ISO/IEC JTC 1/SC 32 N 0262

Date: 1999-05-11

REPLACES: --

ISO/IEC JTC 1/SC 32

Data Management and Interchange

Secretariat: United States of America (ANSI)

Administered by Pacific Northwest National Laboratory on behalf of ANSI

DOCUMENT TYPE	Proposed NP (proposal under review by WG)
TITLE	Proposed NP for Technical Report on Procedures for Achieving Data Registry Content Consistency
SOURCE	U.S. National Body
PROJECT NUMBER	
STATUS	For review at the Matsue WG 2 meeting
REFERENCES	
ACTION ID.	COM
REQUESTED ACTION	For review at the Matsue WG 2 meeting
DUE DATE	
Number of Pages	66
LANGUAGE USED	English
DISTRIBUTION	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

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

Data Element Registry Content Technical Report

New Work Item - Draft for Discussion

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

ISO/IEC 11179, Part 3 and extensions proposed to it in a related New Work Item, have concentrated on the basic attributes of data elements. Much of the work done to date is for the "abstract" level, not "application" level data elements. Well-formed data elements and their domains can be recorded in a metadata registry as "models" for potential reuse. Additional attributes may be required to record essential facts about how a data element is used in an application. Some potential attributes include "purpose for which data is collected", "statistical methodology used in data collection" and other potential data quality attributes. What other attributes are required? There is a need to address the attributes that should be documented at the application level. This will make use of the standard's extensibility, since all application level attributes cannot be established in advance.

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

When a standard data element is created, its value domain is specified. A number of related application elements may be created that are related to this standard element by particularizing the value domain in different ways. The resulting data element's domains might be further constrained to produce still more data elements. How are these relationships to be recorded?

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

ISO/IEC NWI xxx

Information Technology

Procedures for Achieving Data Registry Content Consistency

Working Paper
Draft 2.1

May 1999

Procedures for Achieving
Data Registry Content Consistency

Contents

Foreword

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

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

Introduction

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

ISO/IEC 11179, Part 3 and extensions proposed to it in a related New Work Item, have concentrated on the basic attributes of data elements. Much of the work done to date is for the "abstract" level, not "application" level data elements. Well-formed data elements and their domains can be recorded in a metadata registry as "models" for potential reuse. Additional attributes may be required to record essential facts about how a data element is used in an application. Some potential attributes include "purpose for which data is collected", "statistical methodology used in data collection" and other potential data quality attributes.

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

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

1 Scope

1.1 Background

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

This Technical Report is based on the American National Standard Institute (ANSI) X3.285:1999 Standard, *Metamodel for the Management of Shareable Data.* The standard specifies the structure of a data registry in the form of a conceptual model. The conceptual model is more abstract than a logical data model in that it 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 data registry contains the metadata that is necessary to clearly describe, inventory, analyze, and classify data. It provides an understanding of the meaning, representation, and identification of a unit of data. The ANSI X3.285 standard "outlines the information elements associated with a data element concept that need to be available for determining the meaning of a data element to be shared between systems. The standard is a complement to the six-part International Organization for Standardization/International Electrotechnical Commission (ISO/IEC) 11179 standard that describes the organization of a data registry for managing the semantics of data elements in data systems."[1]

1.2 Purpose

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

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

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

2 Normative References

The following standards contain provisions, which, through reference in the text, constitute provisions for this document. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this document are encouraged to investigate the possibility of applying the most recent editions of standards indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

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

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

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

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

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

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

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

3 Definitions

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

3.1 attribute: A characteristic of an object or entity.

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

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

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

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

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

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

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

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

3.10 data value: An element of a value domain.

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

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

3.13 identifier: See data identifier.

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

3.15 metadata: Data that defines and describes other data.

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

3.17 object class: A set of ideas, abstractions, or things in the real world that can be identified with explicit boundaries and meaning and whose properties and behavior follow the same rules.

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

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

3.20 representation class: A classification of types of representations.

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

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

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

4 Levels of Abstraction

This clause presents a conceptual framework for structuring data elements in a registry and mapping existing data elements to registry components.

Data elements are ideally the result of a process of development, involving several levels of abstraction. Levels progress from the most general (conceptual) to the most specific (physical). The objects at each level are called data element components. Using the Zachman Framework, for instance, the highest levels of definition are contained in the business view; development progresses to the implemented system level.

After the conceptual components are developed by a process of specification from the highest conceptual level, a representation term is assigned which may in turn be derived from a structure set or process. Components are envisioned as a set of building blocks that can be assembled into data elements, and serve to ensure that the end product, the total set of data elements, is as discrete and complete as possible.

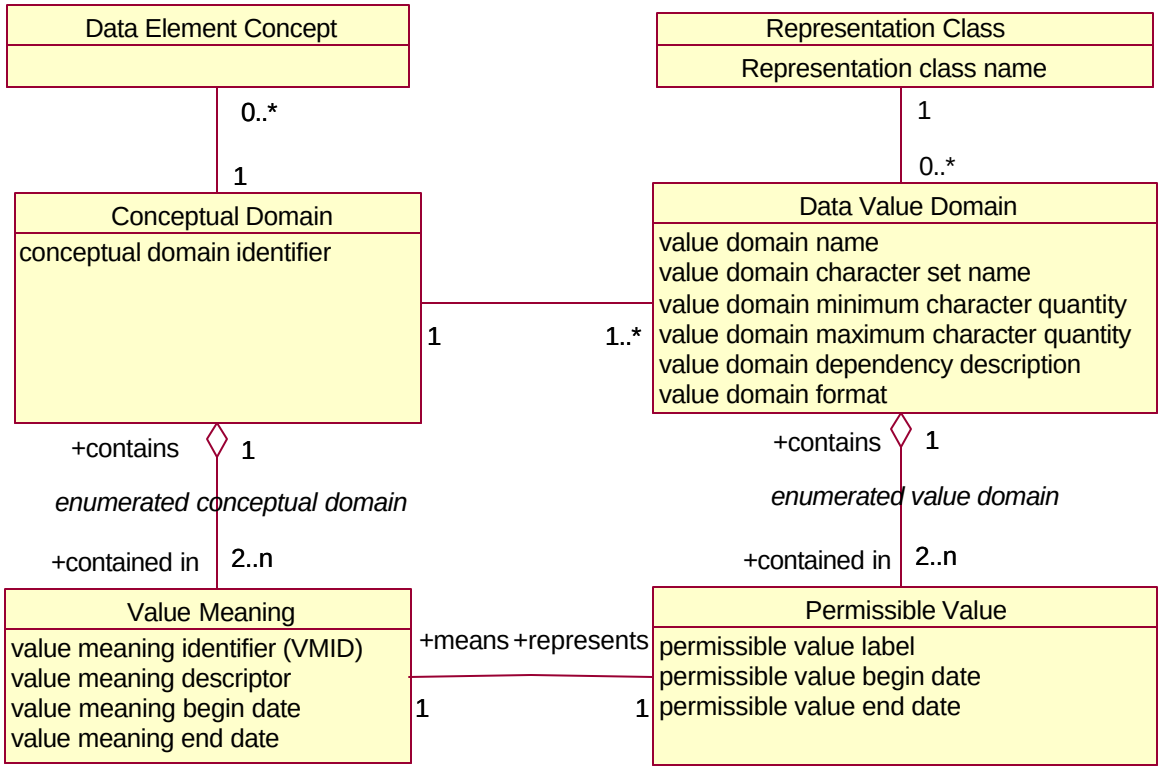
4.1 Developing Levels of Abstraction

Data element development begins at the conceptual level (See Figure 1). At this stage, a set of concepts exists as entities or objects (called object classes), which, with the assignment of properties, become data element concepts (DECs). An object class represents an idea, abstraction or thing in the real world, such as tree or country. A property is something that describes all objects in the class, such as height or identifier.

