ISO/IEC JTC 1/SC 32 N 0455

Date: 2000-02-15
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	Other Document (Open)
TITLE	Study Period draft for technical report on Terminology Management for an ISO/IEC 11179 Metadata Registry
SOURCE	US National Body [J. Carpenter]
PROJECT NUMBER	
STATUS	Interested parties are invited to participate. Approved for Study Period by SC 32 at the Santa Fe meeting 2000-01-28 Resolution 8.
REFERENCES	
ACTION ID.	FYI
REQUESTED ACTION	Interested parties should contact the US National Body
DUE DATE	
Number of Pages	23
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
available from the JTC 1/SC 32 WebSite http://bwonotes5.wdc.pnl.gov/SC32/JTC1SC32.nsf

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

ISO/IEC JTC 1/SC 32/WG 2 N **00202**

Date: 1999-12-10

ISO/IEC PDTR

ISO/IEC JTC 1/SC 32/WG 2

Secretariat: ANSI

Terminology Management for an ISO/IEC 11179 Metadata Registry

Élément introductif — Élément central — Élément complémentaire

---

Warning

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

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

Contents

Foreword

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

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

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

— type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;

— type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;

— type 3, when a technical committee has collected data of different kind from that which is normally published as an International Standard ("state of the art", for example).

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

Technical Reports are drafted in accordance with the rules given in the ISO/IEC Directives, Part 3.

Attention is drawn to the possibility that some of the elements of this Technical Report may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR , which is a Technical Report of type [1, 2 or 3], was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 32, *Data Management Interchange*.

This document is being issued in the Technical Report (type 2) series of publications (according to subclause G.3.2.2 of Part 1 of the ISO/IEC Directives, 1995) as a "prospective standard for provisional application" in the field of [insert description] because there is an urgent need for guidance on how standards in this field should be used to meet an identified need.

This document is not to be regarded as an "International Standard". It is proposed for provisional application so that information and experience of its use in practice may be gathered. Comments on the content of this document should be sent to the ISO Central Secretariat.

A review of this Technical Report (type 2) will be carried out not later than three years after its publication with the options of: extension for another three years; conversion into an International Standard; or withdrawal.

This second/third/... edition cancels and replaces the first/second/... edition (), [clause(s) / subclause(s) / table(s) / figure(s) / annex(es)] of which [has / have] been technically revised.

Introduction

The purpose of this international standard is to specify a uniform way to formulate and manage concepts and terms within the context of an ISO/IEC 11179 metadata registry. This is intended to result in registry content that is consistent and easily located, mapped and shared. For data elements to be shareable, both the users and owners must have a common understanding of meaning, representation, and identification. Classification assists users to find a single data element, facilitates data administration and conveys semantic content. To facilitate the global interchange of data elements, there must be a mechanism in place to enable the mapping between different languages and the different terminology systems. Data elements must not only be adequately defined but users need to have a convenient way to retrieve and deploy these definitions through a variety of technologies.

This document integrates ISO standards addressing terminology, definitions, dictionary, thesaurus, and ontology construction and relates those standards to the data registry context. Further, this standard provides additional guidance for use of terminology in the creation, exchange and retrieval of data elements.

The primary data registry problems addressed by the development of this International Standard include the following:

— Lack of uniform guidance for the formulation of data element definitions

— No universal method of documenting classification structures for data element concepts (keyword lists, thesauri, taxonomies, ontologies)

— Lack of precision in data element definitions to enable mapping and support reuse

— Need for the documentation of methodologies to deploy data element terminology in search engines, EDI messages, intelligent agents, mediators and other structures needed to convey information to software enabling the retrieval of data elements

Terminology Management for an ISO/IEC 11179 Metadata Registry

1   Scope

2   Normative references

The following normative documents contain provisions which, through reference in this text, constitute provisions of this Technical Report. For dated references, subsequent amendments to, or revisions of, any of these publications do not apply. However, parties to agreements based on this Technical Report are encouraged to investigate the possibility of applying the most recent editions of the normative documents indicated below. For undated references, the latest edition of the normative document referred to applies. Members of ISO and IEC maintain registers of currently valid International Standards.

