other attributes may be needed to locate the example data elements in those structures.

### 4.3 Example of decomposition/aggregation

Use decomposition/aggregation to express a relationship among data elements in which the higher layer describes a whole and the lower layers are component parts of that whole. They may or may not be discrete; that is, some of the lower layer components may overlap in their descriptions or domains.

An example of this type of abstraction is the dividing of mailing addresses into their subcomponents (figure 2). These divisions are based on ISO 11180, Postal Addressing. The figure shows some of the subdivisions for the data element group Mailing Address Group.

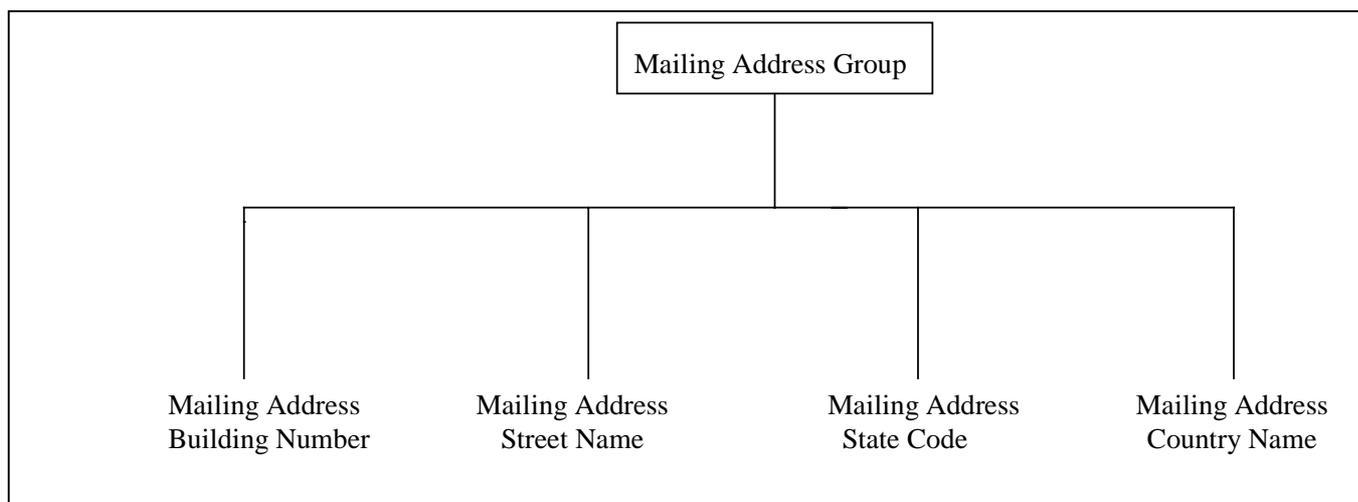


Figure 2 — Decomposition/aggregation example.

Clause 5 and Annexe A show a way to track the layers of aggregation by assigning an attribute called Layer of Abstraction Type to the data element description. This attribute shows both the type of abstraction of the data element in the abstraction structure. In the example above, the type and layer location of the Mailing Address Group is given the Type: Aggregation. This designation indicates that this data element group is at the top of the decomposition/aggregation structure for data elements in this registry. The data elements in the next layer of decomposition are assigned the designation Decomposition.

There may be other structures in the registry in which data elements are placed relative to other registry components, such as data element concepts. In that case, other attributes may be needed to locate the example data elements in those structures.

#### 4.4 Layer relationships

Throughout Clause 5, reference will be made to the layer of abstraction for each example data element. The relationship of elements in any particular layer to those above and below it will help the user to choose which data element contains the appropriate information for a particular use.

### 5 Bottom-up approach to data element registration

Registration of a data element in a registry requires that certain characteristics of the data element be 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 a range of data elements, from those found in 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 technical report 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 registry. This technical report gives examples that demonstrate the population of a registry. It includes attributes that are mandatory and fully defined by the metamodel, as well as those where the Registration Authority (RA) must establish its own profile of required attributes.

In many cases where a data element is submitted for registration, only a limited amount of information (e.g., a name, definition, and a set of permissible values) is provided by the submitter. Other attributes must be determined based on an understanding of the underlying data values and concepts that are implied by those facts. Many registry practitioners find that using a bottom-up approach to registering a data element is most appropriate. A bottom-up approach might also be used where the 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 clause of the report describe how to formulate attributes about a data element, based on a bottom-up procedure. The registration procedures are presented in a logical order for analyzing and formulating attributes for a data element. 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:

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

This report is intended to be used to help registry practitioners to formulate the attributes that describe and define a data element. Subclause 5.1 presents an example of an overall approach to data element registration. Subclauses 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. Annexe A contains a table that concisely summarizes the information registered for each data element in the examples that follow in Clause 5. Annexe B, which is based on ISO/IEC 11179, parts 4, 5, and 6, contains more detailed information and examples to assist the practitioner who is registering data elements. Annexe C provides a crosswalk from the example attributes named in this technical report to the ISO/IEC 11179 part 3 metamodel (Proposal for Comment, 2000-05-23).

The metamodel provides for the recording of administrative data about many components of the registry. These components are considered to be Administered Components. Administered components must contain an identifier, the date the component was created, and the registration and administrative status of the component. Administered components must have a name and a definition in context. An administered component may also contain dates when data about the component were changed, an effective date, and an until date. Explanatory notes and the origin of the component, as well as administrative notes, change description, and unresolved issues may also be recorded. The exact process for determining registration status of administered components other than data element itself, has not been established in the current ISO/IEC 11179-6, where registration is described. For the purpose of this technical report, the characteristics of completeness and data quality that are to be used to determine registration status for all components is considered to be the same as for data elements themselves. In Subclause 5.1 administered components are identified together with the description of the population of those components.

### 5.1 General procedures for registering a data element

Often only a limited amount of information is available about a data element (DE) 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 the following:

- Is the data element described in an existing International, National, or organizational standard?
- Does a data element exist in the registry, or a federation of registries, that has the potential for being reused?

It is necessary to recognize that the registration practitioner must make value decisions when recording metadata into the registry. The practitioner will determine if a data element might be adapted to meet new requirements, or if 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 components 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 or some of its component attributes.

### 5.1.3 Definition and permissible values

#### 5.1.3.1 Data element definition

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. Examples are provided in Annexe B for formulating definitions, based on the rules and guidelines set forth in ISO/IEC 11179-4. Definitions and names (Subclause 5.1.4) are entered in context (i.e., a designation of the discipline in which a name and definition are applied or from which it originates). For many of the examples used in this report, the context is recorded as "Registry." "Registry" in this example indicates that the name and definition have been formulated according to the rules and guidelines established by the registration authority for this particular registry. In addition to designating context, the language represented by a name and definition can also be designated as a Language identifier. For this report, all examples have been provided in English.

#### 5.1.3.2 Permissible values

Part 3 of ISO/IEC 11179 identifies the attributes that describe the domain of permissible values. The permissible values for a data element are the value domain for the data element. Annexe B also contains detailed information about the attributes in value domains and examples of how those attributes are used for both enumerated domains (i.e., established through a list) and non-enumerated domains (e.g., specified through a formula, rule, procedure, or reference).

Different attributes are used depending upon whether the permissible values are enumerated or non-enumerated. Each enumerated permissible value is associated with a valid value meaning that provides meaning to the permissible value, as described in Subclause 5.1.6. Each enumerated permissible value is also entered in the registry with its begin date (i.e., the date when that permissible value became valid for a value meaning in that registry). 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 description of the values that are valid for the data element. This report contains specific examples of registering data elements with enumerated domains (Subclauses 5.2 and 5.4) and with a non-enumerated domain (Subclause 5.3).

At this time in the registration process, it is appropriate to enter the representation class that describes how the data element is represented. A representation class qualifier can be recorded where appropriate. Other value domain attributes are also entered, including the name of the value domain, the datatype by which the data values are represented, field length, format, and the unit of measure and precision where appropriate.

The value domain may, optionally, be an administered component. Where it is administered, administrative data may be recorded for the value domain, including name and definition in context, the origin of the domain, and any explanatory comments, as well as the administrative and registration status of the value domain.

### 5.1.4 Names 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) is unambiguously identified with a unique identifier. Although the standard does not specify the format or content of the data element identifier (DI), the identifier 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 registration authority establishes its own identification scheme, the same identifier might be used to identify a different data element in another registry or even in the same registry for a different registration authority. Therefore, a Registration Authority Identifier (RAI) must be established and used with the data identifier and version identifier 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 an RA and the construction of an RAI. The RAI is composed of an International Code Designator (ICD), an Organization Identifier (OI), an Organization Part Identifier (OPI), and an OPI Source (SRC). The IRDI is discussed further in Part 6 of the standard. For the purpose of this document, the example RAI is constructed, using acronyms for the component parts, rather than the numeric representations that would comprise an official RAI. Therefore, the RAI throughout this example is represented as USA-EPA-OEI-OIC (indicating United States of America-Environmental Protection Agency-Office of Environmental Information-Office of Information Collection).

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. The description of the naming convention is stored in the metadata registry. ISO/IEC 11179, part 5 does not specify a mandatory naming convention, but does explain how to document one. Annex B 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 Subclause 5.1.3 and the identifying attributes described in Subclause 5.1.4, there are administrative, relational, classifying, and other miscellaneous attributes that serve to define and describe a data element. The standard for an ISO/IEC 11179 metadata registry requires that a data element be an administered component.

In addition to the mandatory attributes specified by Part 3 of the standard, a registration authority might establish a profile for a particular registry. The profile specifies such requirements as:

- Where attributes described as optional in the standard are mandatory for that registry.
- Optional attributes in the standard are not included in the registry.
- Attributes not included in the standard might be identified and added to the registry.

The attributes that relate data elements through data element concepts (Subclause 5.1.6) and those that classify data elements (Subclause 5.1.7) are described in subsequent clauses of this report. Many information sources do

not provide concept or classification information about the data element. Some administrative information is related to quality control, and is described in Subclause 5.1.8. Annexe B includes detailed information about these metadata attributes.