From the examples above, we can form the DECs tree height and country identifier. DECs also contain conceptual domains, which are composed of value meanings. These value meanings are defined but do not have a specific form of representation (Figure 2).

The next step in forming data elements takes place at the logical level. A complete logical data element must include a form of representation for the values in its data value domain. For example, name, code, and measure can be applied to the DECs above to produce tree height measure, country identifier name and country identifier code.

Figure 1. Components of a Data Element



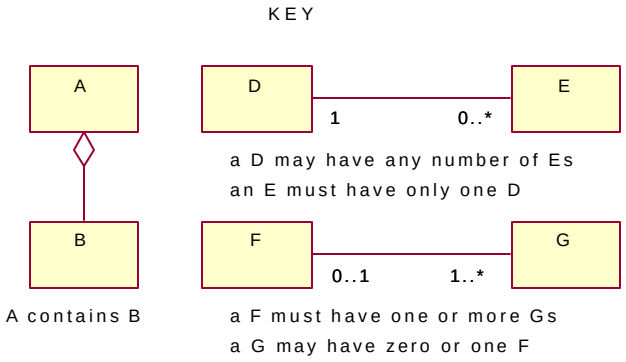
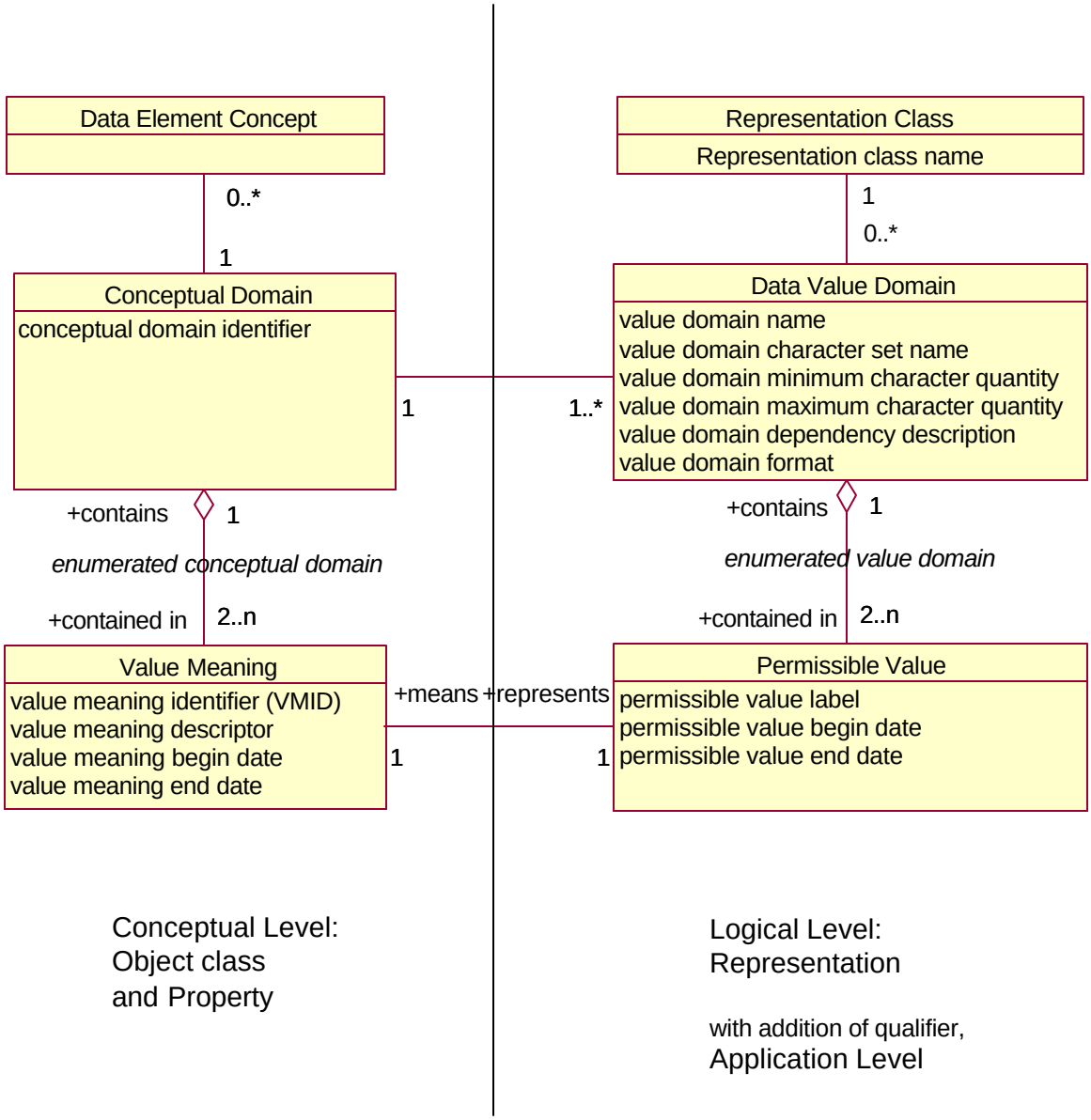


Figure 2. Levels of Abstraction



Some logical data elements can be considered generic elements. These are data elements that have a well-established data value domain and are recognized at the organizational level or above as useful and shared among several systems. Country name and country code are both potential candidates for designation as generic elements. ISO standard 3166, *Codes for the representation of names of countries*, presents a well-established reference list of country names and codes.

Note that this is the highest level at which true data elements, by the definition of ISO 11179, appear: they have an object class, a property, and a representation.

The next level of data element development is the application level. Typically, a data element will be customized to an application by subsetting its data value domain or narrowing the definition (or both) to include only those values of interest to the application. For example, if an application of Country name were to list all the countries a certain organization had trading agreements with, the application data element would be called Trading partner country name. The data value domain would consist of a subset of countries listed in ISO 3166. Note that the qualifier term trading partner is itself an object class. This relationship could be expressed in a hierarchical relationship in the data model.

The last type of data element is the physical data element. These are the elements which actually appear in the database table column headers, file descriptions, EDI transaction file layouts, etc.

