ANSI/NISO Z39.19-2005 **ISBN: 1-880124-65-3**

Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

Abstract: This Standard presents guidelines and conventions for the contents, display, construction, testing, maintenance, and management of monolingual controlled vocabularies. This Standard focuses on controlled vocabularies that are used for the representation of content objects in knowledge organization systems including lists, synonym rings, taxonomies, and thesauri. This Standard should be regarded as a set of recommendations based on preferred techniques and procedures. Optional procedures are, however, sometimes described, e.g., for the display of terms in a controlled vocabulary. The primary purpose of vocabulary control is to achieve consistency in the description of content objects and to facilitate retrieval. Vocabulary control is accomplished by three principal methods: defining the scope, or meaning, of terms; using the equivalence relationship to link synonymous and nearly synonymous terms; and distinguishing among homographs.

An American National Standard
developed by the
National Information Standards Organization

Approved July 25, 2005
by the
American National Standards Institute

Published by the National Information Standards Organization

NISO Press, Bethesda, Maryland, U.S.A.

About NISO Standards

NISO standards are developed by the Standards Committees of the National Information Standards Organization. The development process is a strenuous one that includes a rigorous peer review of proposed standards open to each NISO Voting Member and any other interested party. Final approval of the standard involves verification by the American National Standards Institute that its requirements for due process, consensus, and other approval criteria have been met by NISO. Once verified and approved, NISO Standards also become American National Standards.

This standard may be revised or withdrawn at any time. For current information on the status of this standard contact the NISO office or visit the NISO website at: http://www.niso.org

Published by
NISO Press
4733 Bethesda Avenue, Suite 300
Bethesda, MD 20814
www.niso.org

ISSN: 1041-5653
ISBN: 1-880124-65-3

Contents

Figures

Tables

Foreword

(This foreword is not part of the *Guidelines for the Construction, Format, and Management of Controlled Vocabularies*, ANSI/NISO Z39.19-2005. It is included for information only.)

Acknowledgements

The Z39.19 committee thanks the following publishers of vocabularies or software that were used in the examples of this standard: Energy Technology Data Exchange and International Atomic Energy Agency; Food and Agriculture Organization of the United Nations; Gallery Systems; Getty Research Institute; Institute of Education Sciences and U.S. Department of Education; MACS project of the Conference of European National Librarians; Multisystems; Synaptica; Thomson Gale; U.S. Congressional Research Services Library Services Division; U.S. Defense Technical Information Center; U.S. Library of Congress Prints and Photographic Division; U.S. National Aeronautics and Space Administration; U.S. National Library of Medicine; United Nations Educational, Scientific and Cultural Organization; WordNet; and Xrefer.

The Committee wishes to acknowledge Gale Rhoades of Wise Bytes, Daly City, California for facilitating permission to use Macrex Indexing software to create the index for this standard.

NISO acknowledges the early contributions of Amy Warner to the revision of this standard.

About This Standard

ANSI/NISO Z39.19, now in its fourth edition, is truly "built on the shoulders of giants." The first edition of ANSI/NISO Z39.19, published in 1974, was prepared by Subcommittee 25 on *Thesaurus Rules and Conventions* of the American National Standards Committee Z39 on *Standardization in the Field of Library Work, Documentation, and Related Publishing Practices*—the forerunner to NISO. The authors of that first standard drew heavily on standards of practice developed by experts in the field including the Engineers Joint Council, the Committee on Scientific and Technical Information of the Federal Council for Science and Technology, and UNESCO. In the late 1970's the standard was revised by Madeline Henderson and a new edition was released in 1980.

In 1988, under the leadership of Dr. Bella Hass Weinberg, work began to revise Z39.19 and a significantly expanded standard, taking into account variations in approach, was released in 1993. In 1998 the standard was reviewed by the NISO Members during its regular five-year review cycle. The NISO members supported reaffirming the standard, however, many members advised that there soon would be a need to reexamine the standard in light of the rapidly evolving electronic information environment. In response, NISO organized a national *Workshop on Electronic Thesauri*, held November 4-5, 1999, to investigate the desirability and feasibility of developing a standard for electronic thesauri. This revision of Z39.19 grew out of the recommendations developed by consensus at that Workshop.[1]

Similar revision work is underway internationally on ISO 2788, *Guidelines for the establishment and development of monolingual thesauri*. In the UK, work is under way to revise and extend BS 5723, which is equivalent to ISO 2788. Publication of the new BS 8723 Parts 1 and 2, which will cover similar ground as Z39.19, is expected in 2005. Parts 3, 4, and 5, expected during 2006, will deal respectively with other vocabularies such as taxonomies and classification schemes; interoperability between vocabularies; and protocols and formats for exchange of thesaurus data.

This standard continues to be a reflection of its time. When Z39.19 was first conceived, thesaurus terms were generally applied when indexing collections of documents, most commonly printed

[1] The workshop was co-sponsored with the American Psychological Association (APA), the American Society of Indexers (ASI), and the Association for Library Collections and Technical Services (ALCTS), a division of the American Library Association. The *Report on the Workshop on Electronic Thesauri*, November 4-5, 1999 is available at: http://www.niso.org/news/events_workshops/thes99rprt.html.

resources such as journal articles, technical reports, newspaper articles, etc. As new information storage and retrieval systems have emerged, the concept of *document* has expanded to include patents, chemical structures, maps, music, videos, museum artifacts, and many other types of materials that are not traditional documents. This revision addresses this more inclusive notion of document. Furthermore, the display methods described in earlier editions of the standard were almost entirely for printed products. This revision expands the description of display options and provides new examples to illustrate these display options.