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

- **Submitting Organization:** The submitting organization is the name of the organization that has submitted the data element for addition to the registry. The submission contact for the submitting organization is also recorded in the registry.
- **Stewardship Contact:** The organization that has been delegated the responsibility for managing a set of data resources is the stewardship contact. The stewardship contact may be a position within the organization, or even a hot line answering service. Additional information about the contact, including telephone number, email address, and mail code can be entered at this time.
- **DE Comment:** A data element may have an explanatory comment that can be used to provide remarks about the data element, including usage, procedure, and other explanatory information that is not appropriate to include in the data element definition attribute.
- **DE Origin.** A data element can be associated with the name of any kind of source, including a document, standard, system, group, form, or message set. Only one source is associated with a data element to indicate the origin of information about the data element. Additional information can be entered about the source at this time, including abstract and bibliographic information.

The System Date is generally captured by the application software for the administrative attributes Creation Date and Last Change Date. Other administrative data may be entered manually.

### 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 may relate several data elements that record data about that concept with different representations, e.g., both names and codes that represent provinces of Canada and share the same concept, which in this example 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 always administered components and must be specified with administrative data, as is the data element itself. Therefore, at least one name and definition must be recorded in context for the data element concept. The date when the data element concept is entered into the registry and both registration status and administrative status are required. Comment, origin, and other administrative attributes might also be stored about a data element concept.

Data element concepts may be identified with object class, property, and qualifiers. Data element concepts are specified through a conceptual domain, i.e., the meanings of the set of possible valid values for a data element concept, 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, its begin date (i.e., the date when that value meaning became effective in the conceptual domain for that registry) and end date (i.e., the date when the value meaning became invalid). Permissible values are associated with value meanings, according to the representation defined by the value domain. For example, the permissible value "CA", the 2-character alphabetic code for Canada, would be associated with the value meaning "The primary geopolitical entity known as Canada" in a Conceptual Domain entitled, "Countries of the World."

Derivation of data element concepts and conceptual domains, including value meanings are described in detail in Annexe B.

### 5.1.7 Classification schemes

The classification scheme attributes are recorded, where appropriate, at this time. Classification helps to add information to data elements that is not easily included in definitions, helps to organize the contents of a registry, and helps to provide access by supporting more meaningful queries. Part 2 of ISO/IEC 11179 describes general categories of classification.

Each registration authority should determine the types of classification to be included in a registry. The classification scheme type name is always recorded. The registration authority might choose to classify data elements as groups, such as the following:

- Data elements used in a mailing address.
- Data elements used to identify facilities.
- Data elements that locate a point on the surface of the earth.
- Data elements that are listed in a data standard.
- Data elements included in an application system.
- Data elements identified on a data collection form.

Examples of other types of classification that the registration authority might choose, are to classify data elements with one or more keywords, e.g., altitude, date, facility, industrial, and organization, and by object class. Where appropriate, classification schemes may contain administrative data. In this technical report, only examples of classification scheme types and examples of classification values for those types are provided.

### 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 attributes for a data element have been entered into the registry 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 assigned the status of "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 registry, but only a relatively small number of them might be assigned a "standard" registration status. Annexe B describes the assignment of registration and administrative status throughout the life cycle of a registered data element. ISO/IEC 11179 Part 6 specifies the layers of registration status; the administrative statuses, however, are established by each registration authority. Although Part 6 of ISO/IEC 11179 does not provide for the registration of components other than data element, this technical report has followed the same process for all administered components as has been established for data element.

## 5.2 International standard with enumerated domain

This clause 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 (both English and French) correspond to those given 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*. The names are those published in the *Standard Country or Area Codes for Statistical Use* established by the United Nations Statistics Division. The full name of a country is the formal title as identified 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 by agreement of 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 value 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 English-language short country name as the example. The table in Subclause 5.2.10 contains all of the metadata attributes recorded for this example of an 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 English-language short 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. For example, "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 in preference to the full name, to avoid, wherever possible, any reflection of a country's political status (e.g., "People's Republic of ..." or "Kingdom of...").

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 Latin 1 character set.

### 5.2.2 Content research

Other standards that contain conceptual domains for country identification include the 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 was developed for use by the U.S. Department of State and was maintained by the Office of the Geographer and Global Issues, U.S. Department of State. 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 entities, names, and codes prescribed by ISO 3166, with some qualifications. For example, FIPS 104-1 does not identify freely associated territories and possessions of the United States as countries, as they are in ISO 3166. FIPS 104-1 was last updated on May 12, 1986. The maintenance organization for FIPS 104-1 has been 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.

A standard for data elements that identify countries must be an authoritative international source of value domains with ongoing maintenance. 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 identifying a data element.

### 5.2.3.1 Definition

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

### 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 a registry with the date when that permissible value was valid for that value meaning in that particular registry. In this example the "begin date" is arbitrarily listed as January 10, 1997, the date when the standard became effective. 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 a set of permissible values. It can be labeled with a name and a definition, in context. For this example, the context is considered to be "Registry"; the name of the value domain is "Short English-language Country Names"; and the definition is "All short, ISO-recognized English-language names of all countries." Value domains must also have an identifier and a version number, which can be assigned through application software. Refer to Annexe B.2.2, Value Domain, and record the other value domain attributes for this example at this time, including:

- Datatype: The datatype for country name is "alphanumeric."
- Maximum and minimum characters: Based on prior research (Subclause 5.2.1), the minimum length for data element values in the standard is known to be four, and the known maximum length is 44.
- Format: The format selected by the registration authority for this example is A(44) to accommodate the longest of the English-language short names.
- VD Origin: The standard from which the permissible values are obtained is ISO 3166-1:1997, *Codes for the representation of names of countries and their subdivisions - Part 1: Country codes*.
- Domain type: Although domain type is not an attribute in the ISO/IEC 11179-3 model, some metadata registries have created an attribute for domain type as an extension of the model. Recording of the domain type (enumerated in this example) has value for differentiating between enumerated and non-enumerated domains, rather than differentiating by inference.
- Representation Class: Name
- Representation Class Qualifier: Short
- Identifier: USA-EPA-OEI-OIC 5678:1

### 5.2.4 Identification and naming the data element

Names 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 by which users can 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 Annexe B for the identification of data elements. In the registry for this example, a unique data element identifier and version identifier (DI:VI = 20903:1) are assigned at the time of registry. The identifier is preceded by the identifier for registration authority, as described in Annexe B of this document. For this example, the registration authority identifier (RAI) is represented as 'USA-EPA-OEI-OIC'. EPA is used in this example to represent a specific

registration authority (i.e., Environmental Protection Agency). An automated process usually assigns identification numbers when a record is created, including the RAI for which it has been recorded.

#### 5.2.4.2 Name context and naming convention

ISO/IEC 11179 Part 5 describes the naming of data elements. Annexe B 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 Annexe B 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. EPA 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 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 is Latin 1.
- 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 for this data element 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, according to this registry's naming convention, 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 property "English-language" and qualifier "short" are appropriate to this example. The name that has been formulated for this data element, therefore, is "ISO English-Language Country Short Name."

#### 5.2.5 Other metadata attributes

Examples of other metadata attributes that can be recorded at this time are described in the following paragraphs:

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

DE Example: China

- Identify the origin for this data element as the standard from which the permissible values are obtained. For this example, the origin of the data element is the same as the origin of the value domain.

DE Origin: ISO 3166-1:1997, *Codes for the representation of names of countries and their subdivisions - Part 1: Country codes*

- Record any explanatory comments not included in the definition that might provide additional information about the data element.

DE Explanatory Comment: This data element is included in the mailing address standard.

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

Submitting Organization: Office of Information Collection

- Record the name or title of the organization or other contact assigned the responsibility for monitoring and maintaining the data element as the stewardship contact.

Stewardship Contact: Geopolitical Data Administrator

Administrative metadata, such as Creation Date is recorded or captured automatically by the system where applicable. Other administrative data may be entered manually.

### 5.2.6 Data element concept

Identification of the data element concept, as described in Subclause 5.1.6 is based on the data element name and definition, without the representation. The concept represented in 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 name and definition of a data element concept are recorded in context, in the same manner as the name and definition of a data element. The context of the data element concept name and definition in this example is "Registry."

The object class and property of the data element concept can be entered at this time, with qualifiers if appropriate. For Country Identifier, the object class is recorded as "Country". No property or qualifiers are recorded for this example data element concept. Data element concept is an administered component. Administrative data are required for data element concept, including the name and definition, as described above, and the date when the data element concept was created in the registry. The creation date is generally assigned by application software, captured from the system date.

The conceptual domain is a collection of value meanings that describe the meanings of 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 became valid for the conceptual domain in its particular registry. In this example the date is set arbitrarily to January 10, 1997, the date when the standard became official. End dates will also be entered, when the value meaning becomes invalid within the registry (e.g., when a registry is updated to reflect a change to the usage of that value meaning description).

### 5.2.7 Classification

The following paragraphs provide examples of types of classification by which this example data element might be classified:

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

Concept group: Country Identifiers, where other data elements in the group might include the other six representations for "country identifiers."

Subject group: Geopolitical Entities, where other data elements in the group might include such data elements as state or province name, county name, and city name.

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

Object Class: Country

- Identify the abstraction layer of this data element:

Layer of Abstraction Type: Generalization

### 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." Refer to Annexe B for descriptions of registration and administrative statuses. 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 registry, the registration status will be updated to "Standard" with administrative status to "Final", after the necessary quality review has been completed.

The ISO/IEC 11179-3 metamodel requires that both registration status and administrative status also be recorded for data element concept. When all of the mandatory metadata attributes have been entered for this data element concept, it is assigned the registration status of "Recorded" and the administrative status of "In Quality Review." The descriptions of registration and administrative statuses in Annexe B also apply to data element concepts. Because the data element concept in this example was identified by an international standard, the registration status will be updated to "Standard" with administrative status to "Final", after the necessary quality review has been completed.