Development of data elements and their placement in the correct levels is one of the most important ways to achieve consistency of content among data registries. The concepts introduced here will be described in detail in clause 5.

5.0 REGISTER A DATA ELEMENT

Registration of a data element in a data element registry requires that certain characteristics of the data element are recorded to clearly describe and define it [2]. These characteristics are stored as attributes of the data element. The Metamodel for the Management of Shareable Data, American National Standard for Information Technology, ANSI X3.285, specifies attributes with procedures for formulating those attributes in a data registry. This document gives examples that demonstrate the population of a data registry for a specific registration authority. It includes attributes and procedures that are fully defined by the metamodel, as well as those where the registration authority establishes its own procedures. Each attribute of a data element is described in the following paragraphs, with rules for formulating the attributes and examples. Attributes specific to a data element are:

X Data Element Representation.
- Representation Class.
- Value Domain.
  -- Label (Identifier)
  -- Value Domain Name.
  -- Definition.
  -- Datatype (Character Set).
  -- Domain type.
  -- Maximum and minimum values.
  -- Format.
  -- Unit of measure.
  -- Precision.
  -- Source.

X Data element definition with context.

X Data element name with context.

X Data element identifier and version number.

X Administrative attributes:
- Registration and administrative statuses.
- Submitting organization.
- Data steward.

X Data element concept name, definition, and identifier.

X Conceptual domain name and definition.

---

[2] A Registry can be used to record information about data elements ranging from carefully crafted data standards to those found in applications. The amount and quality of metadata information available can vary from good, complete information to poor, incomplete information. This document is intended to describe the population of a Registry with data elements for which good quality, consistent metadata can be created.

X Value meaning names and definitions.
X Permissible values.
X Example.
X Reference Document (i.e., Origin).

In many cases where a data element is submitted for registration, only limited information (e.g., a name and definition) is provided. Other attributes must be determined based on an understanding of the underlying data values and concepts that are implied by those facts. These are most commonly registered by means of a ABottom Up≅ registration procedure, where the data element is attributed prior to defining the conceptual information about the data element. The examples provided illustrate the logical order for formulating attributes about a data element, based on a ABottom Up≅ procedure.

Registration of three types of data elements is provided in this section, including data elements from:

X An international standard with an enumerated domain.
X An international standard with a non-enumerated domain.
X An application.

Registration of data element groups is also included in this section.
Annex F contains a table that includes the data element attributes for the examples provided in the remainder of this section. Administrative attributes, which are specific to a registration authority, are not listed in the table.

An example of a ATop Down≅ registration, where registration begins with identification of the conceptual domain, is provided in informative Annex G with an example of registration of data elements about biological organisms.

5.1 General Procedures for Registration of a Data Element

The following paragraphs provide general procedures that are used for registration of a data element. The procedures are presented in a logical order for analyzing and formulating attributes for a data element.

5.1.1 Data Element Representation

The first thing to consider when registering a data element is how the data element is to be represented. Data element representation is a data element component consisting of a value domain and a representation class. These are described in the following paragraphs.

5.1.1.1 Representation Class

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

X AmountBthe sum total of two or more quantities; an aggregate.

X CodeBa symbol used to represent something.

X GraphicBdiagrams, graphs, mathematical curves, or the like.

X IconBa sign or representation that stands for its object by virtue of a resemblance or analogy to it.

X MeasureBthe extent, dimensions, quantity, etc. of something ascertained by comparison with a standard.

X NameBa word or combination of words by which a person, place, object, or thought is known.

X NumberBa numeral or group of numerals.

X PictureBa visual representation of a person, object, or scene.

X QuantityBthe property of magnitude of something.

X TextBa unit of connected speech or writing often composed of one or more sentences that form a cohesive whole.

5.1.1.2 Value Domain

The value domain is the set of permissible values that will be stored in the data element. After identifying the representation class, the value domain is formulated, based on an understanding of the data content. A data element is associated with only one value domain, and the definition of the value domain describes all of the data values that are included in that domain.
Value domains can have the following attributes, not all of which are in the standard:

X Label. The record identifier that represents the value domain.

X Name. The name by which a value domain is known. The name should be plural, since a value domain encompasses all values that are included in the domain (e.g., Short English-Language Country Names).

X Definition. A description of the data content of the value domain. The definition should be plural, since a value domain encompasses all the values that are included in the domain (e.g., All short, English-language names of countries assigned by ISO 3166).

X Character Set (i.e., Datatype). The datatype is the natural language character set used by a value domain. 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 an example of the use of character sets, one registration authority has established that datatypes by designated as alphanumeric or numeric. An alphanumeric datatype is composed of either alphabetic characters, numerals, or both. A numeric datatype is composed of numerals. In general, values that are intended to be sorted, whether numerals or alphabetic characters, are described as "alphanumeric." Only

numbers that are used in calculations are given the datatype of "numeric." The character set for "date" has been identified as "date," and whole numbers as "integers." When creating metadata for more complex datatypes (e.g., arrays and bit strings), ISO 11404 provides guidance on datatypes.

X 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 document. Country names are a fixed list of countries, maintained by international standards; therefore, the domain type is enumerated.

Non-enumerated domains have an unspecified set of values. The values, however, must fall within the scope of the definition. Latitude Measures are not restricted to a fixed list. Therefore, the domain type is non-enumerated. Non-enumerated domains are characterized by the process used to determine the domain values. They must be characterized by one of the following:

- Procedure. Measurements and quantities are determined by procedure (e.g., they are calculated, measured, or generated).

- Reference. Telephone numbers and facility names are determined by reference (e.g., they can be validated in some type of directory).

- Range. Percentages and temperatures are examples of range determinations. Maximum and minimum values are always required for range determinations. Examples: 1-100% and 32-212^{o}F.

X Maximum and minimum field lengths.

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

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

X Format. The standard does not specify a format. 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.

X 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 indicates the unit of measure for all data values for the value domain.

X 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).

X Permissible values. For value domains that are enumerated, permissible values must be entered into the registry. The permissible values for country identification in "Short, English-Language Country Name" will be those names that are listed in the ISO 3166 standard for that category.

The permissible values for a value domain are associated with the value meanings (i.e., the names and definitions that are included in the domain of possible values). In this section of the document, only the formulation of value domain has been described. The entry of value meanings and permissible values (i.e., the codes and names that are valid for a data element) is described later in Sections 5.1.8 and 5.1.9.

X Source. The metamodel specifies a reference document for a data element, but does not require that source of data values be recorded. A registration authority might require that enumerated domains be identified with an authoritative source. Example: Short, English language country names, assigned by ISO 3166.

Non-enumerated domains which are determined by reference should also record the source of the reference (e.g., Dun & Bradstreet or telephone directory).

Examples of value domain identifiers (i.e., labels) have been assigned to the examples provided in Annex F to demonstrate uniqueness and reusability.