The revision process was supported by The H.W. Wilson Foundation, the Getty Grant Program, and the National Library of Medicine. Through their generosity and interest, these institutions brought validation and resources that enabled this revision to be launched and arrive at a successful and timely conclusion.

Trademarks, Service Marks

NISO Voting Members

At the time this standard was balloted, NISO had the following voting members:

3M
Susan Boettcher, Roger Larson (Alt)

AIIM International
Betsy A. Fanning

American Association of Law Libraries
Robert L. Oakley, Mary Alice Baish (Alt)

American Chemical Society
Matthew Toussant

American Library Association
Betty Landesman

American Society for Information Science and Technology (ASIS&T)
Gail Thornburg

American Society of Indexers
Judith Gibbs

American Theological Library Association
Myron Chace

ARMA International
Diane Carlisle

Armed Forces Medical Library
Diane Zehnpfennig, Emily Court (Alt)

Art Libraries Society of North America (ARLIS/NA)
Mark Bresnan

Association of Information and Dissemination Centers (ASIDIC)
Marjorie Hlava

Association of Jewish Libraries
Caroline R. Miller, Elizabeth Vernon (Alt)

Association of Research Libraries (ARL)
Duane E. Webster, Julia Blixrud (Alt)

Auto-Graphics, Inc.
Paul Cope

Barnes & Noble, Inc.
Douglas Cheney

Book Industry Communication
Brian Green

California Digital Library
Daniel Greenstein, John Kunze (Alt)

Cambridge Information Group
Michael Cairns, Matthew Dunie (Alt)

Checkpoint Systems
Douglas Karp, Frank Palazzo (Alt)

The Cherry Hill Company
Cary Gordon

College Center for Library Automation (CCLA)
J. Richard Madaus, Ann Armbrister (Alt)

Colorado State Library
Brenda Bailey-Hainer, Steve Wrede (Alt)

CrossRef
Edward Pentz, Amy Brand (Alt)

Davandy, L.L.C.
Michael J. Mellinger

Dynix Corporation
Ed Riding, Gail Wanner (Alt)

EBSCO Information Services
Gary Coker, Oliver Pesch (Alt)

Elsevier
Paul Mostert

Endeavor Information Systems, Inc.
Sara Randall, Shelley Hostetler (Alt)

NISO Voting Members (continued)

Entopia, Inc.
Igor Perisic

Ex Libris, Inc.
James Steenbergen

Fretwell-Downing Informatics
Robin Murray

Gale Group
Katherine Gruber, Justine Carson (Alt)

Geac Library Solutions
Eric Conderaerts, Eloise Sullivan (Alt)

GIS Information Systems, Inc.
Candy Zemon, Paul Huf (Alt)

H.W. Wilson Company
Ann Case, Patricia Kuhr (Alt)

Helsinki University Library
Juha Hakala

Index Data
Sebastian Hammer, David Dorman (Alt)

INFLIBNET Centre
T A V Murthy, Rajesh Chandrakar (Alt)

Infotrieve
Jan Peterson

Innodata-Isogen
Carolyn Muzyka

Innovative Interfaces, Inc.
Gerald M. Kline, Betsy Graham (Alt)

The International DOI Foundation
Norman Paskin

Ithaka / JSTOR / ARTstor
David Yakimischak, Bruce Heterick (Alt)

John Wiley & Sons, Inc.
Eric A. Swanson

Library Binding Institute
Debra Nolan

The Library Corporation
Mark Wilson, Ted Koppel (Alt)

Library of Congress
Sally H. McCallum

Los Alamos National Laboratory
Richard E. Luce

Lucent Technologies
M. E. Brennan

Medical Library Association
Nadine P. Ellero, Carla J. Funk (Alt)

MINITEX
Cecelia Boone, William DeJohn (Alt)

Modern Language Association
Daniel Bokser, B. Chen (Alt)

Motion Picture Association of America (MPAA)
Michael Owens

MuseGlobal, Inc.
Kate Noerr, Clifford Hammond (Alt)

Music Library Association
Mark McKnight, David Sommerfield (Alt)

National Agricultural Library
Eleanor G. Frierson, Gary K. McCone (Alt)

National Archives and Records Administration
Nancy Allard

National Federation of Abstracting and Information Services (NFAIS)
Marjorie Hlava

National Library of Medicine
Barbara Rapp

National Security Agency
Kathleen Dolan

Nylink
Mary-Alice Lynch, Jane Neale (Alt)

OCLC, Inc.
Thomas Hickey

Openly Informatics, Inc.
Eric Hellman

OpenRFP, Inc.
Ralph Shoffner, Donald Chvatal (Alt)

Paratext
Robert Asleson, Eric Calaluca (Alt)

ProQuest Information and Learning
Thomas Hamilton, Carol Brent (Alt)

Random House, Inc.
Laurie Stark

Recording Industry Association of America
Bruce Block, Carlos Garza (Alt)

RLG
Lennie Stovel

Sage Publications
Carol Richman, Richard Fidezuk (Alt)

Serials Solutions, Inc.
Mike McCracken

SIRSI Corporation
Greg Hathorn, Slavko Manojlovich (Alt)

Society for Technical Communication (STC)
Frederick M. O'Hara, Jr.,
Annette D. Reilly (Alt)

Society of American Archivists
Lisa Weber

Special Libraries Association (SLA)
Foster J. Zhang

Synapse Corporation
Trish Yancey, Dave Clarke (Alt)

TAGSYS, Inc.
John Jordon, Anne Salado (Alt)

Talis Information Ltd
Terry Willan, Katie Anstock (Alt)

Thomson ISI
Carolyn Finn

Triangle Research Libraries Network
Mona C. Couts

U.S. Department of Commerce, NIST, Office of Information Services
Mary-Deirdre Coraggio

U.S. Department of Defense, DTIC (Defense Technical Information Center)
Richard Evans, Jane L. Cohen (Alt)

NISO Voting Members (continued)

U.S. Department of Energy, Office of Scientific & Technical Information
Ralph Scott, Karen Spence (Alt)

U.S. Government Printing Office
Judith Russell, T.C. Evans (Alt)

VTLS, Inc.
Carl Grant

WebFeat
Todd Miller, Paul Duncan (Alt)

Zone & Zone Co. Ltd.
Ahn Gye Sung

NISO Board of Directors

At the time NISO approved this standard, the following individuals served on its Board of Directors:

Carl Grant, Chair
VTLS, Inc.