When administrative information is recorded for a value domain and for a conceptual domain as in this example, registration status and administrative status are also required. For this example, which includes permissible values and value meanings specified by an international standard, the registration status will be set to Standard and the administrative status to Final, after all mandatory value domain attributes and conceptual domain attributes have been entered and quality review has been performed.

### 5.2.9 Other names and codes from ISO 3166

Other names, i.e., full English names, and French names (both short and full), and all codes from ISO 3166 can also be registered with their individual value domains, representation, data element definitions, and data element names. All of the data elements associated with ISO 3166 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 share the same value meanings. They do, 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, i.e., the English-language country short name identified by the ISO 3166:1997 standard, are summarized in the following table, and in the first column of the table in Annexe A.

Metadata Attribute Name	ISO 3166, Enumerated
<b>1</b>	<b>Data Element Definition and Permissible Values</b>
<b>Data Element (DE) Definition Context</b>	Registry
<b>DE Definition</b>	The short, ISO-recognized name of a country, represented in the English language.
<b>Permissible Values (PV)</b>	All English-language short country names from ISO 3166, matched with value meanings. (recorded as Afghanistan, Albania,...., Zimbabwe)
<b>PV Begin Date</b>	19970110

<b>Metadata Attribute Name</b>		<b>ISO 3166, Enumerated</b>
	<b>PV End Date</b>	(Not applicable)
	<b>Value Domain (VD) Context</b>	Registry
	<b>VD Name</b>	ISO English-Language Country Short Names
	<b>VD Definition</b>	All short, ISO-recognized English-language names of all countries.
	<b>VD Component Identifier</b>	USA-EPA-OEI-OIC 5678:1
	<b>Datatype</b>	Alphanumeric
	<b>Minimum Characters</b>	4
	<b>Maximum Characters</b>	44
	<b>Format</b>	A(44)
	<b>Unit of Measure</b>	(Not applicable)
	<b>Precision</b>	(Not applicable)
	<b>VD Origin (Enumerated)</b>	ISO 3166-1:1997
	<b>VD Explanatory Comment</b>	(Not applicable)
	<b>Representation Class</b>	Name
	<b>Representation Class Qualifier</b>	Short
<b>2 Data Element Name and Identifier</b>		
	<b>DE Name Context</b>	Registry
	<b>DE Name</b>	ISO English-Language Country Short Name
	<b>DE Component Identifier</b>	USA-EPA-OEI-OIC 209033:1
<b>3 Other Metadata Attributes</b>		
	<b>DE Example</b>	China
	<b>DE Origin</b>	ISO 3166-1:1997, Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes
	<b>DE Explanatory Comment</b>	This data element is included in the mailing address standard.
	<b>Submitting Organization</b>	Office of Information Collection
	<b>Stewardship Contact</b>	Geopolitical Data Administrator
<b>4 Data Element Concept and Conceptual Domain</b>		
	<b>Data Element Concept (DEC) Context</b>	Registry
	<b>DEC Name</b>	Country Identifier
	<b>DEC Definition</b>	An identifier for a primary geopolitical entity of the world.
	<b>Object Class</b>	Country
	<b>Object Class Qualifier</b>	(Not applicable)
	<b>Property</b>	(Not applicable)
	<b>Property Qualifier</b>	(Not applicable)
	<b>DEC Component Identifier</b>	USA-EPA-OEI-OIC 12345:1
	<b>Conceptual Domain (CD) Context</b>	Registry
	<b>CD Name</b>	Countries of the World
	<b>CD Definition</b>	The primary geopolitical entities of the world.

<b>Metadata Attribute Name</b>		<b>ISO 3166, Enumerated</b>
	<b>CD Component Identifier</b>	USA-EPA-OEI-OIC 2468:1
	<b>Value Meaning (VM)</b>	The primary geopolitical entity known as <China>
	<b>VM Begin Date</b>	19970110
	<b>VM End Date</b>	(Not applicable)
	<b>VM Identifier</b>	<Assigned by system as 1001...1230: one to each VM>
<b>5</b>	<b>Classification Type Examples</b>	<b>Classification Value for Classification Types</b>
	<b>Keyword</b>	Country
	<b>Group</b>	Geopolitical Entities, Country Identifiers
	<b>Object Class</b>	Country
	<b>Layer of Abstraction Type</b>	Generalization
<b>6</b>	<b>Quality Control</b>	
	<b>DE Registration Status</b>	Standard
	<b>DE Administrative Status</b>	Final
	<b>VD Registration Status (includes PV)</b>	Standard
	<b>VD Administrative Status (includes PV)</b>	Final
	<b>DEC Registration Status</b>	Standard
	<b>DEC Administrative Status</b>	Final
	<b>CD Registration Status (includes VM)</b>	Standard
	<b>CD Administrative Status (included VM)</b>	Final

### 5.3 International standard with non-enumerated domain

This clause 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 Subclause 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: 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 registration authority of the 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 registration authority 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 60 degrees (0 to 59). 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, where the measurement includes decimal seconds. In data transfer, 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

ISO 19111, Spatial referencing by coordinates, describes the 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 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 smokestack). 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 Annexe B, based on ISO/IEC 11179-4. The rules require that a data element definition be unique within a registration authority, 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 for latitude, 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 name for sexagesimal latitude is "Sexagesimal Measures of Latitude," and the value domain 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 description, 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.

Note that the ISO/IEC 11179-3 metamodel provides for the optional collection of administrative data for value domains. In this example, using ISO 6709 as an example of a non-enumerated domain, no administrative data is recorded.

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

Other value domain attributes for this example include:

- Identifier: A system generated identifier for the value domain and its version identifier.
- Domain Type: Non-enumerated. This attribute is an extension of 11179-3. It has been found to be useful by the registration authority in this example.
- Description: All sexagesimal measures of the distance of an angle north or south of the equator. The range limit for degrees is 0-90, for minutes is 0-59, and for seconds is 0-59. Latitude measures on or north of the equator are denoted as "+", and latitude measures south of the equator are denoted as "-".
- Datatype: The datatype for latitude measure is "alphanumeric" to explicitly include the directional symbol and decimal point, where appropriate.
- Minimum and maximum characters: The known minimum field length at this time is seven (+/- DDMMSS) where no decimal seconds are recorded. The maximum field length is 13 (+/- DDMMSS.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.
- Unit of Measure: Latitude for this example is measured in sexagesimal units.
- Precision: The number of decimal places recorded indicates the precision of the measure.

### 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 registration authority in this example, the data element identifier and version identifier (DI:VI = 312345:1) are assigned automatically by the registry software. The RAI is USA-EPA-OEI-OIC.

#### 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 Annexe B, 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 class and a property, where appropriate. Qualifiers shall be used to differentiate between names that would otherwise be the same. The representation class term shall 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 English language alphabet.
- **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.

DE Example: +674532 and +674531.85435

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

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

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

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

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

Submitting Organization: Office of Information Collection

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

Stewardship Contact: Geographic Information Systems Administrator

- Administrative metadata, such as Creation Date are recorded or captured automatically by the system where applicable. Other administrative data may be entered manually.

### 5.3.6 Data element concept

The methodology to be used for deriving a data element concept is described in Subclause 5.1.6 and Annexe B of this technical report. A data element concept is the data element without representation. 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. Administrative information must be collected about a data element concept. Therefore, both name and definition are recorded in context, which for this example is the context of the Registry (i.e., the rules that have been established by the registration authority for this example registry).

Both object classes and properties may be stored for data element concepts. For this example, Latitude is recorded as the object class and Distance as its property. No qualifiers are recorded.

A conceptual domain is a set of possible value meanings for a data element concept. The value meanings for this conceptual domain, however, are not enumerated. For non-enumerated domains, such as latitude distance 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 explicit value meanings are stored in the registry for latitude coordinates. Recording of administrative information about conceptual domains is optional according to the ISO/IEC 11179-3 metamodel. Therefore, for this example, names and definitions have not been assigned to the conceptual domain attribute, and no administrative data has been recorded.

### 5.3.7 Classification

This data element might be classified according to a source or by one or more of 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 Latitude Sexagesimal Measure with similar data elements according to concept for translation or by general subject matter.

Subject group: Geographic Point Location.

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

Object Class: Latitude

- Identify the abstraction layer of this data element:

Layer of Abstraction Type: Generalization

**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 a higher status of the data element. The data element, however, would not be expected to be assigned the status of "Standard." The data element is not expected to become 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 "Certified" with an administrative status of "No further action."

The ISO/IEC 11179-3 metamodel requires that both registration status and administrative status also be recorded for data element concept. When all of the mandatory metadata attributes have been entered for this data element concept, it is assigned the registration status of "Recorded" and the administrative status of "In Quality Review." The descriptions of registration and administrative statuses in Annexe B also apply to data element concepts. Because the data element concept in this example was identified by an international standard, the registration status will be updated to "Standard" with administrative status to "Final", after the necessary quality review has been completed.

For this example, no registration or administrative statuses are required for the value domain or the conceptual domain. Quality review of these attributes will be incorporated into the reviews of the data element itself and its corresponding data element concept.

**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 ISO 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 longitude data element concept and 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 6709 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."

**5.3.10 Summary of metadata attributes**

The following table summarizes the metadata attributes assigned to latitude sexagesimal measure in the preceding paragraphs in Subclause 5.3. The table in Annexe A also contains this data in the second metadata column.