5.1.2 Data Element Definition

The purpose of a data element definition is to define a data element with words or phrases that describe, explain, or make definite and clear its meaning. Precise and unambiguous data element definitions are one of the most critical aspects of ensuring data shareability. The value domain, described in Section 5.1.1, defines the complete set of values that can be contained in a data element. The data element definition defines one instance of those domain values.

ISO/IEC 11179-4 provides the standard for formulating data element definitions. There are mandatory rules, to which all data element definitions must comply, and there are guidelines which should be followed in formulating a definition. The standard does not specify syntactical requirements (i.e., word order and structure), which may be established by the registration authority. The rules and guidelines applicable to the Registry Definition (i.e., the unique definition that has been assigned to the data element for registration in a metadata registry) follow. A syntax that has been adopted by one registration authority is also included in this section.

5.1.2.1 Mandatory Rules

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

X Unique (within any data dictionary in which it appears).
X Singular.
X State what the concept is, not only what it is not (i.e., never exclusively in the negative).
X Descriptive phrase or sentence.
X Contain only commonly used abbreviations.
X 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.

5.1.2.1.1 Uniqueness

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

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 an event started.
Sample Collection Start Date: The date when an event started.

5.1.2.1.2 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.

5.1.2.1.3 State the Concept; Not Only its Negative

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

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

Poor: Not the long name of a country.

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

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

5.1.2.1.4 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.

5.1.2.1.5 Contain Only Commonly Used Abbreviations

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

Example 1:
Good: The code that represents the economic activity of a company as specified by the Standard Industrial Classification (SIC) of Establishments.

Poor: The SIC code for a company.

Example 2:
Good: The code that represents the unit for measuring the mass per unit (m.p.u.) volume.

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

5.1.2.1.6 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.

5.1.2.2 Guidelines for Definitions

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

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

Examples of these guidelines are provided in the following paragraphs.

5.1.2.2.1 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.

5.1.2.2.2 Precise and Unambiguous

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

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

Poor: The code that represents a country.

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

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

Example 2:
Other examples of **good** definitions that clearly distinguish between similar data elements are:

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

X The complete, long name that identifies a country.

5.1.2.2.3 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."

5.1.2.2.4 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.

5.1.2.2.5 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.

5.1.2.2.6 Avoid Circular Reasoning

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

Poor: A code number assigned to an object.

Poor: An object identified by a code number.

5.1.2.2.7 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 ConsistencyB 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 names of the methods 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 ConsistencyB 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.

5.1.2.3 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:

X 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.

X Begin the phrase 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 ...."

5.1.2.4 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:

X DefineBTo set forth the meaning of a word or phrase.

X DepictBTo represent by, or as if by painting, or to characterize by words with vividness of detail.

X DescribeBTo convey in words the appearance, nature, or attributes of something.

X DesignateBTo select or nominate for a purpose.

X IdentifyBTo recognize or establish as being a particular person or thing; to verify the identity of something.

X IndicateBTo 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.

5.1.3 Data Element Name

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

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

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. Detailed information about formulating a data element name is provided in Section 5.1.3.1, based on ISO/IEC 11179-5.

Multiple names may be appropriate for a data element based on the context for which the data element name has been derived. Contexts for names are described in Section 5.1.3.2. Each data registry establishes its own naming convention. Suggestions for a naming convention are provided in Section 5.1.3.3

5.1.3.1 Formulating a Data Element Name

The examples used in this document are based on a naming convention for registry name context, established by one registration authority. The example requires that the data element name be constructed to reflect both the logical entity which includes the data element (i.e., the object) and the attribute which identifies the type of data value to be contained in the data element (i.e., the property). Although the entity is not always required to be a term in the name, the attribute (i.e., type of data value) is a requirement. For the registration authority used in this example, data element name would always include the representation class term, such as name, measure, amount, number, code, quantity, text, or others, as defined in Section 5.1.1.

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

Object	Property (Data Values)	Representation	Qualifier	Resultant Data Element Name
[1]	Country Identifier	Name		Country Name
Address	Country Identifier	Name	Mailing	Mailing Address Country Name
Address	Country Identifier	Code	Geographic	Geographic Address Country Code
Address	Person Name	Name[2]	Mailing	Mailing Address Person Name
Facility	Legal Name	Name[2]		Facility Legal Name
[1]	Latitude	Measure		Latitude Measure
Location	Latitude	Measure	Facility	Facility Location Latitude Measure
Location	Latitude	Measure	Stack	Stack Location Latitude Measure
Geographic Coordinates[3]	Collection Method	Code	Horizontal	Horizontal Collection Method Code
Geographic Coordinates[3]	Collection Method	Code	Vertical	Vertical Collection Method Code

[1] No value for entity is necessary for the ISO and ANSI standard data elements. These standard data elements are the basis for many data elements that use the same set of value meanings.
[2] "Name Name" is redundant, so only one "Name" is used in the data element name.
[3] "Geographic Coordinates" is an implied entity not included in the data element name.

Exhibit 5.1. Data Element Names

5.1.3.2 Name Context

Data element names, according to ISO/IEC 11179-5, are assigned within a registry at the discretion of the Registration Authority. Each data element must have one name, and more than one name can be assigned depending on the context in which the data element is used. Context names are not listed in the standard. Examples of name contexts that might be used for a registration authority include:

X Legacy - a name that has been used in the past.

X Standard - a name that has been used in a standard (e.g., ANSI, ISO, or other standard).

X Short Abbreviation - a name that is used in a computer system.

X <*source system name*> - the name that is used by the source that submitted the data element for registration.

X Registry - the unique name that has been assigned to the data element for registration by a registration authority.

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

Registry: Vertical MeasureBThe vertical measure, in meters, of the measured point, above or below a reference point.

Legacy: Vertical MeasureBThe measure of elevation (i.e., the altitude), in meters, above or below a reference datum.

Standard: AltitudeBThe 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 AltitudeBThe 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: The registry includes an attribute for "Unit of Measure," which is the appropriate attribute to indicate the unit by which the data value is to be recorded. In the ANSI X3.61-1986 standard, however, unit of measure has been included in the definition, so it has been replicated in this example. The registry also includes an attribute for the precision required for recording the data value.

5.1.3.3 Naming Convention

Each registration authority should establish its own naming convention to be used for a registry. Where data element names are provided from other sources, the naming convention may not be known (e.g., the names assigned to data elements in an application software system). Each data element context in a registry should have its own naming convention. The naming convention shall be constructed according to ISO/IEC 11179-5 naming conventions, as explained in the following paragraphs.

X **The Scope of the Naming Convention**. The scope of the naming convention for this document is limited to the Registry Name. For example, a data element might have the name Regulation Abstract Text as a Registry Name and the name Abstract in another context. The conventions used for names in contexts other than for the Registry Name may not be known to the registration authority.

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

X **Semantic Rules for Source and Content of Terms**. Semantic rules enable meaning to be conveyed. Each registry shall develop guidelines governing 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. In the data element name, Employee Last Name (Name), Employee could come from the object class and Last Name could be the property terms. The word Name given in parenthesis is the representation term, which is not used because it would be redundant.