ISO ab-c:199x, *General title of series of parts — Part c: Title of part.*

ISO xyz (all parts), *General title of the series of parts .*

ISO 704 :1987 *Principles and methods of terminology*

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

ISO 1087:1990 *Terminology - Vocabulary*

ISO/DIS 1087-1:1996 *Terminology work -Vocabulary - Part 1 Theory and applications* (Partial revision of ISO 1087:1990). To be published

ISO/DIS 1087-2:1996 *Terminology work -Vocabulary - Part 2 Computer applications* (Partial revision of ISO 1087:1990). To be published

ISO 10241:1992 *International terminology standards - Preparation and layout*

ISO/DIS 860:1996 *Terminology work - Harmonization of concepts and terms*

ISO 5964:1985 Documentation - *Guidelines for the establishment of multilingual thesauri*

ANSI/NISO Z39.19-1993 *Guidelines for the Construction, Format, and Management of Monolingual Thesauri*

ISO/FDIS 12620 *Terminology - Computer applications - Data Categories*

ISO 5127-1:19XX *Documentation and information. Vocabulary - Part 1:Basic concepts.*

3   Terms and definitions

For the purposes of this Technical Report, the following terms and definitions apply/the terms and definitions given in … and the following apply.

3.1
term
text of the definition

3.2
term
text of the definition

4   Summary of required functions for terminology management in an ISO/IEC 11179 metadata registry

5   Methods for management of terminology for an ISO/IEC 11179 metadata registry

6   Registration and use of classification schemes

7   The guts of the standard

7.1   Integration with metadata registries

7.2   Organization structures for concepts

7.3   Terminology attributes for 11179 (modifications to Part 2 and 3)

This table presents terminology attributes that are proposed for inclusion in 11179. Inclusion of this minimum set of data elements is compliant with ISO 2788:1986 Documentation - Guidelines for the establishment and development of monolingual thesauri.

Table 1 — Terminology Attributes for 11179 - A Starter Set

Data Element Name	Data Element Definition
Classification Identifier	Number that uniquely identifies the Classification Scheme.
Classification Type Code	The code that indicates the type of classification scheme (thesaurus, glossary,etc.)
Classification Abbreviation Text	The abbreviated name of the Classification Scheme.

Classification Name	The term used to identify the classification scheme.
Classification Definition Text	The descriptive text about the classification scheme.
Classification Scope Notes Text	Text that describes the scope of the classification scheme.
Term Identifier	The number that uniquely identifies a term (word or phrase) in a classification scheme.
Term Name	The word used to identify the term.
Term Definition Text	The descriptive text that defines the meaning of the term.
Term Language Context Code	The language that provides context for the term.
Term Association Context Code	The code that identifies the context of the type of association between terms.
Source Name	The name of the document that identifies the source of the term.
Source Date	The calendar date that is associated with the source document.
Source url	The uniform resource locator that is the Internet address of the source document.
Source Organization ID	The number that uniquely identifies the organizational source of a term.
Source Organization Name	The name that identifies the organizational source of a term.
Source Point of Contact Name	The name of the person who is the point of contact for the organizational source.
Source Point of Contact email	The email address of the person who is the point of contact for the organizational source.
Source Point of Contact Phone	The telephone number of the person who is the point of contact for the organizational source.
Source Point of Contact	The mailing address of the person who is the point of

Address	contact for the organizational source (composed of component data elements).
Responsible Organization ID	The number that uniquely identifies the organization that is the steward for a term.
Responsible Organization Name	The name that identifies the organization that is the steward for a term.
Responsible Point of Contact Name	The name of the person who is the point of contact for the stewardship organization.
Responsible Point of Contact email	The electronic mail address of the person who is the point of contact for the stewardship organization.
Responsible Point of Contact Phone	The telephone number of the person who is the point of contact for the stewardship organization.
Responsible Point of Contact Address	The mailing address of the person who is the point of contact for the stewardship organization (composed of component data elements).

Annex A
(informative)

Framework for Semantics Management in Metadata Registries

Annex B
(informative)

Use Cases

B.1 Access to concepts

B.1.1 BRIEF DESCRIPTION

The meta-model specifies that for every data element, there will be a data element concept. Each data element concept may be related to one or more data elements that differ only in representation. To find data that is sharable at the data element level, it is necessary to find the associated data element concept.

B.1.2 ACTOR(S)

A registry user who wishes to register data elements or share data.

B.1.3 GOAL

Easy access to data element concepts.

B.1.4 FUNCTIONAL TRAITS

Data element concepts should be accessible by object class, property, data element, conceptual domains and value domains

B.2 Establish object classes and properties

B.2.1 BRIEF DESCRIPTION

A controlled vocabulary for assigning Object Classes, Properties, Modifiers and Qualifiers is needed.

B.2.2 ACTOR(S)

A registry user who wishes to register data elements

The Registrar.

B.2.3 GOAL

A limited, organized and well understood vocabulary of Object Classes, Properties, Modifiers and Qualifiers.

B.2.4 FUNCTIONAL TRAITS

The ability to search, create, and update data elements and data element concepts in the registry requires the limited, organized and well understood vocabulary of Object Classes, Properties, Modifiers and Qualifiers.