Metadata Attribute Name		ISO 6709, Non-enumerated
<b>1</b>	<b>Data Element Definition and Permissible Values</b>	
	Data Element (DE) Definition Context	Registry

<b>Metadata Attribute Name</b>		<b>ISO 6709, Non-enumerated</b>
	<b>DE Definition</b>	The sexagesimal measure of the angular distance of a position on the earth on a meridian north or south of the equator.
	<b>Permissible Values (PV)</b>	Measures of latitude in degrees, minutes, and seconds. (informational only; not recorded <i>per se</i> )
	<b>PV Begin Date</b>	(Not applicable)
	<b>PV End Date</b>	(Not applicable)
	<b>Value Domain (VD) Context</b>	(Not applicable)
	<b>VD Name</b>	Sexagesimal Measures of Latitude
	<b>VD Description</b>	All sexagesimal measures of the distance of an angle north or south of the equator. Latitude measure on or north of the equator is positive; latitude measure south of the equator is negative. The range limit for degrees is 0-90, for minutes is 0-59, and for seconds is 0-59.
	<b>VD Component Identifier</b>	(Not applicable)
	<b>VD Identifier</b>	9876:1
	<b>Datatype</b>	Alphanumeric
	<b>Minimum Characters</b>	7
	<b>Maximum Characters</b>	13
	<b>Format</b>	A(13) +/-DDMMSS.SSSSS
	<b>Unit of Measure</b>	Sexagesimal
	<b>Precision</b>	Number of decimal places recorded
	<b>VD Origin (Enumerated)</b>	(Not applicable)
	<b>VD Explanatory Comment</b>	(Not applicable)
	<b>Representation Class</b>	Measure
	<b>Representation Class Qualifier</b>	Sexagesimal
<b>2</b>	<b>Data Element Name and Identifier</b>	
	<b>DE Name Context</b>	Registry
	<b>DE Name</b>	Latitude Sexagesimal Measure
	<b>DE Component Identifier</b>	USA-EPA-OEI-OIC 312345:1
<b>3</b>	<b>Other Metadata Attributes</b>	
	<b>DE Example</b>	+674532 and +674531.85435
	<b>DE Origin</b>	ISO 6709-1983 (E), Standard representation of latitude, longitude and altitude for geographic point locations.
	<b>DE Explanatory Comment</b>	Latitude sexagesimal converts to latitude degrees by the following formula: seconds x 60 = decimal minutes, total minutes x 60 = decimal degrees.
	<b>Submitting organization</b>	Office of Information Collection
	<b>Stewardship Contact</b>	Geographic Information Systems Administrator
<b>4</b>	<b>Data Element Concept and Conceptual Domain</b>	
	<b>Data Element Concept (DEC) Context</b>	Registry
	<b>DEC Name</b>	Latitude Distance

<b>Metadata Attribute Name</b>		<b>ISO 6709, Non-enumerated</b>
	<b>DEC Definition</b>	A measure of the angular distance of a point on the surface of the earth north or south of the equator.
	<b>Object Class</b>	Latitude
	<b>Object Class Qualifier</b>	(Not applicable)
	<b>Property</b>	Distance
	<b>Property Qualifier</b>	(Not applicable)
	<b>DEC Component Identifier</b>	USA-EPA-OEI-OIC 23456:1
	<b>Conceptual Domain (CD) Context</b>	(Not applicable)
	<b>CD Name</b>	(Not applicable)
	<b>CD Definition</b>	(Not applicable)
	<b>CDComponent Identifier</b>	(Not applicable)
	<b>Enumerated Value Meaning (VM)</b>	(Not applicable)
	<b>VM Begin Date</b>	(Not applicable)
	<b>VM End Date</b>	(Not applicable)
	<b>VM Identifier</b>	(Not applicable)
<b>5</b>	<b>Classification Type Examples</b>	<b>Classification Values for Classification Types</b>
	<b>Keyword</b>	Horizontal Coordinate, Spatial, Latitude
	<b>Group</b>	Geographic Point Location
	<b>Object Class</b>	Latitude
	<b>Layer of Abstraction Type</b>	Generalization
<b>6</b>	<b>Quality Control</b>	
	<b>DE Registration Status</b>	Certified
	<b>DE Administrative Status</b>	No further action
	<b>VD Registration Status</b>	(Not applicable)
	<b>VD Administrative Status</b>	(Not applicable)
	<b>DEC Registration Status</b>	Standard
	<b>DEC Administrative Status</b>	Final
	<b>CD Registration Status</b>	(Not applicable)
	<b>CD Administrative Status</b>	(Not applicable)

#### 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 Subclause 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 the data element can

reuse domain and conceptual information based on a standard data element, as in this example, the data element can be registered as Recorded. Many data elements, however, must be registered as Incomplete, and all metadata attributes identified in the model as Mandatory, might never be complete.

The table in Subclause 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 a data element for the mailing address country name is registered, the registry for the registration authority 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 to 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 Subclause 5.2, and the conceptual domain for this data element is the same. Therefore, the data element will be registered, reusing the conceptual domain and permissible values 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 Annexe B of this technical report, with additional examples that will provide assistance in formulating 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 application data element is described in Subclause 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). The scope of the permissible values for this data element includes the short English-language name for all countries. The registration authority has determined that this data element is represented by the same value domain as exists for the data element USA-EPA-OEI-OIC 209033:1, ISO English-Language Country Short Name. It is not necessary to record additional permissible values or other value domain information for this data element, since value domain USA-EPA-OEI-OIC 5678:1 is being reused.

### 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 registry. The Registration Authority for this data element in this example is USA-EPA-OEI-OIC.

#### 5.4.4.2 Name context and naming convention

In this example, the data element is assigned two names, each with its own context. First is the system name context, since this data element was identified as being part of an application system, and retention of 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 (i.e., table name.column 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 Latin 1.
- **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 Annexe B.

- **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 Latin 1. There are no controlled word lists for naming application data elements.
- **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 its 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

representation term. The representation is the same as for the ISO standard data element. For the application data element (e.g., country name in a mailing address entity) the entity is "address" qualified by "mailing." 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 Subclause 5.1.3 should be followed. The Registry name of this data element, based on ISO/IEC 11179-5 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.  
DE Example: Denmark
- Identify the origin for this data element as the standard from which the permissible values are obtained.  
DE Origin: Facility Data System, Environmental Protection Agency
- Record any notes or comments that might provide additional information about the data element that is not included in the definition.  
DE Explanatory Comment: 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 Assurance
- Record the name of the individual or organization assigned the responsibility for monitoring and maintaining the data element as the stewardship contact.  
Stewardship Contact: Facility Data Systems Hot Line

Other administrative metadata, such as Creation Date are recorded or captured automatically by the system where applicable. Other administrative data may be entered manually.

#### 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 Subclause 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 (CD ID 2468:1) and the value domain (5678:1) that were established for the standard data element, ISO English-Language Short Name (IRDI USA-EPA-OEI-OIC 209033:1), described in Subclause 5.2.

**5.4.7 Classification**

This data element might be classified by source or according to one or more of 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, which might include such data elements as street address name, city name, and postal code.

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

Object Class: Country, Mailing Address

- Identify the abstraction layer(s) of this data element:

Layer of Abstraction Type: Specification; Decomposition

**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 it often would 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. This includes 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 Subclause 5.4. The table in Annexe A also contains this metadata.

Metadata Attribute Name		Application, Enumerated	
<b>1</b>	<b>Data Element Definition and Permissible Values</b>		
	<b>Data Element (DE) Definition Context</b>	Registry	Facility Data System
	<b>DE Definition</b>	The name of the country where a mailpiece is delivered.	The name of a country where the addressee is located.
	<b>Permissible Values (PV)</b>	All English-language short country names from ISO 3166, matched with value meanings. (recorded as: Afghanistan, Albania,....., Zimbabwe)	

<b>Metadata Attribute Name</b>		<b>Application, Enumerated</b>	
	<b>PV Begin Date</b>	19970110	
	<b>PV End Date</b>	(Not applicable)	
	<b>Value Domain (VD) Context</b>	Registry	
	<b>VD Name</b>	ISO English-Language Country Short Names	
	<b>VD Definition</b>	All short, ISO-recognized English-language names of all countries.	
	<b>VD Component Identifier</b>	USA-EPA-OEI-OIC 5678:1	
	<b>Datatype</b>	Alphanumeric	
	<b>Minimum Characters</b>	4	
	<b>Maximum Characters</b>	44	
	<b>Format</b>	A(44)	
	<b>Unit of Measure</b>	(Not applicable)	
	<b>Precision</b>	(Not applicable)	
	<b>VD Origin (Enumerated)</b>	ISO 3166-1:1997	
	<b>VD Explanatory Comment</b>	(Not applicable)	
	<b>Representation Class</b>	Name	
	<b>Representation Class Qualifier</b>	Short	
<b>2 Data Element Name and Identifier</b>			
	<b>DE Name Context</b>	Registry	Facility Data System
	<b>DE Name</b>	Mailing Address Country Name	Mailing_Address.Country_Name
	<b>DE Component Identifier</b>	USA-EPA-OEI-OIC 5394:1	
<b>3 Other Metadata Attributes</b>			
	<b>DE Example</b>	Denmark	
	<b>DE Origin</b>	ISO 3166-1:1997, Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes	
	<b>DE Comment</b>	This data element is required when mail is intended to be delivered outside the country of origin.	
	<b>Submitting organization</b>	Office of Enforcement and Compliance Assurance	
	<b>Stewardship Contact</b>	Facility Data Administrator	
<b>4 Data Element Concept and Conceptual Domain</b>			
	<b>Data Element Concept (DEC) Context</b>	Registry	
	<b>DEC Name</b>	Address Country Identifier	
	<b>DEC Definition</b>	An identifier for a primary geopolitical entity of the world which indicates an address.	
	<b>Object Class</b>	Address	
	<b>Object Class Qualifier</b>	(Not applicable)	
	<b>Property</b>	Country Identifier	
	<b>Property Qualifier</b>	(Not applicable)	
	<b>DEComponent Identifier</b>	USA-EPA-OEI-OIC 12468:1	