X **Syntactic Rules for Word Order**. Syntactic principles specify the arrangement of components within a name. The specific syntactic rules for a registry should be developed by the registration authority. In the examples in this document, the only convention for syntax is to include the representation class term as the last term in the name, as in Regulation Abstract *Text*. Representation class terms are defined in Section 5.1.1.

X **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 are part of the design of a registry. A registration authority might choose to establish controlled, well defined word lists for formulating a name.

X **Name Uniqueness**. Each registration authority determines whether a name 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).

5.1.3.4 Example of a Naming Convention

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

X Scope. The scope of this example naming convention is registry name.

X Authority. The authority for this example is the U.S. Environmental Protection Agency for its Environmental Data Registry.

X Semantic Rules. Names shall include an object and a property. 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.

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

X Name Uniqueness. Names shall be unique within a registration authority.

5.1.4 Administrative Attributes

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. An example of the type of administrative attributes that might be used for a registration authority is:

X Registration status. The standard values for registration status include the following:

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

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

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

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

X Administrative status. There is no list of values in the standard for administrative status. 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." Other administrative status changes are determined by the Registration Authority.

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

X Data steward. The data steward is the individual who has been assigned by a submitting organization to be responsible for authorizing and maintaining one or more data elements.

No administrative data attributes have been assigned to the examples described in the text of this document or in the table provided in Annex F.

5.1.5 Data Element Identifier and Version

Each data element is required to have a unique data identifier within the register of a Registration Authority. The combination of Registration Authority Identifier (RAI), Data Identifier (DI), and Version Identifier (VI) shall constitute the International Registration Data Identifier (IRDI).

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.

For the examples listed in Annex F, DI and VI have been recorded to demonstrate uniqueness.

5.1.6 Data Element Concept

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

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

X IdentifierBSomething that represents to be, regards, or treats as the same or identical.

X LabelBA short word or phrase descriptive of a person, group, or intellectual movement, or indicating that what follows belongs in a particular category or classification.

X TagBA descriptive word or phrase applied to a person, group, organization, etc., as a label or means of identification or epithet.

X IndicatorBAnything that serves to point out or direct attention to, as of a measuring device.

X DiscriminatorBA distinction that differentiates one from another.

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

X Singular. Each data element concept represents only one concept.

X 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.

X 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.

X Can be associated with multiple data elements, each with its own representation and value domain.

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

X Can be associated with only one conceptual domain.

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

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

Data element concept identifiers have been assigned to the examples provided in Annex F to demonstrate uniqueness and reusability.

5.1.7 Conceptual Domain

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

Characteristics of conceptual domains include:

X Plural. Whether enumerated or non-enumerated, a conceptual domain includes the entire body of information that might be included as meanings of the data values in a particular data element for a particular concept. Therefore, the name and definition are always described as plural.

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

- Lacking representation. The definitionidentifies 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."

X Conceptual domains can be, and often are, associated with more than one data element concept. Data element concepts that "Countries of the World" could be associated with include, but are not limited to:

- Address Country Identifier.
- North American Country Identifier.
- NATO Country Identifier.
- Geographic Country Identifier.

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

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

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

5.1.8 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:

X Cannot be a representation. The name and definition do not contain representation class terms such as name, number, text, code, or other representation terms.

X Must be associated with at least one conceptual domain.

X Can be associated with more than one conceptual domain.

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

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

X Begin and End Dates. The dates when a value meaning was entered into a conceptual domain and when a value meaning was no longer valid for a conceptual domain are required in a data registry.

X Unique Identifier. Each value meaning has a unique identifier (VMID) in a registry. The VMID and the data element unique identifier (IRDI) provide unique identification of a particular data element item occurrence. This combination of identifiers is valuable for data transfer. No identifiers have been assigned to the examples provided in Annex F of this document.

X 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.

5.1.9 Permissible Values

Permissible values are the exact names, codes, and text that can be stored in a data field in an information management system. The permissible values for the data element "Short English-Language Country Name" are the actual names listed in the ISO 3166 standard for names of countries, where those names and definitions have value meanings associated with the conceptual domain "Countries of the World." The permissible values are entered in association with the value domain, "Short, English-Language Country Names," defined as "All short, English-language country names of the world."

A different data element (i.e., Country Numeric Codes) with the value domain of "3-character numeric codes that represent countries of the world" will be associated with permissible values that are the listed 3-character numeric codes for those same countries, associated with the same value meanings.

For data elements with non-enumerated value domains, the permissible values are not explicitly defined. The values that are to be stored in the data field, however, must conform to the definition of the value domain.

5.1.10 Example

Each data element should have 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 also 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:

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

Example for "Country Name" B 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" B 036 (Australia)

X 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" B 87.123456
Example for "Location Comments Text" B The coordinates reference the flag pole in the North parking lot of the installation. This location is near the center of the facility.

5.1.11 Reference Document

A data element can be associated with any kind of source, including a document, standard, system, group, partner, or message set. One source, as a minimum, must be associated with a data element to indicate the origin of information about the data element. Rules for registering source information are outside the scope of this document.

One registration authority records the origin for every data element in the registry. The origin is intended to identify the source of information about the data element that was the reason for its entry into the data registry. That source is then marked appropriately to indicate that it is the source of the data element.

For the examples in this document, only the name of the origin has been provided. For the Short English-Language Country Name, ISO 3166 is the origin.

5.1.12 Note/Description

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

5.2 Examples of Registration of Data Elements

Registration of three types of data elements is provided in this section, including data elements from:

X An international standard with an enumerated domain.
X An international standard with a non-enumerated domain.
X An application.

Each of the examples follows the logical order for Bottom Up registration that was described in Section 5.1.

5.2.1 International Standard B Enumerated Domain

The International Organization for Standardization (ISO) 3166:1997 standard is used as a source of data elements to demonstrate the registration of data elements with enumerated domains from an international standard. ISO 3166 includes the following domains: short country name in English, 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 official country name in French. The following paragraphs are presented in the logical order for formulating attributes for a standard, enumerated data element, using the short English-language country name as the example.

5.2.1.1 Data Element Representation

For the ISO 3166 data element for short, English-language country name, the representation class is "name."

The scope of the permissible values for the data element includes all countries. The name of the value domain, therefore, could be "Short English-Language Country Names," and the definition "All short, English-language names of all countries."

Other domain attributes for this example include:

X Character Set. The character set for Short English-Language Country Names is "alphanumeric."

X Domain Type. Country names are a fixed list of countries, maintained by international standards; therefore, the domain type is "enumerated."

X Maximum and minimum field lengths. 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).

X Format. The format selected by the registration authority for this example is A(44) to accommodate the longest of the English-language short names.

X Permissible values. The permissible values for country identification in "Short, English-Language Country Name" are those names that are listed in the ISO 3166 standard for that category. These are associated with the appropriate value meanings for country identifiers.