B.3 Use the Terminology Registry to Support Searching Documents and Databases

B.3.1 BRIEF DESCRIPTION

One may believe that the generation of an agency thesaurus would share a functional relationship to loading a controlled keyword list to search agency data. However, very often disparate organizations within an agency are responsible for managing a terminology system than those charged to operate and maintain the agency's search engine. The ability of a terminology system to export terms that can support the generation of weighted topic sets to be used by a search engine application to support users search needs is commonly problematic, and at best, clumsy. Nonetheless, it is logical to assume that any terminology system should be able to export terms to a topic editor for weighting. The ability to maintain and store this master file should reside in the terminology system. Ideally, the terminology system could be used to create and maintain a thesaurus, provide functionality to assign weighting to terms, and export terms for loading into search engine applications. This use case documents the need for a terminology system to provide the ability to support search engine technologies.

B.3.2 ACTORS

Primary Actors: Agency staff responsible for the implementation and maintenance of an agency-wide search engine, and agency staff responsible for the operation and maintenance of an agency thesaurus.

Other Actors: Users accessing agency text and database data from intranet/internet sites through the use of controlled vocabulary keyword searches.

B.3.3 GOAL

The ability of the terminology system to support the generation and maintenance of controlled vocabulary keyword lists/thesauri to import into the agency's search engine technology.

B.3.4 FUNCTIONAL TRAITS

The agency staff responsible for creating and maintaining an agency thesaurus through a terminology system need to share this data with those staff that feed the agency's keyword list into the search engine. Typically, the terminology system will generate an output file that can be used to support the creation of a topic set or knowledge base through the addition of weighting factors and operators. This topic set will be imported into the search engine application to drive the selection of keywords available to search agency text and databases, as well as to rank the retrieval results.

B.4 Generate controlled vocabulary for input into search engine

B.4.1 BRIEF DESCRIPTION

B.4.2 ACTORS

B.4.3 GOAL

B.4.4 FUNCTIONAL TRAITS

B.5 Controlled vocabulary used for input to data element design

B.5.1 BRIEF DESCRIPTION

In the course of designing a data element concept, the specialized controlled vocabulary is used within the definition as dictated by the associated classification scheme.   The use of the controlled vocabulary eases the identification and/or prevention of redundancies.

B.5.2 ACTORS

Functional Experts, Standards Developers, Data Submitters, Data Stewards, Registrar

B.5.3 GOAL

To assure the descriptions of data element concepts can be understood in the same way by users and organizations in the community of discourse, and that the data to be shared between organizations is interpreted correctly.

B.5.4 FUNCTIONAL TRAITS

The use of a well understood, controlled vocabulary in the definition of a data element concept:

- Promotes precise, clear, consistent and unambiguous data element concepts that assure correct interpretation by users.

- Results in appropriates 'finds' by the search engine applications utilizing that vocabulary

B.5.5 PRE-CONDITIONS

A specialized controlled vocabulary / classification scheme

B.6 Relate Existing or New Data Element to Existing or New Legislative/Regulatory Requirement

B.6.1 BRIEF DESCRIPTION

Laws and regulations enable the collection of information. They refer directly or indirectly to the elements of information needed to support compliance and enforcement activities. Law and regulation may use specialized terminology, either by definition within the law or by reference to an established standard. Over time, terminological confusion can lead multiple problems in data collection and definition. Traceability between an information collection and the (multiple) regulations justified to collect it may be lost. Terminology used within information management to name and define data elements may become different from terminology used in enabling regulation, forming a language barrier between regulation writer and information systems analysts at worst and introducing unnecessary confusion at least. Writers of new legislation have no effective way of determining whether the information requirements in new legislation have already been met. Enterprise data stewards and data registration authorities cannot effectively advise regulation writers on suitable names and definitions for new data collections.

B.6.2 ACTORS

Primary Actor: Data Registration Authority/Enterprise Data Steward

Other Actors: Regulation Writers

B.6.3 GOAL

All actors can trace information requirements from inception in law and regulation through storage in enterprise information systems. Common terminology is employed as much as judged practicable, and differences in terminology do not become obstacles to identifying common information requirements and providing common information solutions.

B.6.4 FUNCTIONAL TRAITS

The actor begins by knowing that a data element is used in an enterprise system and needs to verify the regulatory basis for collecting it.

 The actor needs to ensure that the data element is named and defined consistent with the terminology used to describe the information required in the law and/or regulation that justified the information collection.

The actor needs to research law, regulation, or industry/information discipline to identify the alternative ways of naming and defining similar or related concepts, with a reasonable assurance of identifying all of the alternative terminologies used in an information discipline of interest.