<b>Metadata Attribute Name</b>		<b>Application, Enumerated</b>
	<b>Conceptual Domain (CD) Context</b>	Registry
	<b>CD Name</b>	Countries of the World
	<b>CD Definition</b>	The primary geopolitical entities of the world.
	<b>CD Component Identifier</b>	USA-EPA-OEI-OIC 2468:1
	<b>Value Meaning (VM)</b>	The primary geopolitical entity known as <China>
	<b>VM Begin Date</b>	19970110
	<b>VM End Date</b>	(Not applicable)
<b>5</b>	<b>Classification Type Examples</b>	<b>Classification Values for Classification Types</b>
	<b>Keyword</b>	Country, Address, Mailing
	<b>Group</b>	Mailing Address
	<b>Object Class</b>	Address, Country
	<b>Layer of Abstraction Type</b>	Specification; Decomposition
<b>6</b>	<b>Quality Control</b>	
	<b>DE Registration Status</b>	Recorded
	<b>DE Administrative Status</b>	In Quality Review
	<b>VD Registration Status</b>	Standard
	<b>VD Administrative Status</b>	Final
	<b>DEC Registration Status</b>	Recorded
	<b>DEC Administrative Status</b>	In Quality Review
	<b>CD Registration Status</b>	Standard
	<b>CD Administrative Status</b>	Final

### 5.5 Classification of data elements

The registration authority may determine that is appropriate to group certain data elements, out of some observed relationship among the data elements or a perceived value in identifying those data elements together. This grouping is referred to as a classification of the data elements. Data elements may be classified by grouping them according to the documents, systems or other sources where the data elements are identified. After the data elements that are to be associated with the source have been identified, the source itself is registered with the metadata that provides certain information about the source/classification. The metadata answers the following questions: How is the source identified? Why has the group of data elements been established? What is the authority for this classification or grouping of data elements? What is the potential use for the group of data elements? The examples provided in this clause describe how to formulate attributes about a document, system, or other source of data element classification. First a general procedure for registering data element classifications is described, followed by examples of registration of three types of data element classifications, including data elements from:

- A document that contains a list of data elements.
- A standard that identifies a group of standard data elements.
- A composite data element where individual data elements are used together to form a single data element.

Other classifications of data elements might be identified in a registry, such as the following examples:

- An entity for a computer information system (e.g., mailing address data elements).
- A classification of data elements (e.g., geopolitical data elements).
- Data elements that comprise an information system (e.g., data elements in a hazardous waste management system).
- Data elements that are used in a data collection form.

These other classifications of data element would also be registered according to the general procedures that follow.

### 5.5.1 General procedures for registering a classification of data elements

The steps required to register a group of data elements by a classification are similar to those required for registering a data element. First the practitioner must understand the need for the group, identify the data elements to be contained in the group, and the source of the group of data elements.

#### 5.5.1.1 Understanding the classified group

The perceived need for the classified group must first be identified. Will the data elements to be grouped serve a purpose such as facilitating the development of an information system, identification of all of the data elements that comprise a standard, or documentation of the data that exists in an information system?

Are there other classifications in the registry or federation of registries that provide similar groupings of data elements? Can the other groups be modified so that they meet the newly identified need for a group of data elements? For example, where there is a group of data elements contained in an international standard, is there a need for a group of data elements that meets a specific organizational need for the same or similar data elements?

#### 5.5.1.2 Specify the data elements

Where the need for classification of a group of data elements has been established, the next step is to identify the specific data elements to be grouped together in the registry. Are the data elements to be used together for purposes of data conversion (e.g., conversion of codes to names or conversion of data values based on units of measure)? Are all of the data elements in a group required to be present to represent a particular type of data (i.e., mandatory), or are some of the data elements optional? Are some of the data elements used conditionally (i.e., required if some other data element is or is not included when the group is used together)?

#### 5.5.1.3 Understanding the source of the group of data elements

Classification of a group of data elements in a registry requires that certain characteristics of the group be 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 registration authority are:

- Type of Source: Document, system, standard, or other source of the data elements to be grouped. The name of the type of source is the classification scheme type name.
- Name/Title: The name or title of the classification for a group of data elements.
- Definition: Text that describes the features of, specifies relationships of, or establishes the context for classification of a group of data elements.
- Organization: The originating point of information that provides an authoritative reference for a group of data elements.
- Explanatory Comment: The text that explains the reasons for using the selected source materials in development of a group of data elements and describes how a group of data elements can be used.

- Component Identifier: The system generated component identifier for a classified group of data elements, including the RAI, group ID and VI. Other administrative information, such as Creation date is recorded or captured from the system when the classification is entered into the registry. Registration status and administrative status are recorded, following the same general guidelines as for other data element components. Subclause 5.5 is not comprehensive for recording a fully defined classification scheme. Therefore, for the examples in this technical report, registration status is assigned the value of "Incomplete," and administrative status is assigned the value of "Awaiting information."

### 5.5.2 Data elements in a document

The Federal Geographic Data Committee document that records the proposed data elements for unique identification of a facility is an example of data elements in a document that might be grouped together.

#### 5.5.2.1 Understanding the classified group

The document, *Facility Location and Identification Data Standard*, is being reviewed by the Federal Geographic Data Committee (FGDC) as a proposed standard to be used for locating and identifying facilities of interest to the Federal Government and to the private sector of the United States. The document contains a core set of the minimum amount of information necessary to uniquely identify a facility. The group of data elements can be used by other organizations to ensure that their facility information management systems support the identification of place-based objects that are generally known as facilities.

There are no other national or international standards that contain the same core set of data elements as the FGDC proposed standard. The standard for the Agency responsible for the registry used for this example includes more detailed data elements than those required by the FGDC proposed standard.

#### 5.5.2.2 Specify the data elements

The core set of data elements that are used to locate and identify facilities includes the following:

- Facility Name: The public or commercial name of a facility, such as is commonly recognized by the local community.
- Facility Category Type: A descriptor that indicates the type of activity that takes place at the facility or the general purpose of the facility.
- Facility Identification Number: The non-intelligent, unique identification number assigned to a facility.
- Latitude Measure: The measure of the angular distance on a meridian north or south of the equator.
- Longitude Measure: The measure of the angular distance on a meridian east or west of the prime meridian.

All data elements in the above core set are listed as mandatory (i.e., they must all be present to uniquely identify a facility).

#### 5.5.2.3 Understanding the source of the group of data elements

The characteristics of the group of facility identification data elements includes the following:

- Type of Source/Classification Type: Document
- Name/Title: Federal Geographic Data Committee *Facility Location and Identification Standard*.
- Definition: The document provides a core set of standardized data elements that supports the location and identification of place-based objects that are generally known as facilities. Facilities incorporate the properties of being objects established at specific places for specific purposes.
- Organization: Federal Geographic Data Committee, Facility Working Group

- Explanatory Comment: The FGDC facility standard was selected for developing this core set of data elements, because it is intended to become a standard for U.S. Federal Agencies and the private sector. The purpose of the document is to facilitate the development and promulgation of a standard that supports identification of place-based objects generally known as facilities.
- Component Identifier: The system generated component identifier for this group USA-EPA-OEI-OIC 1243:1.
- Registration status: Incomplete.
- Administrative status: Awaiting information.

### 5.5.3 Data elements in a standard

The data elements listed in the ISO 6709 standard representation of latitude, longitude and altitude for geographic point locations was selected for this example.

#### 5.5.3.1 Understanding the classified group

An example of a group of data elements that are used together, both for the purpose of data translation and for identifying a precise location, is the ISO 6709 group of data elements that specify the standard representation of latitude, longitude, and altitude for geographic point locations.

#### 5.5.3.2 Specify the data elements

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 Measure. 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 (i.e., conversion) 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, and they must precede the altitude measure).

#### 5.5.3.3 Understanding the source of the group of data elements

The characteristics of the group of locational representation data elements includes the following:

- Type of Source/Classification Type: International data standard
- Name/Title: *Standard representation of latitude, longitude and altitude for geographic point locations*, International Organization for Standards, ISO 6709, 1983-05-15.
- Definition: The standard includes the horizontal and vertical coordinates that define a point on earth.

- Organization: International Organization for Standardization.
- Explanatory Comment: 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. 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.
- Component Identifier: The system generated component identifier for this group is USA-EPA-OEI-OIC 1345:1.
- Registration status: Incomplete.
- Administrative status: Awaiting information.

### 5.5.4 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. Each of the data elements in the urban style street address 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.

#### 5.5.4.1 Understanding the classified group

The U.S. Postal Service (USPS) has identified the components of an urban street address used as a delivery point for mailing addresses. There are seven components, each of which can be defined as a data element with its own value domain. Four of the components, represented as codes, have enumerated domains.

#### 5.5.4.2 Specify the data elements

Urban style street address 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.

#### 5.5.4.3 Understanding the source of the group of data elements

The characteristics of the urban style street address group of composite data elements includes the following:

- Type of Source/Classification Type: Complex data element.
- Name/Title: Urban-style Street Address.
- Definition: A set of precise and complete data elements that cannot be subdivided and that can be combined into an urban-style street address.
- Organization: U.S. Postal Service, Publication 28: Postal Address Standards.
- Explanatory Comment: 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. 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.
- Component Identifier: The system generated component identifier for this group is USA-EPA-OEI-OIC 2543:1.
- Registration status: Incomplete.
- Administrative status: Awaiting information.

## 5.6 Linking of data elements

The ISO/IEC 11179-3 metamodel does not provide for linking of data elements. A registration authority, however, might choose to extend the model to link data elements based on their layers of abstraction, including generalization to specialization, and other relationships. Linkages can occur in both vertical relationships (e.g., from general to more specific) and horizontal relationships (e.g., with equivalent layers of specialization). They can also be linked according to other relationships (e.g., data elements that are always used together). These are defined as follows:

- Vertical relationships are those where a specialized data element that has been registered for a particular purpose is related to a generalized data element that is intended for a general purpose. For example, the following data elements can be linked vertically in parent/child relationships, based on 3 layers of generalization/specialization where 1 is the most general. The vertical linkages are incremented by one for each layer of specialization:
  - 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 the same layer of specialization, with equivalent data value domains. For example, the following data elements can be linked horizontally as equivalent layers of specialization in Envirofacts, a data warehouse of EPA environmental systems.