X Source. The source of short, English language country names is ISO 3166.

5.2.1.2 Data Element Definition

Using the mandatory rules and guidelines described in Section 5.1.2, the data element definition formulated for the short, English-language country name in ISO 3166 is "The short name of a country represented in the English language."

5.2.1.3 Data Element Name

The data element name for Registry Name should give some indication of the data values to be contained in the value domain (i.e., the property) associated with the data element. It should also include the name of the representation class, in this example "Name." For this standard data element is it necessary to qualify the "Country Name," since there are four value domains of country names in the ISO 3166 standard. For this particular example, the qualifiers are "short" and English-language."

The name that has been formulated for this data element is "Short English-Language Country Name."

5.2.1.4 Data Element Concept

Identification of the data element concept, as described in Section 5.1, is based on the data element name and definition, without the representation. The concept represented by the data element "Short English-Language Country Name" is "Country Identifier, defined as "An identifier for a primary geopolitical entity of the world." 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.

5.2.1.5 Conceptual Domain and Value Meanings

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

5.2.1.6 Permissible Values

The permissible values for this data element, the ISO 3166 short English country name, are those same names that are listed in the ISO standard under that title. The permissible values range from 4-character names to 44-character names, as defined in the value domain. Each permissible value is associated with a valid value meaning that provides meaning to the permissible value. Each permissible value is also entered in the registry with its begin date (i.e., the date when that permissible value became valid for that value meaning). End dates will also be entered, when the permissible value for a value meaning becomes invalid.

5.2.1.7 Example

An example of a valid value for this short name of a country, in the English language, is "Denmark."

5.2.1.8 Origin

The origin of information about this example data element is "ISO 3166."

5.2.1.9 Other Codes and Names from ISO 3166

Other codes, official English names, and French names (both official and short) from ISO 3166 are registered with their individual value domains, representation, data element definitions, and data element names. All of the data elements associated with ISO 3166 will share the same data element concept (i.e., Country Identifier, defined as "An identifier for a primary geopolitical entity of the world.") and the same conceptual domain (i.e., Countries of the World, defined as "The primary geopolitical entities of the world."). All of the ISO 3166 data elements will share the same value meanings. They will, however, have different permissible values associated with the value meanings, depending upon the data element, its representation, and its value domain. The table in Annex F provides examples of the metadata associated with three data elements from the ISO 3166 standard (i.e., Short English-Language Country Name, Official English-Language Country Name, and Country Numeric Code).

5.2.2 International Standard with NonBEnumerated Domain

The ISO 6709 international standard, *Standard representation of latitude, longitude and altitude for geographic point locations* has been selected as the source for an example of a data element with a non-enumerated domain. Latitude is the example that is used in this document for registering a non-enumerated data element. The standard includes two units of measure for latitude: degrees and sexagesimal (i.e., degrees, minutes, and seconds). Both data elements are included as examples in the table in Annex F of this document.
The methodology for registering latitude is the same as was described for registering the name of a country in Section 5.1. First the representation and value domain are described, followed by data element definition, data element name, and other attributes. These descriptions are in this section.

5.2.2.1 Data Element Representation

The value domain for latitude includes "all measures of the distance of an angle north or south of the equator." For this example, the representation class is defined as "measure." Note in the example in Annex F that the value domains for the different latitude formats have different units of measure, and are therefore different data elements. The units of measure for the two data elements are degree and sexagesimal. The value domains are then differentiated by the form of measurement (i.e., degree and sexagesimal).

The domain type is "non-enumerated" for both, and the determinant type is "range," indicating that the data values must fall within a given range. Latitude degrees always fall within the range of 0-90 degrees. Minutes and seconds fall within the range of 0-60 degrees, described as sexagesimal. The two data elements vary in edit format and minimum number of characters, as noted in Annex F. Both require that the number of decimals recorded indicate the precision of the data value.

For latitude measured in degrees, the range of format is +/-DD.DD to +/-DD.DDDDD or a minimum number of 6 characters and a maximum number of 9 characters. For latitude measured in degrees, minutes, and seconds, the range of format is +/-DDMMSS to +/-DDMMSS.S or a range of 7-9 characters. The datatype is listed as alphanumeric for both, because data transfer requires the +/- directional symbols and the decimal point to be explicitly included.

5.2.2.2 Data Element Definition

The data element definition is formulated according to the rules and guidelines described in Section 5.1.2, based on ISO/IEC 11179-4. The rules require that a data element definition be unique within the registry, so the unit of measure has been included in both of the definitions: "The measure in degrees of the angular distance of a position on earth on a meridian north or south of the equator" and "The sexagesimal measure of the distance of a position on earth on a meridian north or south of the equator."

5.2.2.3 Data Element Name

There is no requirement in the standard that data element names be unique in a registry. It is advisable, however, to use a qualifier in the data element name that differentiates between data elements that might otherwise have the same name. The guidelines for formulating a name according to ISO/IEC 11179-5 are described in Section 5.1.3.

For this example, registering non-enumerated data elements from an international standard, the names of the latitude data elements carry the qualifiers "degree" and "sexagesimal" as discriminators. The names that have been derived for the latitude data elements are "Latitude Degrees Measure" and "Latitude Sexagesimal Measure."

5.2.2.4 Data Element Concept

The methodology to be used for deriving a data element concept is described in Section 5.1.6 of this document. A data element concept is the data element without representation. We have indicated previously that latitude is a distance measure, where measure is its representation. The data element concept for both latitude data elements 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 representation class term "measure." The concept of latitude is the measure of a distance. Therefore, it is appropriate in this instance to use the term measure, even though representation is not appropriate for a data element concept. The different formats represented by the units of measure for latitude can be considered to express representation. The two latitude data elements from ANSI 6709 are translatable at the concept level, based on their unit of measure representations.

The data element concept identifier listed on the table in Annex F indicates that the data element concept is shared by the two latitude data elements.

5.2.2.5 Conceptual Domain and Value Meanings

A conceptual domain, as described in Section 5.1.7, is a collection of value meanings. This collection must be identified with a name and a definition. The latitude is one of the horizontal coordinates that fix a position on the surface of the earth either north or south of the equator. For this example, the name of the conceptual domain for both latitude data elements is "Latitude Coordinates" with the definition "The coordinates that indicate the distance measurement of points north or south of the equator."

Note: Longitude, which is the horizontal coordinate that fixes a position east or west of the prime meridian, requires its own conceptual domain, even though the concept and the value meanings are similar.

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

The conceptual domain identifier listed on the table in Annex F indicates that the conceptual domain is shared by the two latitude data elements.

5.2.2.6 Permissible Values

There are no stored permissible values in a registry for non-enumerated domains. The values that are permissible for either of the ISO 6709 latitude data elements are those values that conform to the value domain for datatype, format, unit of measure, and precision. Latitude values that are measured in degrees must conform to the format +/-DD.DD to +/-DD.DDDDD. Latitude values that are measured as degrees, minutes, and seconds must conform to the format +/-DDMMSS to +/-DDMMSS.S. The precision of the value is indicated by the number of decimal places recorded.