Robin Murray, Vice Chair and Chair-Elect
Fretwell-Downing Informatics

Jan Peterson, Immediate Past Chair
Infotrieve

Douglas Cheney, Treasurer
Barnes & Noble, Inc.

Patricia Stevens, Chair of SDC
OCLC, Inc.

Patricia R. Harris, Executive Director/Secretary
NISO

Directors:

Nancy Davenport
Council on Library and information Resources

Lorcan Dempsey
OCLC, Inc.

Daniel Greenstein
California Digital Library

James Neal
Columbia University

Oliver Pesch
EBSCO Information Services

Bruce Rosenblum
INERA, Inc.

Eric Swanson
John Wiley & Sons, Inc.

Committee Members

Emily Fayen, Project Lead
MuseGlobal

Vivian Bliss
Microsoft

Carol Brent
ProQuest

Dave Clarke
Synapse, the Knowledge Link Corporation

John Dickert
DTIC

Lynn El-Hoshy
Library of Congress

Margie Hlava
Access Innovations

Patricia Harpring
Getty Research Institute

Stephen Hearn
American Library Association

Cynthia Hodgson
NISO

Sabine Kuhn
CAS

Pat Kuhr
H.W. Wilson

Diane McKerlie
Design Strategy

Paula McCoy
ProQuest

Peter Morville
Semantic Studios

Stuart Nelson
National Library of Medicine

Allan Savage
National Library of Medicine

Diane Vizine-Goetz
OCLC, Inc.

Trish Yancey
Synapse, the Knowledge Link Corporation

Marcia Lei Zeng
Special Libraries Association

Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

1 Introduction

Vocabulary control is used to improve the effectiveness of information storage and retrieval systems, Web navigation systems, and other environments that seek to both identify and locate desired content via some sort of description using language. The primary purpose of vocabulary control is to achieve consistency in the description of content objects and to facilitate retrieval.

1.1 Need for Vocabulary Control

The need for vocabulary control arises from two basic features of natural language, namely:

- Two or more words or terms can be used to represent a single concept

  Example: **salinity/saltiness**
  VHF/Very High Frequency

- Two or more words that have the same spelling can represent different concepts

  Example: **Mercury (planet)**
  Mercury (metal)
  Mercury (automobile)
  Mercury (mythical being)

1.2 How Vocabulary Control is Achieved

Vocabulary control is achieved by three principal methods:

- Defining the scope, or meaning, of terms;
- Using the equivalence relationship to link synonymous and nearly synonymous terms; and
- Distinguishing among homographs.

2 Scope

This Standard presents guidelines and conventions for the contents, display, construction, testing, maintenance, and management of controlled vocabularies. It covers all aspects of constructing controlled vocabularies including extensive rules and guidelines for term selection and format, the use of compound terms, and establishing and displaying various types of relationships among terms.

This Standard focuses on controlled vocabularies that are used for the representation of content objects. Controlled vocabularies covered by this Standard include lists of controlled terms, synonym rings, taxonomies, and thesauri. The guidelines apply to all four types unless noted otherwise.

This Standard should be regarded as a set of recommendations based on preferred techniques and procedures. Optional procedures are, however, sometimes described, e.g., for the display of terms in a controlled vocabulary.

The conventions used in this Standard to indicate the force of recommendations are: ***must*** (required for meeting the Standard), ***should*** (recommended), and ***may*** (optional). The Standard also uses the conventions ***must not*** (not allowed in order to be in compliance with the Standard) and ***should not*** (not recommended).

A summary of the recommendations in this Standard can be found in Appendix A.

2.1 Applying the Standard

This Standard is primarily intended to be applied to controlled vocabularies for use with knowledge organization systems. "The term *knowledge organization systems* is intended to encompass all types of schemes for organizing information and promoting knowledge management. Knowledge organization systems include classification schemes that organize materials at a general level (such as books on a shelf), subject headings that provide more detailed access, and authority files that control variant versions of key information (such as geographic names and personal names). They also include less-traditional schemes, such as semantic networks and ontologies. Because knowledge organization systems are mechanisms for organizing information, they are at the heart of every library, museum, and archive." [29]

This Standard does not cover numerical classification schemes (except as they correlate to topics such as Dewey, for example), ontologies or semantic networks.

Guidelines for the development of controlled vocabularies intended for use in postcoordinated retrieval systems are central to the Standard but the guidelines are also applicable to development of controlled vocabularies for precoordinated retrieval systems (e.g. Web navigation systems, browsing schemes, taxonomies, and the like).

The Standard does not provide guidelines for human or computer-aided content description (indexing) based on controlled vocabularies. For indexing procedures and practices, see NISO TR02-1997 [22].

2.2 Monolingual Controlled Vocabularies

The recommendations contained in this Standard relate specifically to monolingual controlled vocabularies, with particular reference to English. The section on Interoperability considers some of the issues surrounding multi-lingual controlled vocabularies, such as the special requirements of other languages and issues surrounding the display of parallel sequences showing conceptual equivalences in more than one language. For more on multilingual vocabularies, see ISO 5964 [30].