The third layer: 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)

- Related Through Use

An example of data elements that are related through use are data elements where values are recorded that must always be accompanied by a data element that records the unit of measure for that value. Data elements of this type are the following:

- Sample Quantity. The volume of sample that was collected.
- Sample Quantity Units Code. The code that represents the unit of measure for the volume of collected sample.

## 6 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 annexe, 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.

### 6.1 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.

#### 6.1.1 Data element concepts

One data element concept that would be associated with Biological Organisms would be "Biological Organism Identifier," where "Biological Organism" would be the object class, and "Identifier" the property. The definition of this data element concept would be "An identifier for a biological organism."

#### 6.1.2 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:

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

- ITIS Taxonomic Serial Number: The unique number assigned to a biological organism by the Integrated Taxonomic Information System (ITIS)<sup>1</sup>.
- Biological Identification Number: The unique number assigned to a biological organism by the Biological Registry System.

### 6.1.3 Permissible values

Permissible values for these data elements would not be enumerated, as described above in Subclause 6.1. The permissible values, however, will all be names, numbers, and codes that represent an implied value meaning of "The biological organism known as...".

## 6.2 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:

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

The conceptual domain for biological organism type would be enumerated, and value meanings entered into the metadata registry for the domain.

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.

### 6.2.1 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 class, and Type is the property. The definition of the data element concept might be "A type of a biological organism."

### 6.2.2 Data elements

Data elements associated with this data element concept might be:

- Biological Organism Type Name — The name of the type of a biological organism.

<sup>1</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.

- Biological Organism Type Code — The code that represents a type of biological organism.

### **6.2.3 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.

### 6.3 Top down registration

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

<b>Order of Registering Components for Top Down Registration of a Data Element</b>	
<b>1</b>	<b>Conceptual Domain (CD)</b>
	CD Context
	CD Name
	CD Definition
	CD Identifier
	CD Registration Status
	CD Administrative Status
<b>2</b>	<b>Value Meaning (VM)</b>
	VM Description
	VM Begin Date
	VM End Date
	VM Identifier
<b>3</b>	<b>Data Element Concept (DEC)</b>
	DEC Context
	DEC Name
	DEC Definition
	Object Class
	Object Class Qualifier
	Property
	Property Qualifier
	DEC Identifier
	DEC Registration Status
	DEC Administrative Status
<b>4</b>	<b>Value Domain (VD)</b>
	VD Context
	VD Name
	VD Definition/Description
	VD Identifier
	Datatype
	Minimum Characters
	Maximum Characters
	Unit of Measure
	Precision
	VD Origin
	VD Explanatory Comment
	Permissible Values (PV)
	PV Begin Date
	PV End Date

Order of Registering Components for Top Down Registration of a Data Element	
	Representation Class
	Representation Class Qualifier
	VD Registration Status
	VD Administrative Status
<b>5</b>	<b>Data Element (DE) Definition and Name</b>
	DE Context
	DE Definition
	DE Name
	Registration Authority Identifier
	DE Identifier:Version Identifier
	DE Example
	DE Origin
	DE Comment
	DE Registration Status
	DE Administrative Status

## 7 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.

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.

## Annexe A

## Table of data element attributes for examples

Annexe A contains a table that includes the data element attributes for the examples provided earlier in this technical report. The table provides examples of the metadata associated with the English-language short name from the ISO 3166 standard, a data element from the ISO 6709 standard, and an application data element. 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.

Metadata Attribute Name	ISO 3166, Enumerated	ISO 6709, Non-enumerated	Application, Enumerated	
<b>1 Data Element Definition and Permissible Values</b>				
<b>Data Element (DE) Definition Context</b>	Registry	Registry	Registry	Facility Data System
<b>DE Definition</b>	The short, ISO-recognized name of a country, represented in the English language.	The sexagesimal measure of the angular distance of a position on the 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.
<b>Permissible Values (PV)</b>	All English-language short country names from ISO 3166, matched with value meanings. ( recorded as: Afghanistan, Albania,....., Zimbabwe)	Measures of latitude in degrees, minutes, and seconds. (informational only; not recorded <i>per se</i> )	All English-language short country names from ISO 3166, matched with value meanings. (recorded as: Afghanistan, Albania,....., Zimbabwe)	
<b>PV Begin Date</b>	19970110	(Not applicable)	19970110	
<b>PV End Date</b>	(Not applicable)	(Not applicable)	(Not applicable)	
<b>Value Domain (VD) Context</b>	Registry	(Not applicable)	Registry	
<b>VD Name</b>	ISO English-Language Country Short Names	Sexagesimal Measures of Latitude	ISO English-Language Country Short Names	
<b>VD Definition</b>	All short, ISO-recognized English-language names of all countries.	(Not applicable)	All short, ISO-recognized English-language names of all countries.	
<b>VD Description</b>	(Not applicable)	All sexagesimal measures of the distance of an angle north or south of the equator. Latitude measure on or north of the equator is positive; latitude measure south of the equator is negative. The range limit for degrees is 0-90, for minutes is 0-59, and for seconds is 0-59.	(Not applicable)	

Metadata Attribute Name	ISO 3166, Enumerated	ISO 6709, Non-enumerated	Application, Enumerated	
VD Component Identifier	USA-EPA-OEI-OIC 5678:1	(Not applicable)	USA-EPA-OEI-OIC 5678:1	
VD Identifier	(Not applicable)	9876:1	(Not applicable)	
Datatype	Alphanumeric	Alphanumeric	Alphanumeric	
Minimum Characters	4	7	4	
Maximum Characters	44	13	44	
Format	A(44)	A(13) +/-DDMMSS.SSSSS	A(44)	
Unit of Measure	(Not applicable)	Sexagesimal	(Not applicable)	
Precision	(Not applicable)	Number of decimal places recorded	(Not applicable)	
VD Origin (Enumerated)	ISO 3166-1:1997	(Not applicable)	ISO 3166-1:1997	
VD Explanatory Comment	(Not applicable)	(Not applicable)	(Not applicable)	
Representation Class	Name	Measure	Name	
Representation Class Qualifier	Short	Sexagesimal	Short	
<b>2 Data Element Name and Identifier</b>				
DE Name Context	Registry	Registry	Registry	Facility Data System
DE Name	ISO English-Language Country Short Name	Latitude Sexagesimal Measure	Mailing Address Country Name	Mailing_Address.Country_Name
DE Component Identifier	USA-EPA-OEI-OIC 209033:1	USA-EPA-OEI-OIC 312345:1	USA-EPA-OEI-OIC 5394:1	
<b>3 Other Metadata Attributes</b>				
DE Example	China	+674532 and +674531.85435	Denmark	
DE Origin	ISO 3166-1:1997, Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes	ISO 6709-1983 (E), Standard representation of latitude, longitude and altitude for geographic point locations.	ISO 3166-1:1997, Codes for the representation of names of countries and their subdivisions -- Part 1: Country codes	
DE Comment	This data element is included in the mailing address 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 Collection	Office of Information Collection	Office of Enforcement and Compliance Assurance	
Stewardship Contact	Geopolitical Data Administrator	Geographic Information Systems Administrator	Facility Data Systems Administrator	

Metadata Attribute Name	ISO 3166, Enumerated	ISO 6709, Non-enumerated	Application, Enumerated
<b>4</b>	<b>Data Element Concept and Conceptual Domain</b>		
<b>Data Element Concept (DEC) Context</b>	Registry	Registry	Registry
<b>DEC Name</b>	Country Identifier	Latitude Distance	Address Country Identifier
<b>DEC Definition</b>	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 a primary geopolitical entity of the world which indicates an address.
<b>Object Class</b>	Country	Latitude	Address
<b>Object Class Qualifier</b>	(Not applicable)	(Not applicable)	(Not applicable)
<b>Property</b>	(Not applicable)	Measure	Country Identifier
<b>Property Qualifier</b>	(Not applicable)	(Not applicable)	(Not applicable)
<b>DEC Component Identifier</b>	USA-EPA-OEI-OIC 12345:1	USA-EPA-OEI-OIC 23456:1	USA-EPA-OEI-OIC 12468:1
<b>Conceptual Domain (CD) Context</b>	Registry	(Not applicable)	Registry
<b>CD Name</b>	Countries of the World	(Not applicable)	Countries of the World
<b>CD Definition</b>	The primary geopolitical entities of the world.	(Not applicable)	The primary geopolitical entities of the world.
<b>CD Component Identifier</b>	2468:1	(Not applicable)	2468:1
<b>Value Meaning (VM)</b>	The primary geopolitical entity known as <China>	(Not applicable)	The primary geopolitical entity known as <Denmark>
<b>VM Begin Date</b>	19970110	(Not applicable)	19970110
<b>VM End Date</b>	(Not applicable)	(Not applicable)	(Not applicable)
<b>VM Identifier</b>	<Assigned by system as 1001...1230: one to each VM>	(Not applicable)	<Assigned by system as 1001...1230: one to each VM>
<b>5</b>	<b>Classification Type Examples</b>		
<b>Keyword</b>	Country	Horizontal Coordinate, Spatial, Latitude	Country, Address, Mailing
<b>Group</b>	Geopolitical Entities, Country Identifiers	Geographic Point Location	Mailing Address
<b>Object</b>	Country	Latitude	Address, Country
<b>Layer of Abstraction Type</b>	Generalization	Generalization	Specification; Decomposition
<b>6</b>	<b>Quality Control</b>		
<b>DE Registration Status</b>	Standard	Certified	Recorded
<b>DE Administrative Status</b>	Final	No further action	In Quality Review
<b>VD Registration Status</b>	Standard	(Not applicable)	Standard
<b>VD Administrative Status</b>	Final	(Not applicable)	Final
<b>DEC Registration Status</b>	Standard	Standard	Recorded
<b>DEC Administrative Status</b>	Final	Final	In Quality Review

<b>Metadata Attribute Name</b>	<b>ISO 3166, Enumerated</b>	<b>ISO 6709, Non-enumerated</b>	<b>Application, Enumerated</b>
<b>CD Registration Status</b>	Standard	(Not applicable)	Standard
<b>CD Administrative Status</b>	Final	(Not applicable)	Final

## Annexe B

### Application of ISO/IEC 11179, Parts 4-6

This annexe includes additional detailed examples of using Parts 4-6 for registering data element metadata. It is intended to be used to supplement to the description of data element registration provided in Subclause 5.1 of this technical report.