5.2.2.7 Example

An example, as described in Section 5.1.10, must be one of the permissible values for enumerated value domains or must conform to the value domain attributes and data element definition for non-enumerated domains. Examples of latitude measured in degrees and indicating precision are +67.85 and +67.84623. Examples of latitude measured in degrees, minutes, and seconds and indicating precision are +674532 and +674531.8.

5.2.2.8 Origin

The origin of a data element is the standard, document, application system, document or other source where it was identified for registration. The source of the latitude data elements used as examples in this document is the international standard ISO 6709.

5.2.3 Application Data Element

An application data element might be described as a data element that is used in a computer system application. It is associated with an entity and might be identified with a qualifier. The country name attribute in the mailing address

entity has been selected from an information management system that contains data about facilities (i.e., the Facility Data System). This data element was selected to illustrate the relationship between an application data element and a standard data element with the same data values. It also illustrates how a well defined data element might differ from one that is identified from a computer application system. The methodology is the same as that described in Sections 5.1.

5.2.3.1 Value Domain and Representation Class

The United States Postal Service mailing address standard requires that the country name be included as the last line of a mail piece. The address standard further specifies that each address line be limited to 30 characters, including alphabetic and numeric characters. A mail piece might be delivered to any country in the world. Therefore, names of all countries must be included in the value domain. These constraints also indicate that the appropriate value domain for country name to be used in a mailing address, should be the short name from the ISO 3166 standard. All value domain information for this application data element (i.e., country name used in a mailing address) is the same as for the ISO standard Short English-Language Country Name, described in Section 5.1.

5.2.3.2 Data Element Definition

The definition for the country name attribute in the mailing address entity includes the concepts for country identifier, mailing address, and representation. The rules and guidelines described in Section 5.1.2 are used to formulate the data element definition as "The name of the country where a mail piece is delivered."

5.2.3.3 Data Element Name

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

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

Application data elements might be registered because they exist in a computer application system. 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. 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. The entity in which it is stored gives information, even though the context definition "The name of a country" might not convey that application information. This information is recorded as an alternate name (i.e., Country_Name) with the *context* of Facility Data System, Mailing_Address Entity, and the context definition, "The name of a country."

5.2.3.4 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. This concept is not the same as the concept for the standard Country Short Name data element, which is limited to the concept of country identifier. The name of this data element concept, following the guidelines described in Section 5.1.6, is "Address Country Identifier" and the definition is "An identifier for a primary geopolitical entity of the world which indicates an address."

The data element concept identifier listed in the Annex F table demonstrates that the data element concept is different than for the standard data element. This data element concept would 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.

5.2.3.5 Conceptual Domain and Value Meanings

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. This is indicated by the reuse of the conceptual domain identifier in the Annex F table.

5.2.3.6 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.

5.2.3.7 Example

The example for an enumerated data element must be one of the permissible values enumerated for that value domain. The example "Denmark" has been used for all country name examples in the Annex F table, to demonstrate that the permissible values are the same as for the standard data element, Country Short Name, and that the data for all of the country name examples can be translated at the conceptual domain level.

5.2.3.8 Origin

The application data element for country name in a mailing address is taken from a computer application, the Facility Data System. This same data element might be reused in many applications, where mailing address information is included. The data element name might differ across applications, as is recorded in the name context. The meaning of the data element across applications must, however, be defined to reflect the exact same concept for the data to be shared.

5.3 Register a Group of Data Elements

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

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

X Group NameXThe name of a group of data elements.

X Group DefinitionXText that describes the features of, specifies relationships of, or establishes the context for a group of data elements.

X Authoritative SourceXThe originating point of information that provides an authoritative reference for a group of data elements.

X Source RationaleXThe text that explains the reasons for using the selected source materials in development of a group of data elements.

X Potential Usage CommentsXThe text that describes how a group of data elements can be used.

X Group IdentifierXThe system generated identifier for a group of data elements.

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

X System architecture, where the data elements make up a logical entity (e.g., mailing address).

X Data element components, where the individual data elements are grouped to make another data element (e.g., urban style street address).

X Usage, where the elements have a common usage (e.g., data elements in a data standard).

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

5.3.1 System Architecture Group

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

X Urban-style Street Address Text. The text that describes the urban-style street name and number where the mail is delivered.

X Post Office Box Number. The number of the post office box where mail for the addressee is delivered.

X Mailing Address City Name. The name of the city, town, or village where the mail is delivered.

X Mailing Address State Code. The alphabetic code assigned by the U.S. Postal Service that represents the state where the mail is delivered.

X Mailing Address Postal Code. The code that represents the code assigned by a postal service that provides information about the location of a place where mail is delivered.

X Mailing Address Country Name. The name of the country where a mail piece is delivered.

An example of metadata for the Mailing Address group is provided in Exhibit 5.2.

5.3.2 Composite Data Element

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

X Building Number. The number assigned to a building or a land parcel along the street to identify location and to ensure accurate mail delivery.

X Pre-Directional Code. The code that represents the direction the street has taken from some arbitrary starting point, and that precedes the street name.

X Street Name. The name assigned to a street or road, not including other urban-style street address components.

X Street Suffix Code. The code that represents the qualifier that follows the street name in a street address.

X Post-Directional Code. The code that represents the direction the street has taken from some arbitrary starting point, and that follows the street suffix.

X 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.

X Suite Number. The number that represents the specific room, apartment, or other secondary component of an address.

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

Another example of a composite data element group is telephone number, which groups data elements for area code, access code, unit number, and extension number.

5.3.3 Use Group

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

X 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.

X 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.

X Altitude. The measure of the distance in meters of a position above or below the surface of a reference datum.

X Latitude Sexagesimal Measure. The sexagesimal measure of the angular distance of a position on earth on a meridian north or south of the equator.

X Longitude Sexagesimal Measure. The sexagesimal measure of the angular distance of a position on earth on a meridian east or west of the prime meridian.

The latitude and longitude data elements provide information about the formats and units of measure that enable translation of the data for data sharing. The rules associated with the standard provide instructions for grouping the data elements for data sharing (e.g., latitude and longitude must be measured by the same unit when grouped together for data transfer).

The following table contains examples of the metadata that should be captured about a group of data elements when the group is registered.

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

Exhibit 5.2. Metadata for Groups

6 Complex Data

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

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

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

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

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

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

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

• **Data Group -** A set of data elements considered as a logical unit.

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

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

• Have the Rule entity account for the transformation formula.

• Put an attribute on the relationship between Data Element and Rule, called role, to distinguish data elements which are input to a transformation and the data elements which are output from a transformation (derived).

• A lookup table entity, such as Derivation Type, is needed to keep track of the type of transformation used.

• A recursive or hierarchical relationship on Rule is necessary to account for combinations of transformations.