2.3 Types of Terms Covered by the Standard

This Standard provides guidelines for the selection, formulation, organization, and display of terms that together make up a controlled vocabulary. These terms form a controlled subset of natural language—a subset for a designated purpose or community.

The Standard does not suggest procedures for organizing and displaying headings consisting of mathematical and chemical formulas, nor for establishing proper names. (For information on these, refer to appropriate standards, such as *Anglo-American Cataloguing Rules*, 2nd ed. [20]) Such headings may be included in controlled vocabularies, however, and relationships among them may be indicated through the devices presented in this Standard.

This is not a standard for the creation of authority files or authority lists as defined by the U.S. Library of Congress or others.

2.4 Interoperability

In the decades since this Standard was first issued there has been tremendous growth in the number of databases, information retrieval systems, and websites that employ controlled vocabularies at some stage in the process. Users frequently need to locate information in more than one system or database. Therefore, it is important to facilitate interoperability wherever possible. This Standard describes the issues, especially those arising from the occurrence of controlled vocabularies in multiple languages, and suggests some guidelines to help overcome the problems.

2.5 Maintenance

A controlled vocabulary can be as simple as a short list of terms or as complex as a thesaurus containing tens of thousands of terms with a complex hierarchical structure and many different types of relationships among the terms. In addition, because language changes over time, a controlled vocabulary is not frozen in time—new concepts emerge, terminology and usage change, and some terms go out of fashion or become obsolete. Thus, there is a continuing need to maintain the controlled vocabulary once it has been constructed. This Standard recommends features and procedures that make the task easier.

2.6 Testing and Evaluation

Finally, because a controlled vocabulary exists to serve its users, both those who employ it to describe content and those who use it to locate needed information, it is vital to conduct periodic tests with follow-on evaluation to ensure that the controlled vocabulary is as useful as possible. This Standard recommends some test and evaluation procedures.

3 Referenced Standards

ISO 5127, *Information and documentation — Vocabulary*

4 Definitions, Abbreviations, and Acronyms

4.1 Definitions

The following terms, as used in this Standard, have the meanings indicated. A more extensive Glossary of controlled vocabulary terms can be found in the Appendix.

Term	Definition
associative relationship	A relationship between or among terms in a controlled vocabulary that leads from one term to other terms that are related to or associated with it; begins with the words SEE ALSO or related term (RT).
asymmetric	Lacking symmetry. In the context of controlled vocabularies, reciprocal relationships are asymmetric when the relationship indicator used between a pair of linked terms is different in one direction than it is in the reverse direction, e.g. BT / NT. See also *symmetric* and *reciprocity*.

Term	Definition
authority file	A set of established headings and the cross-references to be made to and from each heading, often citing the authority for the preferred form or variants. Types of authority files include name authority files and subject authority files.
broader term	A term to which another term or multiple terms are subordinate in a hierarchy. In thesauri, the relationship indicator for this type of term is BT.
browsing	The process of visually scanning through organized collections of representations of content objects, controlled vocabulary terms, hierarchies, taxonomies, thesauri, etc.
candidate term	A term under consideration for admission into a controlled vocabulary because of its potential usefulness. Also known as *provisional term*.
category	A grouping of terms that are semantically or statistically associated, but which do not constitute a strict hierarchy based on genus/species, parent/child, or part/whole relationships. See also *tree structure*.
classification scheme	A method of organization according to a set of pre-established principles, usually characterized by a notation system and a hierarchical structure of relationships among the entities.
classified display	See *tree structure*.
compound term	A term consisting of more than one word that represents a single concept. Compound terms ***must*** be constructed according to the guidelines of this Standard. **Examples:** **duct tape** **New Jersey** **onion domes** See also *bound term* (in the glossary) and *precoordination*.
concept	A unit of thought, formed by mentally combining some or all of the characteristics of a concrete or abstract, real or imaginary object. Concepts exist in the mind as abstract entities independent of terms used to express them.
content object	An entity that contains data/information. A content object can itself be made up of content objects. For example, a journal is a content object made up of individual journal articles, which can each be a content object. The text, figures, and photographs included in a journal article can also be separate content objects. Paintings, sculpture, maps, photographs, and other non-textual objects are also content objects. The metadata for a content object can itself be a content object.

Term	Definition
controlled vocabulary	A list of terms that have been enumerated explicitly. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary ***must*** have an unambiguous, non-redundant definition. *NOTE: This is a design goal that may not be true in practice; it depends on how strict the controlled vocabulary registration authority is regarding registration of terms into a controlled vocabulary.* At a minimum, the following two rules ***must*** be enforced: 1. If the same term is commonly used to mean different concepts, then its name is explicitly qualified to resolve this ambiguity. *NOTE: This rule does not apply to synonym rings.* 2. If multiple terms are used to mean the same thing, one of the terms is identified as the preferred term in the controlled vocabulary and the other terms are listed as synonyms or aliases.
cross-reference	1. A direction from one term to another. See *associative relationship*; *equivalence relationship*; *hierarchical relationship*.
descriptor	See *preferred term.*
document	Any item, printed or otherwise, that is amenable to cataloging and indexing. The term applies not only to written and printed materials in paper or microform versions (e.g., books, journals, maps, diagrams), but also to non-print media (e.g., machine-readable records, transparencies, audiotapes, videotapes) and, by extension, to three-dimensional objects or realia (e.g., museum objects and specimens). A *document* is a *content object*.
entry term	The non-preferred term in a cross reference that leads to a term in a controlled vocabulary. Also known as "lead-in term." In thesauri, the relationship indicator for this type of term is U (USE); its reciprocal is UF (USED FOR). See also *preferred term.*
entry vocabulary	The set of non-preferred terms (USE references) which lead to the preferred terms in a controlled vocabulary. *NOTE: This term is used by some controlled vocabulary designers to represent the preferred as well as the non-preferred terms in a controlled vocabulary.*
equivalence relationship	A relationship between or among terms in a controlled vocabulary that leads to one or more terms that are to be used instead of the term from which the cross-reference is made; begins with the word SEE or USE.
flat format	An alphabetical display format of controlled vocabularies in which only one level of broader terms and one level of narrower terms are shown for each term.
focus	In a compound term, the noun component that identifies the class of concepts to which the term as a whole refers. Also known as *head noun*. See also *modifier*.