B.6.5 PRE-CONDITIONS

B.6.6 POST-CONDITIONS

B.6.7 EXPLANATORY TERMS

B.7 Generate a Controlled Vocabulary for Cataloging of Documents or Data

B.7.1 BRIEF DESCRIPTION

Agencies employ the services of library science personnel to catalog agency resources. These resources may include traditional paper-based products, graphics, websites, and agency data. Cataloging is usually conducted through a software application which manages the assignment of keywords selected from a controlled vocabulary list. The application of these keywords to agency resources not only facilitates traditional retrieval of these products, but may also drive the efficiency with which a search engine locates these resources on an agency intranet/internet. The ability of a terminology system to provide for the creation, maintenance, and export of this controlled vocabulary list would alleviate the need for a separate application, and provide library science personnel access via the internet. Additionally, this functionality would support the standardization of vocabulary content. This use case documents the need for a terminology system to provide an internet based application to support the creation and maintenance of a controlled vocabulary for cataloging of agency resources.

B.7.2 ACTORS

Primary Actors: Agency library science personnel responsible for managing the cataloging activities of agency resources.

Other Actors: Users who benefit from efficient retrieval of agency information.

B.7.3 GOAL

Develop the ability in a terminology management system to create, maintain, and export a controlled vocabulary list.

B.7.4 FUNCTIONAL TRAITS

Agency library science personnel would have the ability to perform cataloging activities via the intranet/internet using a common controlled vocabulary list. An export file from the terminology system could be created to support the creation of topic sets used to drive agency search engine applications.

B.7.5 PRE-CONDITIONS

B.7.6 FLOW OF EVENTS

B.7.7 POST-CONDITIONS

B.7.8 EXPLANATORY TERMS

B.8 Extract Multiple Contexts for a Single Term

B.8.1 BRIEF DESCRIPTION

A user entering terms into a terminology system needs the ability to determine if a term has been entered previously and to view the context of stored terms. If the terms already exist in the terminology system, the context with which they have been used may reside in the source of the term, or in definitional differences in the term. A terminology management system needs the ability to store and display multiple sources for a term, as well as multiple definitions for a term. Terms, source fields and definition fields should all be searchable for ease of term entry and to minimize duplication of terms from the same source with the same definition. This use case documents the need for a terminology system to store and display multiple definitions and sources for a term.

B.8.2 ACTORS

Primary Actors: A registry user registering his own terms.

Other Actors:

B.8.3 GOAL

Develop the ability in a terminology management system to store and display multiple sources and definitions for a term.

B.8.4 FUNCTIONAL TRAITS

A user would like to enter a subject matter glossary into the terminology system. Prior to importing his file into the terminology system, he wants to check to see if any of the terms have been used before in the same context. He can either view, or download the list of current terms, their sources, and definitions to compare against those to be loaded. Current registry terms would be displayed with sources and definitions, and possibly with a registry preferred definition appearing first in the list. This ordering may help to encourage users to adopt and reuse a preferred definition. Individual terms may be accessed from the system by searching by term name, source, or a keyword that in the definition of the term.

B.8.5 PRE-CONDITIONS

B.8.6 POST-CONDITIONS

B.8.7 EXPLANATORY TERMS

B.9 Define terms used in data element definitions

B.9.1 BRIEF DESCRIPTION

Specialized language is frequently used in the formulation of data element and data element concept definitions in a data element registry.

B.9.2 ACTOR(S)

A registry user who wishes to understand data elements and data element concepts in a registry.

B.9.3 GOAL

Reduce ambiguity and enhance understanding of data element and data element concept definitions.

B.9.4 FUNCTIONAL TRAITS

Ideally, definitions would be provided for all specialized language used in data element definitions and data element concept definitions. The collection of all these definitions of specialized terms would constitute the basis for registry glossary.

B.10   Retrieve a classification scheme

B.10.1 SUMMARY

An X3.285 registry can contain a classification scheme. A classification scheme is a set of classes and relationships, or in X3.285 terms, classified components and component classification associations. It is assumed that all components of a scheme are administered as a unit, i.e., have identical properties as X3.285 administered component types.

In this use case, the user desires to retrieve the entire classification scheme for display by some client software. The user submits a query for a named classification scheme and the registry returns all classes and relationships in the scheme and the administered component properties of the classification scheme.

B.10.2 ACTOR(S)

A user who wants to display an entire classification scheme and its administrative properties.

User's software, called the client software, which interacts with the registry.

B.10.3 GOAL

Display a classification scheme in tabular or graphical form along with a textual description of the administrative properties of the scheme.

B.10.4 PRE-CONDITIONS

The registry contains a classification scheme.

B.10.5 BEGINS WHEN

User selects the function in the client software to retrieve a classification scheme.

B.10.6 ENDS WHEN

The entire classification scheme is displayed.

B.10.7 POST-CONDITIONS

None.

B.10.8 EXPLANATORY TERMS