#### B.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 Clause B.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 that should be followed in formulating a definition. The standard does not specify syntactical requirements (i.e., word order and structure) that 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 Subclause 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 registry) follow. A syntax that has been adopted by one registration authority is also included in this clause.

##### B.1.1 Rules for definitions

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 within a specific context).
- 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.

##### B.1.1.1 Uniqueness

According to the standard rules for formulating data definitions, a data definition and context shall be unique for a specific context within any 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 Subclause 5.6 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.

### B.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.

### B.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 ISO-recognized short name of a country.

**Poor:** The ISO-recognized 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 ISO-recognized, short name of a country *that is not its long name*.

### B.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.

### B.1.1.5 Contain only commonly used abbreviations or acronyms

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.

#### **B.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.

#### **B.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.

##### **B.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.

### B.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 is documented by the Origin attribute in the value domain. 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.

### B.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."

### B.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.

### B.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.

### **B.1.2.6 Avoid circular reasoning**

Two definitions should not be defined in terms of each other. A definition should not use the definition of another concept 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.

### **B.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.

### **B.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:

- 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.

- 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 ...."

#### B.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.

## **B.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 field as well as other representational attributes.

### **B.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 "ISO English-Language Country Short 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 Annexe as B.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 Subclause B.2.2.

### **B.2.2 Value domain**

The value domain is defined as a set of permissible values. It 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.

- Data Identifier and Version Number. The identifier and version number that represents the set of permissible values in 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., English-Language Country Short Names). Note that a definition can also be used to describe the value domain.
- Definition. A phrase expressing the essential value of the value domain. The definition is plural, because it encompasses all values for that domain.
- Name and Definition Context. A designation or description of the application environment or discipline in which a name is applied or from which it originates. A registration authority will determine the contexts in which names and definitions are appropriate for a value domain. Context is required whenever administrative data are recorded for a value domain.
- 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 as part of the description of the datatype.

For example, a registration authority might define datatypes as alphanumeric, numeric, date, and integer. An alphanumeric datatype is composed of 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." ISO 11404 provides guidance on language independent datatypes.

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

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 technical report. Country names are a fixed list of countries, maintained by international standards; therefore, the domain type is enumerated. Enumerated domains are identified by their association with specified permissible values.

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 identified by a definition that clearly describes the values that are permissible for the data element.

Domain type is not an attribute in the 11179-3 model, but it has been found to be a useful attribute for differentiating between enumerated and non-enumerated domains in a registry.

- Comment. An explanatory comment might be used to describe the procedure, rule, reference, range, or other process for which a set of all permissible values is defined for the non-enumerated value domain.
  - 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°B.
- Maximum and minimum characters.

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, alphanumerics 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 Annexe A to demonstrate uniqueness and reusability of the value domain.

Closely related to the value domain is the **Rule and the Rule Description**. The rule is the logical, mathematical, or other operation that specifies the derivation of the data element values. The rule can identify data elements upon which a mathematical process is used to calculate or determine the value of a data element. For non-enumerated value domains, the rule description can be used to describe the procedure, the reference, or the maximum and minimum values for the range that limits the permissible values for a data element.

### B.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 technical report 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.

### B.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 flagpole in the North parking lot of the installation. This location is near the center of the facility.

### B.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 registration authority identifier, data element identifier, and version 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 Subclause B.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 Subclause B.3.1. Each registry establishes its own naming convention. Suggestions for establishing a naming convention are provided in Subclause B.3.2.

#### B.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.

An attribute for Language context is also contained in the model. This attribute is useful for registries that represent more than one language. For example, a registry might have a Standard name context for both French and Spanish.

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 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 registry model also includes an attribute for the precision required for recording the data value.

### B.3.2 Establish a naming convention

The Registration Authority 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 technical report, the scope is limited to the Registry name context. For example, a data element might have the name Regulation Abstract Text with the context Registry and the name Abstract 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 Unknown.
- **The Authority That Establishes Names.** The registration authority establishes the naming convention for a registry for the context of Registry. The Environmental Data Registry (EDR) has as its registration authority the Environmental Protection Agency (EPA). The stewardship contact appointed by that agency is the final authority for the assignment of names. Other registries will establish their own naming conventions.
- **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 technical report, 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 Subclause B.2.3 of this Annexe.
- **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 registration authority 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).

### B.3.3 Example of a naming convention

An example of a naming convention for the context "Registry," and its adaptation for a specific registration authority is provided in this clause. 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 Country in its name. Properties and 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 English language alphabet.
- **Name Uniqueness.** Names shall be unique within a registration authority for the context Registry.

### B.3.4 Formulating a data element name

The examples used in this technical report are based on a naming convention for name context "registry," established by one registration authority. The example requires that the data element name be constructed to reflect both the logical entity (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 Subclause B.2.3.

The data element names in the following table 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 (Entity)	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 Identifier	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

Object (Entity)	Property (Data Values)	Representation	Qualifier	Resultant Data Element 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.

**B.4 Identification**

**B.4.1 Data element identifier**

Part 5 of ISO/IEC 11179 gives principles for naming and identification of data elements. Each data element registered within a Registration Authority, i.e., an organization authorized to register metadata, is unambiguously identified with a unique identifier. This number takes the form ICD-OI-OPI-OPI\_SRC, where ICD = international code designator, OI = organization identifier, OPI = organization part identifier, and OPI\_SRC = OPI source. At the time a data element is registered into a 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 Annexe A, 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 clause. No administrative data attributes have been assigned to the examples described in the text of this technical report or in the table provided in Annexe A.

**B.4.2 Versioning**

Any administered component in a metadata registry may need to be versioned: value domains, data elements, classification schemes (models, systems, message sets, standards, documents, groups, forms, term sets, and others), conceptual domains, data concepts, object classes, properties, representation classes, and rules.

**B.4.2.1 General rules**

1. In order to ensure versioning is appropriately applied, it cannot be decided by software, but requires interpretation of the business rules by a data analyst.

2. Versions would be incremented only for non-trivial changes (not typos). In some cases, the stewardship contact and the registrar would need to agree on changes. In the case of standards, the working group would have to approve changes.

**B.4.2.2 Value domains**

1. Value meanings and permissible values will be date stamped with begin and end dates for valid values.

2. A change in a permissible value or value meaning would result in adding the new value to the table with an effective (or beginning) date. For example, if the country name for the Belgian Congo is changed to Zaire, Zaire will be added to the file with a begin date, and an end date would be recorded for the date that the Belgian Congo ceased to be a country.

3. Value domains will be reviewed for potential versioning when a new value domain set is published. For example, if the State Department releases a new set of country codes, the new country code values would need to be reflected in the metadata registry. When information about changes to individual values in the domain is available, permissible values and value meanings will be individually updated and date stamped. In the event that change information is unavailable, the entire new domain would be published in the metadata registry and the old domain retained. The new value domains will have a new version number and carry their source information. Each value domain can be linked to the related data element, such as Country Code.

4. Value domain changes would prompt the need to review related data elements to determine whether or not they should be versioned. For example, if the Country Name Value Domain changed, the related data element of Country Name would be reviewed for potential versioning.

### **B.4.2.3 Data elements**

1. Data elements may be versioned based on changes to definition, representation, format, or a related value domain.

2. All changes made to standard data elements require some documentation of authorization. This could be indicated within a text field for each standard.

3. Data elements would be reviewed for versioning based on changes in related groups, documents, applications, or standards.

4. 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 record the earlier versions (6125:1, as well as 6125:2).

### **B.4.2.4 Classification schemes**

1. A classification scheme would be versioned when a new release is entered into the registry.

2. Terms would not be versioned, but definitions would be versioned when a source submitted a revised definition.

3. Changing versions of classification scheme or term definitions would trigger the need to review existing classifications of data elements and other objects in the model.

## **B.5 Conceptual relationships**

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

### **B.5.1 Data element concept**

The data element concept 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 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 by codes (e.g., "Country Name" or "Country Code"). There is 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.

The system software can create a data element concept identifier. It provides unique identification and versioning for data element concepts. It is an identifier that can be used to indicate the domain for translation of data values.

### B.5.2 Conceptual domain

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. A non-enumerated conceptual domain constrains the perception meaning of a data element concept, describing the possible valid values of a non-enumerated set of value meanings, without representation.

Characteristics of conceptual domains include:

- Plural. Whether enumerated or non-enumerated, a conceptual domain encompasses 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 of government, type of government, economic activities) would each 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 clause. Conceptual domain identifiers have been recorded for the examples provided in Annexe A to demonstrate uniqueness and reusability.

### B.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 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 date when a value meaning was entered into a conceptual domain and the date when a value meaning was no longer valid for a conceptual domain are required in a registry.
- Unique Identifier. Each value meaning has a unique identifier (VMID) in a registry. The VMID and the data element unique identifier (IRD1) provide unique identification of an occurrence of a particular data element item. 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.

### B.6 Classification

Classification helps to add information not easily included in definitions, helps to organize the contents of a registry, and helps to provide access by supporting more meaningful queries. Part 2 of ISO/IEC 11179 describes general categories of classification, including:

- Keywords (i.e., controlled word lists), which can be applied to object classes, properties, representations, data elements, and data element concepts.
- Thesaurus terms, which can be associated with data elements and data element concepts.
- Taxonomy and ontology taxa, which may be related to the classified data registration components: object class, property, representation class, and data element concept.

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 Annexe Subclause B.2.3.

A registry might also choose to classify data elements as one or more of the following:

- 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, e.g., altitude, date, facility, industrial, and organization.
- Documents, e.g., all the data elements listed in a document.
- Computer information system, e.g., all the data elements used in a computer system.
- Data Standard, e.g., data elements defined in a data standard.
- Form, e.g., forms that identify data elements for a particular data collection.
- Other sources of data elements.