Term	Definition
generic posting	In controlled vocabularies, the treatment of narrower terms as equivalents, e.g., **furniture** UF beds; UF sofas. See also *up-posting*, in glossary.
generic structure	A controlled vocabulary format that indicates all hierarchical levels of terms within an alphabetic display by means of codes, indentation, and/or punctuation marks.
heading	A preferred name or term. Types of headings include proper name headings (which can be called identifiers), subject headings, and terms. A heading may include a *qualifier*.
hierarchical relationship	A relationship between or among terms in a controlled vocabulary that depicts broader (generic) to narrower (specific) or whole-part relationships; begins with the words broader term (BT), or narrower term (NT).
hierarchy	Broader (generic) to narrower (specific) or whole-part relationships, which are generally indicated in a controlled vocabulary through codes or indentation. See also *broader term*; *narrower term.*
homograph	One of two or more words that have the same spelling, but different meanings and origins. In controlled vocabularies, homographs are generally distinguished by *qualifiers*.
indexing	A method by which terms or subject headings from a controlled vocabulary are selected by a human or computer to represent the concepts in or attributes of a content object. The terms may or may not occur in the content object.
indexing language	A controlled vocabulary or classification system and the rules for its application. An indexing language is used for the representation of concepts dealt with in documents [content objects] and for the retrieval of such documents [content objects] from an information storage and retrieval system. [ISO 5127/1]
indexing term	The representation of a concept in an indexing language, generally in the form of a noun or noun phrase. Terms, subject headings, and heading-subheading combinations are examples of indexing terms. Also called descriptor.
information storage and retrieval system	A set of operations and the associated equipment, software, and documentation by which content objects are indexed and the data are stored, so that selected content objects can be retrieved in response to requests employing commands that can be handled by the system.
keyword	A word occurring in the natural language of a document that is considered significant for indexing and retrieval.
lexeme	A fundamental unit of the vocabulary of a language.
literary warrant	Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of its frequent occurrence in the literature. See also *organizational warrant* and *user warrant*.

Term	Definition
modifier	In a compound term, one or more components that serve to narrow the extension of a focus and specify one of its subclasses. Also known as *difference*.
narrower term	A term that is subordinate to another term or to multiple terms in a hierarchy. In thesauri, the relationship indicator for this type of term is NT.
natural language	A language used by human beings for verbal communication. Words extracted from natural language texts for indexing purposes without vocabulary control are often called keywords.
navigation	The process of moving through a controlled vocabulary or an information space via some pre-established links or relationships. For example, navigation in a controlled vocabulary could mean moving from a broader term to one or more narrower terms using the predefined relationships.
near synonym	A term whose meaning is not exactly synonymous with that of another term, yet which may nevertheless be treated as its equivalent in a controlled vocabulary. Example: **salinity**, **saltiness**
node label	A "dummy" term, often a phrase, that is not assigned to documents when indexing, but which is inserted into the hierarchical section of some controlled vocabularies to indicate the logical basis on which a class has been divided. Node labels may also be used to group categories of related terms in the alphabetic section of a controlled vocabulary.
non-preferred term	See *entry term*. See also *preferred term*.
organizational warrant	Justification for the representation of a concept in an indexing language or for the selection of a preferred term due to characteristics and context of the organization. See also *literary warrant* and *user warrant*.
permuted display	A type of index where individual words of a term are rotated to bring each word of the term into alphabetical order in the term list. See also *KWIC* and *KWOC* (in the glossary).
pick list	A graphical user interface device that allows the user to select from a pre-set list of terms. Typically the list of terms is shown when the user clicks on a down arrow next to the entry box for the term.
polyseme	A word with multiple meanings. In spoken language, polysemes are called homonyms; in written language they are called homographs. Only the latter are relevant to controlled vocabularies designed for textual information.
postcoordination	The combining of terms at the searching stage rather than at the subject heading list construction stage or indexing stage. See also *precoordination*.

Term	Definition
precision	A measure of a search system's ability to retrieve only relevant content objects. Usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the total number of content objects retrieved. A high-precision search ensures that, for the most part, the content objects retrieved will be relevant. However, a high-precision search may not retrieve all relevant content objects. See also *recall*. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down.
precoordination	The formulation of a multiword heading or the linking of a heading and subheadings to create a formally controlled, multi-element expression of a concept or object. Precoordination is often used to ensure logical sorting of related expressions. Examples of precoordinated headings: **New England—Genealogy—Handbooks, Manuals, etc.** **Searching, Bibliographic** **United States—History—Civil War, 1861-1865** See also *postcoordination*.
preferred term	One of two or more synonyms or lexical variants selected as a term for inclusion in a controlled vocabulary. See also *non-preferred term*. *NOTE: In the previous version of this Standard, a preferred term was known as a descriptor.*
provisional term	See *candidate term*.
qualifier	A defining term, used in a controlled vocabulary to distinguish homographs. A qualifier is considered part of a term, subject heading, or entry term, but is separated from it by punctuation. The qualifier is generally enclosed in parentheses. Example: **Mercury (metal)**
quasi-synonym	See *near synonym*.
recall	A measure of a search system's ability to retrieve all relevant content objects. Usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the number of all relevant content objects in a collection. A high recall search retrieves a comprehensive set of relevant content objects from the collection. However, high recall increases the possibility that less relevant content objects will also be retrieved. See also *precision*. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down.
reciprocity	Semantic relationships in controlled vocabularies must be reciprocal, that is each relationship from one term to another must also be represented by a reciprocal relationship in the other direction. Reciprocal relationships may be symmetric, e.g. RT / RT, or asymmetric e.g. BT / NT. See also *asymmetric* and *symmetric*.
related term	A term that is associatively but not hierarchically linked to another term in a controlled vocabulary. In thesauri, the relationship indicator for this type of term is RT.