Annex A
(informative)

References

[1] ISO 1087:1990 Terminology - Vocabulary.

[2] ISO/DIS 1087-1 Terminology - Vocabulary - Part 1: Theory and application (Partial revision of ISO 1087:1990).

[3] ISO/DIS 1087-2 Terminology - Vocabulary - Part 2: Computer applications (Partial revision of ISO 1087:1990).

[4] ISO/IEC 2382:1979-1998 Parts 1-32 Information technology - vocabulary.

[5] ISO 2788:1986 Documentation - Guidelines for the establishment of monolingual thesauri.

[6] ISO 3166-1:1997 Codes for the representation of names of countries and their sub-divisions.

[7] ISO 5964:1985 Documentation - Guidelines for the establishment of multilingual thesauri.

[8] ISO 6709, 1983-05-15 Standard representation of latitude, longitude and altitude for geographic point locations.

[9] ISO/IEC 7826-1:1994 Information technology - General structure for the interchange of code values - Part 1: Identification of coding schemes.

[10] ISO/IEC 7826-2:1994 Information technology - General structure for the interchange of code values - Part 2: Registration of coding schemes.

[11] ISO/IEC 11404:1996 Information technology B Programming languages, their environments and system software interfaces B Language-independent datatypes.

[12] SC32 N0147 Horizontal Issues and Encodable Value Domains in Electronic Commerce.

[13] ANSI X3.61B1986 Representation of Geographic Point Locations for Information Interchange.

[14] Zachman, John A., *The Framework for Enterprise Architecture: Background, Description and Utility*, 1997, http://www.ozemail.com.au/~visible/papers/zachman3.htm

Annex B

(informative)

Definitions of representation class terms

X Amount - the sum total of two or more quantities; an aggregate.

X Code - a symbol used to represent something.

∃ Discriminator - A distinction that differentiates one from another.

X Graphic - diagrams, graphs, mathematical curves, or the like.

• Identifier - Something that represents to be, regards, or treats as the same or identical.

∃ Indicator - Anything that serves to point out or direct attention to, as of a measuring device.

∃ Label - A short word or phrase descriptive of a person, group, or intellectual movement, or indicating that what follows belongs in a particular category or classification.

X Measure - the extent, dimensions, quantity, etc. of something ascertained by comparison with a standard.

X Name - a word or combination of words by which a person, place, object, or thought is known.

X Number - a numeral or group of numerals.

X Picture - a visual representation of a person, object, or scene.

X Quantity - the property of magnitude of something.

X Text - a unit of connected speech or writing often composed of one or more sentences that form a cohesive whole.

∃ Tag - A descriptive word or phrase applied to a person, group, organization, etc. as a label or means of identification or epithet.

Annex C
(informative)

Principles of managing shared data

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

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

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

C.1 Data

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

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

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

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

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

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

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

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

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

C.2 Concept

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

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

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

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

C.3 Object class

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

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

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

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

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

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

C.4 Property

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

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

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

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

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

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

C.4.7 It is desirable to describe properties without redundancy within the universe of discourse. The same property class, but with different names and/or wording of definitions, should eventually be normalized (a.k.a. harmonized or rationalized).
C.4.8 Properties are sometimes called "characteristics".

C.5 Data element concept

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

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

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

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

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

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

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

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

C.6 Attribute

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

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

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

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

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

C.7 Representation

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

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

C.8 Data element representation

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

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

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

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

C.8.5 A data element representation may have a "compound datatype" that separates the representation into constituent parts. A compound datatype would only be plausible where the data element representation could be used as the representation of a single data element concept.
C.8.6 A value domain may be an aggregation of a set of smaller value domains.

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

C.9 Data element

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C.10 Enumerated domain

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

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

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

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

C.11 Identifier

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

C.11.2 Identifiers will carry no intelligence.

C.12 Name

C.12.1 A name will not be used as an identifier.
C.12.2 Various names for various contexts where the names are used and have meaning are important metadata.

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

C.13 Quality

C.13.1 Data elements have several levels of quality.

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

C.14 Registration Authority

C.14.1 A Registration Authority is self-nominated.

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

C.14.3 A Registration Authority manages a data registry.

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

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

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

C.15 Data Registry

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Annex D
(informative)

Data registry uses and users

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

With this approach, the key to sharing data is thus to share and reuse metadata. We can put this metadata in a catalog that is organized in a way that all stakeholders can use it. Users can have direct access to items in the catalog with convenient retrieval procedures.
When we catalog all the data used in an enterprise, we are confronted with several ways to represent the same "fact". Information in the catalog can be organized to assist data administrators to identify redundancy. Data administrators can use the metadata catalog to standardize preferable data descriptions. By labeling well-described and sanctioned units of data in the catalog, other users will know which form of data representation to use.

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

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

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

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

supports administration and retrieval of registered data. A data registry is definitely more than just another data dictionary.

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

11

Annex E
(informative)

Conceptual and logical data models

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

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

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

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

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

Annex F
(Informative)

Table of data elements for examples

To be added

13

Annex G
(Informative)

Top down approach to data element registration

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

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

1) Biological Organisms XAll life forms considered as entities.

2) Biological Organism Types XAll ways of typing biological organisms.

G.1 Biological Organisms

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

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

G.1.1 Data Element Concepts

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

G.1.2 Data Elements

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

X Biological Organism Taxonomic Name XThe systematic name that provides a definitive classification for a biological organism.

X Biological Organism Vernacular Name XThe common name that is associated with a biological organism.

X ITIS Taxonomic Serial NumberXThe unique number assigned to a biological organism by the Integrated Taxonomic Information System (ITIS)[3].

X Biological Identification NumberXThe unique number assigned to a biological organism by the Biological Registry System.

G.1.3 Permissible Values

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

G.2 Biological Organism Types

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

X BiotaXAn animal, plant, fungus, or other biological organism of a region or period.

X VirusXAn ultramicroscopic agent that replicates only within the cells of living hosts, which are mainly bacteria, plants, and animals.

[3]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.

3
X GroupXA collection of biological organisms that are related in some way.

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

G.2.1 Data Element Concepts

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

G.2.2 Data Elements

Data elements associated with this data element concept might be:

X Biological Organism Type NameXThe name of the type of a biological organism.

X Biological Organism Type CodeXThe code that represents a type of biological organism.

G.2.3 Permissible Values

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

G.3 Top Down Population of a Registry

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

Conceptual Domain (CD) Name
Conceptual Domain Definition
CD ID
Value Meanings
    VM Begin Date
    VM End Date
    VM ID
Data Element Concept (DEC) Name
Data Element Concept Definition
    DEC ID
Representation
Value Domain (VD)
    VD ID
    Domain Type
        Determinant type
        Range limits
    Datatype

4

- Format
- Minimum
- Maximum
- Unit of Measure
- Precision

Data Element Name Context
Data Element Definition

Data Element Name
DI:VI
Permissible Values
PV Begin Date
PV End Date
Example
Origin
Note/Description