One of the sources that classify a data element is also considered to be the origin of that data element or value domain in a registry.

### B.7 Quality review

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

#### B.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." The registration authority determines other status changes.

### **B.7.2 Administrative status**

There is no list of values in the ISO/IEC 11179 standard for administrative status; the registration authority establishes these. One registration authority uses the following values for administrative status:

- Awaiting information. The data element is not described by the minimum required attributes.
- In Quality Review. The data element attributes are under review for quality.
- Interim. The data element is designated for use as a standard during final review.
- Final. The data element is a standard.
- Inactive. The data element is not currently active.
- No Further Action. The data element will not receive further review for standardization.
- Proposed for Certified. The data element is complete and has undergone quality review by the Data Registrar for certification.
- Proposed for Standard. The data element has undergone review and is proposed as a standard.
- Proposed to Retire. The data element is no longer used within the registration authority and is proposed for retirement.
- Review for Standard. The data standard is under review to become a standard.
- Superseded. The data element was replaced by another data element.

For this example, the administrative status for a new data element that has a registration status of "Incomplete" is always "Awaiting Information." When all mandatory attributes are complete and the registration status has been changed to "Recorded," the administrative status is updated to "In Quality Review." The registration authority determines all changes in administrative status.

## Annexe C

## Crosswalk of data element component names in technical report to ISO/IEC 11179-3 metamodel, revision 2000-05-34

Optionality is indicated as: Mandatory, Optional, Conditional

Opti on- ality	Metadata Attribute Name in this document	Entity/Attribute Names in Metamodel	Context of Metamodel Entities
	<b>1.Data Element Definition and Permissible Values</b>		
O	Data Element Definition Context	Context.context_name	Data Element is always an Administered Component
M	Description of the Context	Context.context_description	Describes Context for an Administered Component.
M	Data Element Definition	Definition.definition_text	Data Element is always an Administered Component
C <sup>1</sup>	Permissible Values	Value:value_item	Associated with Permissible_Value(s) for Enumerated_Domain(s)
C <sup>1</sup>	PV Begin Date	Permissible_Value. permissible_value_begin_date	Used when the Value Domain has an Enumerated_Domain
O	PV End Date	Permissible_Value. permissible_value_end_date	Used when the Value Domain has an Enumerated_Domain
O	Value Domain Context	Context.context_name	Used when Value Domain is an Administered Component
C <sup>2</sup>	Value Domain Name	Designation.name <i>or</i> Value_Domain.value_domain_name	If Value Domain is not administered, the name must be recorded in the Value_Domain entity
O	Value Domain Definition	Definition.definition_text	Used when Value Domain is an Administered Component
C <sup>3</sup>	VD Description	Non-enumerated_Domain. non-enumerated_domain_description	Associated with Value Domain and required for non-enumerated domains.
C <sup>2</sup>	Value Domain Component Identifier	Administered_Component.identifier	If Value Domain is an Administered Component, the identifier must be recorded in the Administered Component entity.
C <sup>2</sup>	Value Domain Identifier	Value_Domain. value_domain_identifier	If Value Domain is not administered, the identifier must be recorded in the Value_Domain entity.
M	Datatype	Value_Domain.datatype (5/14/00)	
O	Minimum Characters	Value_Domain. minimum_character_quantity (5/14/00)	
O	Maximum Characters	Value_Domain. maximum_character_quantity (5/14/00)	
O	Format	Value_Domain.format (5/14/00)	
O	Unit of Measure	Unit_of_Quantity. unit_of_quantity_name	Value_Domain.value_domain_unit_of_quantity
O	Precision	Unit_of_Quantity.precision	Value_Domain.value_domain_unit_of_quantity
O	VD Origin	Administered_Component.origin	Used when Value Domain is an Administered Component

Opti- on- ality	Metadata Attribute Name in this document	Entity/Attribute Names in Metamodel	Context of Metamodel Entities
O	VD Explanatory Comment	Administered_Component. explanatory_comment	Used when Value Domain is an Administered Component
M	Representation Class	Data_Element. data_element_representation_class	
O	Representation Class Qualifier	Data_Element. representation_class_qualifier	
<b>2</b>	<b>Data Element Name and Identifier</b>		
O	Data Element Name Context	Context.context_name	Data Element is always an Administered Component
M	Data Element Name	Designation.name	Data Element is always an Administered Component
M	Data Element Identifier	Administered_Component.identifier	Data Element is always an Administered Component
<b>3</b>	<b>Other Metadata Attributes</b>		
M	DE Example	Data_Element_Example. data_element_example_item	Every data element must be exemplified by a data element example.
O	DE Origin	Administered_Component.origin	Every Data Element is an Administered Component.
O	DE Explanatory Comment	Administered_Component. explanatory_comment	Every Data Element is an Administered Component.
M	Submitting Organization	Organization.organization_name	
M	Stewardship Contact	Stewardship.stewardship_contact	The stewardship contact is usually represented as an organization or a hot line telephone number.
M	Creation Date	Administered_Component. creation_date	Usually captured as the system date.
<b>4</b>	<b>Data Element Concept and Conceptual Domain</b>		
O	Data Element Concept Context	Context.context_name	Data Element Concept is always an Administered Component.
O	Data Element Concept Name	Designation.name	Data Element Concept is always an Administered Component.
O	Data Element Concept Definition	Definition.definition_text	Data Element Concept is always an Administered Component.
O	Object Class	Data_Element_Concept. data_element_concept_object_class	
O	Object Class Qualifier	Data_Element_Concept. object_class_qualifier	
O	Property	Data_Element_Concept. data_element_concept_property	
O	Property Qualifier	Data_Element_Concept. property_qualifier	
O	Conceptual Domain Context	Context.context_name	Used when the Conceptual Domain is administered.
O	Conceptual Domain Name	Designation.name	Used when the Conceptual Domain is administered.
O	Conceptual Domain Definition	Definition.definition_text	Used when the Conceptual Domain is administered.
C <sup>1</sup>	Enumerated Value Meaning Identifier	Value_Meaning. value_meaning_identifier	Contained in a Conceptual_Domain and used by Permissible_Values for enumerated domains.
C <sup>1</sup>	Enumerated Value Meaning Text	Value_Meaning. value_meaning_description	Contained in a Conceptual_Domain and used by Permissible_Values for enumerated domains

Opti- on- ality	Metadata Attribute Name in this document	Entity/Attribute Names in Metamodel	Context of Metamodel Entities
C <sup>1</sup>	VM Begin Date	Value_Meaning. value_meaning_begin_date	Contained in a Conceptual_Domain for enumerated domains.
O	VM End Date	Value_Meaning. value_meaning_end_date	Contained in a Conceptual_Domain for enumerated domains.
<b>5</b>	<b>DE Classification</b>		
O	Classification Types	Classification_Scheme_Item. csi_type_name	Classification Scheme
C <sup>4</sup>	Examples of Classification Types	Classification_Scheme_Item. csi_value	Classification Scheme
<b>6</b>	<b>Quality Control</b>		
M	Registration Status	Administered_Component. registration_status	Required for all administered components. Always required for Data Element and Data Element Concept.
M	Administrative Status	Administered_Component. administrative_status	Required for all administered components. Always required for Data Element and Data Element Concept.
<b>7</b>	<b>Classification Scheme for Groups</b>		
M	Type of Source/Classification Scheme Type	Classification_Scheme. classification_scheme_type_name	
O	Name/Title	Designation.designation_name	Classification scheme is an administered component.
O	Definition	Definition.definition_text	Classification scheme is an administered component.
M	Organization	Organization.organization_name	
O	Explanatory Comment	Administered_Component. explanatory_comment	Administered component for the classification scheme.
M	Component Identifier	Administered_Component. component_identifier	Administered component for the classification scheme.
M	Registration Status	Administered_Component. registration_status	Administered component for the classification scheme.
M	Administrative Status	Administered_Component. administrative_status	Administered component for the classification scheme.

<sup>1</sup> Mandatory for enumerated domains.

<sup>2</sup> One of the alternative fields is mandatory.

<sup>3</sup> Mandatory for non-enumerated domains.

<sup>4</sup> Mandatory when classification types are recorded.

## 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 2382-1-32:1979-1998 *Information technology - vocabulary*.
- [5] ISO 2788:1986 *Documentation - Guidelines for the establishment of monolingual thesauri*.
- [6] ISO/IEC 3166-1:1997, *Codes for the representation of names of countries and their subdivisions Part 1: Country codes*.
- [7] ISO/IEC 3166-2:1997, *Codes for the representation of names of countries and their subdivisions Part 2: Country subdivision codes*.
- [8] ISO 5964:1985 *Documentation - Guidelines for the establishment of multilingual thesauri*.
- [9] ISO 6709:1983, *Standard representation of latitude, longitude and altitude for geographic point locations*.
- [10] ISO/IEC 7826-1:1994 *Information technology - General structure for the interchange of code values - Part 1: Identification of coding schemes*.
- [11] ISO/IEC 7826-2:1994 *Information technology - General structure for the interchange of code values - Part 2: Registration of coding schemes*.
- [12] ISO 11180:1993 *Postal Addressing*.
- [13] ISO/IEC 11404:1996 *Information technology - Programming languages, their environments and system software interfaces - Language-independent datatypes*.
- [14] ISO 19111, *Spacial referencing by coordinates*.
- [15] SC32 N0147 *Horizontal Issues and Encodable Value Domains in Electronic Commerce*.
- [16] ANSI X3.61-1986 *Representation of Geographic Point Locations for Information Interchange*.
- [17] ANSI X3.285-1998 *Metamodel for the Management of Shareable Data*.
- [18] Firesmith, Donald G., *Object-Oriented Requirements Analysis and Logical Design*, John Wiley and Sons, New York, 1993.
- [19] 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.
- [20] Zachman, John A., "The Framework for Enterprise Architecture: Background, Description and Utility," 1997, <http://www.ozemail.com.au/~visible/papers/zachman3.htm>