Term	Definition
relationship indicator	A word, phrase, abbreviation, or symbol used in thesauri to identify a semantic relationship between terms. Examples of relationship indicators are UF (USED FOR), and RT (related term).
scope note	A note following a term explaining its coverage, specialized usage, or rules for assigning it.
sibling	A term that shares the same broader term (one level higher) as other terms.
subheading	A term appended to a heading in order to modify or delimit the heading by indicating a particular aspect or relationship pertaining to it. A term with a subheading may be subject to further modification. See also *precoordination*.
subject heading	A word or phrase, or any combination of words, phrases, and modifiers used to describe the topic of a content object. Precoordination of terms for multiple and related concepts is a characteristic of subject headings that distinguishes them from controlled vocabulary terms. See also *precoordinated term* and *precoordination*.
symmetric	Having symmetry. In the context of controlled vocabularies reciprocal relationships are symmetric when the relationship indicator used between a pair of linked terms is the same in one direction as it is in the reverse direction, e.g. RT / RT. See also *asymmetric* and *reciprocity*.
synonym	A word or term having exactly or very nearly the same meaning as another word or term.
synonym ring	A group of terms that are considered equivalent for the purposes of retrieval.
taxonomy	A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy.
term	One or more words designating a concept. See also *compound term, entry term, and precoordinated term.*
thesaurus (plural: thesauruses, thesauri)	A controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators. Relationship indicators ***should*** be employed reciprocally.
top term	The broadest term in a controlled vocabulary hierarchy, sometimes indicated by the abbreviation TT.
tree structure	A controlled vocabulary display format in which the complete hierarchy of terms is shown. Each term is assigned a tree number or line number which leads from the alphabetical display to the hierarchical one. The hierarchical display is also known as *systematic display*.

Term	Definition
user warrant	Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of frequent requests for information on the concept or free-text searches on the term by users of an information storage and retrieval system. See also *literary warrant* and *organizational warrant*.
vocabulary control	The process of organizing a list of terms (a) to indicate which of two or more synonymous terms is authorized for use; (b) to distinguish between homographs; and (c) to indicate hierarchical and associative relationships among terms in the context of a controlled vocabulary or subject heading list. See also *controlled vocabulary*.

4.2 Abbreviations and Acronyms

The following abbreviations and acronyms, as used in this Standard, have the meanings indicated.

Abbreviation / Acronym	Definition
BT	Broader Term
BTG	Broader Term (generic)
BTI	Broader Term (instance)
BTP	Broader Term (partitive)
HN	History Note
NT	Narrower Term
NTG	Narrower Term (generic)
NTI	Narrower Term (instance)
NTP	Narrower Term (partitive)
RT	Related Term
SN	Scope Note
TT	Top Term
U	USE
UF	USED FOR
X SN	See Scope Note for…

5 Controlled Vocabularies – Purpose, Concepts, Principles, and Structure

5.1 Purpose

The purpose of controlled vocabularies is to provide a means for organizing information. Through the process of assigning terms selected from controlled vocabularies to describe documents and other

types of content objects, the materials are organized according to the various elements that have been chosen to describe them.

Controlled vocabularies serve five purposes:

1. **Translation:** Provide a means for converting the natural language of authors, indexers, and users into a vocabulary that can be used for indexing and retrieval.
2. **Consistency:** Promote uniformity in term format and in the assignment of terms.
3. **Indication of relationships:** Indicate semantic relationships among terms.
4. **Label and browse:** Provide consistent and clear hierarchies in a navigation system to help users locate desired content objects.
5. **Retrieval:** Serve as a searching aid in locating content objects.

5.2 Concepts

5.2.1 Terms

Controlled vocabularies are generally used to describe content by assigning terms to represent metadata associated with content objects, organizing content on websites, and the like.

For the purposes of this Standard, a term is defined to be one or more words used to represent a concept. Terms are selected from natural language for inclusion in a controlled vocabulary. Section 6 provides detailed rules and guidelines for the selection and formatting of terms.

In this Standard, the word “term” is used to refer generically to both “preferred terms” and “entry terms” (also called non-preferred terms).

5.2.2 Content Objects

A content object is any item that is to be described for inclusion in an information retrieval system, website, or other source of information. Typical content objects are journal articles, technical reports, and other types of documents. Content objects ***may*** exist solely in electronic form and also ***may*** be entities such as paintings, sculpture, and realia such as Dr. Seuss’s hat.

There are two classes of content objects, primary and secondary, although this distinction is rarely made.

- A primary content object is the item itself, whether it exists in physical form (e.g. print, audiotape, DVD, movie) or exists solely in electronic form.
- A secondary content object is the metadata that describes the primary content object. Many data stores combine the primary content object and its metadata into a single, hybrid content object.

Figure 1 illustrates this concept.

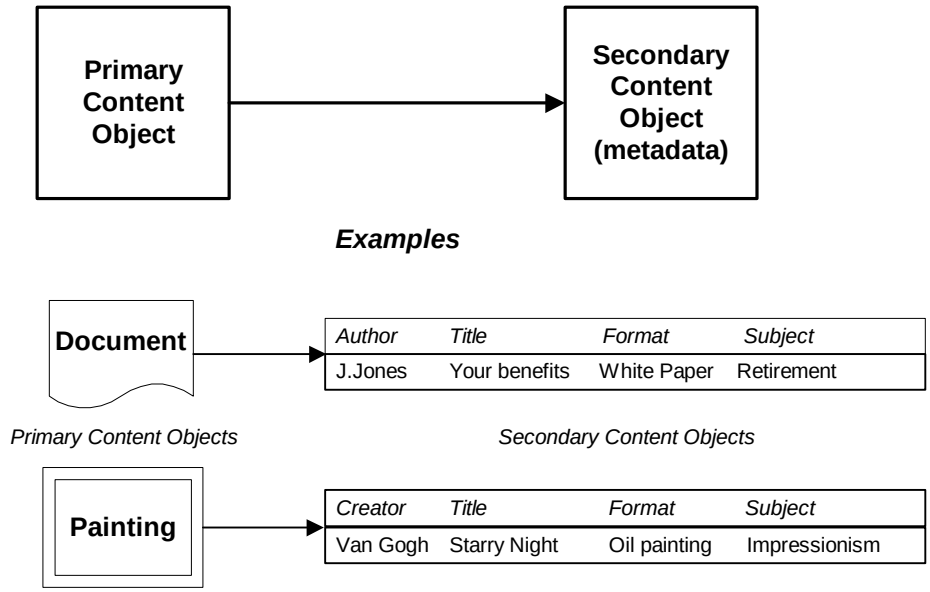


Figure 1: Content objects

5.2.3 Indexing

Indexing is the process of assigning preferred terms or headings to describe the concepts and other metadata associated with a content object. Indexing covers any system or procedure in which the selection and organization of terms requires human intellectual decisions at some point in the process. Computer processing ***may*** also be a part of the process for storing and manipulating the terms in a controlled vocabulary or to identify content objects to which certain terms or combinations of terms have been assigned or ***should*** be assigned.

The process of indexing, therefore, involves selecting preferred terms from one or more controlled vocabularies or other sources to describe a content object.

The effectiveness of indexing as a means for identifying and retrieving content objects depends upon a well-constructed indexing language. Research in the field of information science has shown that controlled vocabularies improve both precision and recall in searching. For example, they improve precision by defining the scope of terms and they increase recall by retrieving documents that employ different terms for the same concept.

5.3 Principles

There are four important principles of vocabulary control that guide their design and development. These are:

- eliminating ambiguity
- controlling synonyms
- establishing relationships among terms where appropriate
- testing and validation of terms

A major goal of vocabulary control is to ensure that each distinct concept is described by a single linguistic form. If multiple forms exist, these ***should*** be controlled or regularized so that information or content that is provided to a user is not spread across the system under multiple access points, but is gathered together in one place. Eliminating ambiguity and compensating for synonymy through vocabulary control assures that each term has only one meaning and that only one term can be used to represent a given concept or entity.

Note that a very large organization may use the same terms in different contexts. In such cases, care must be taken to ensure that users—both indexers and searchers—can make the necessary distinctions among terms.

5.3.1 Ambiguity

Ambiguity occurs in natural language when a word or phrase (a homograph or polyseme) has more than one meaning. Figure 2 provides an example and shows how a single word can be used to represent multiple, very different concepts.

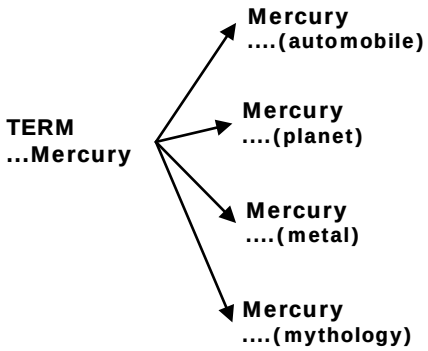


Figure 2: Ambiguity caused by homographs and polysemes

A controlled vocabulary ***must*** compensate for the problems caused by ambiguity by ensuring that each term has one and only one meaning.

5.3.2 Synonymy

Synonymy occurs when a concept can be represented by multiple terms having the same or similar meanings. The desired content may be difficult to retrieve because it can be described by different but equivalent terminology. Figure 3 illustrates this case

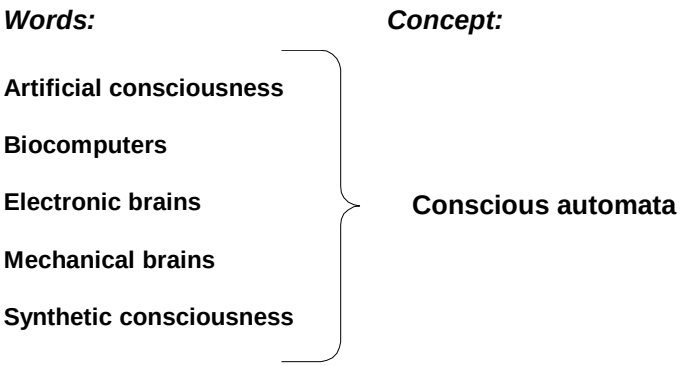


Figure 3: Information scatter caused by synonyms

A controlled vocabulary ***must*** compensate for the problems caused by synonymy by ensuring that each concept is represented by a single preferred term. The vocabulary ***should*** list the other synonyms and variants as non-preferred terms with USE references to the preferred term.

Note: A synonym ring is an exception to the above rule. See section 5.4.2 for more information on this type of

vocabulary.

There are other types of “equivalent” terms besides synonyms which require vocabulary control. Section 8.2 includes a full discussion of equivalence control.

5.3.3 Semantic Relationships

Various types of semantic relationships ***may*** be identified among the terms in a controlled vocabulary. These include equality relationships, hierarchical relationships, and associative relationships, which ***may*** be defined as required for a particular application. Section 8 includes a full discussion of the various types of relationships that may be included in controlled vocabularies.

5.3.4 Facet Analysis

Controlled vocabularies—especially large ones consisting of thousands of terms—may be easier to use if they are organized in some way other than hierarchically. Faceted analysis, based on the work of Ranganathan and refinements by the Classification Research Group, is another way of organizing knowledge. It takes a bottom-up approach, forming areas of knowledge after first having pieced together their parts and determining the areas of knowledge they form, rather then the discipline-directed, top-down approach of hierarchies. Facet analysis is particularly useful for:

- new and emerging fields where there is incomplete domain knowledge or where relationships between the content objects are unknown or poorly defined;
- interdisciplinary areas where there is more than one perspective on how to look at a content object or where combinations of concepts are needed;
- vocabularies where multiple hierarchies are required but can be inadequate due to difficulty in defining their clear boundaries; or
- classifying electronic documents and content objects where location and collocation of materials is not an important issue.

Facet analysis is sometimes used to indicate the attributes of content objects as shown in Figure 4.

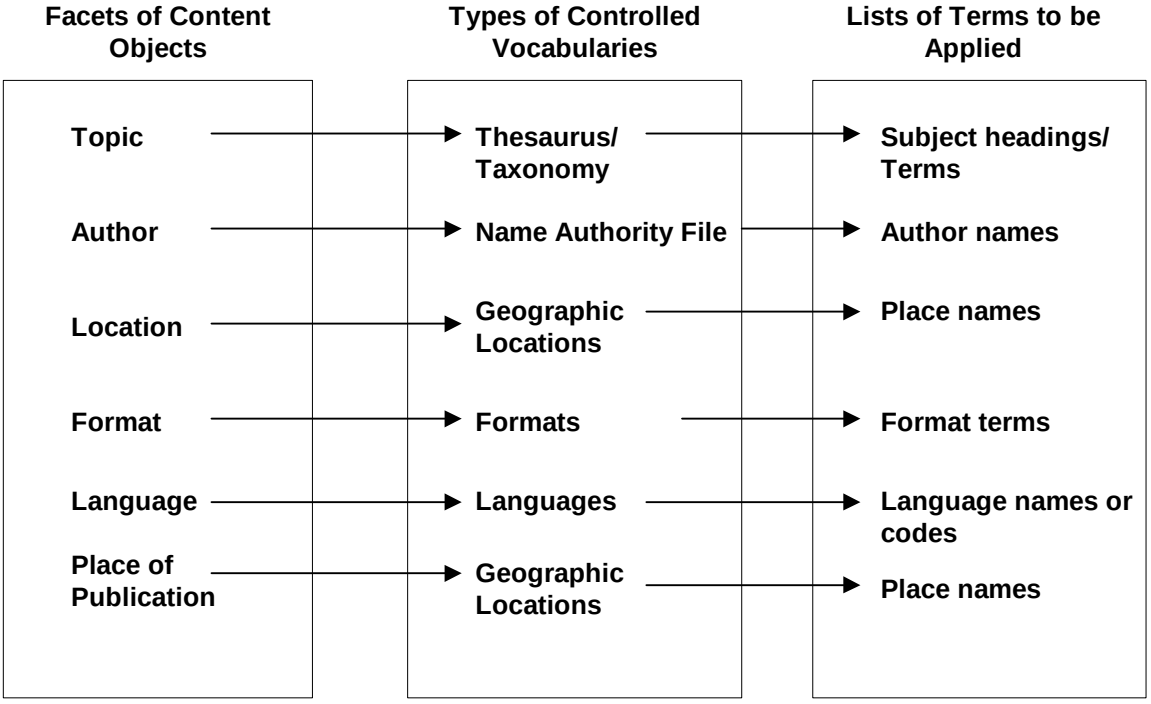


Figure 4: Facet analysis applied to content objects and controlled vocabularies

As shown in Figure 4, facets assigned to content objects represent one or more of the attributes that can be used to locate the item by searching or browsing. For example, *Format* appears in the diagram as a *facet* of a content object. Terms belonging to this facet might be the various format indicators; for example:

Format:
Print
CD-ROM
Online

Facets are a kind of structural metadata. They ***may*** be applied (as indicated in the diagram above) either to content objects themselves or to the metadata that represent them.

Attributes that might be selected as facets for content objects are:

- **Topic** – the subject of the content object
- **Format** – the format of material (e.g., text, image, sound, etc.).
- **Target audience** – the appropriate reader for the content (e.g., Children, Adults)
- **Intellectual level** – the type of the content (e.g., K-12, news, scholarly journal).

Information retrieval systems ***may*** use facets to represent at least part of the structure of the underlying database. That is, facets ***may*** be defined that correspond to some of the attributes of the content objects. In a printed or online tool, such as a catalog, for example, the facets might appear as the generic terms used to organize the more specific information.

Whether to define facets for content objects is the responsibility of the database designer or information architect. This Standard does not address guidelines for determining which aspects or attributes to create as facets for content objects.

Whether to use facet analysis for the controlled vocabulary itself depends to a large extent on the size of the vocabulary. In some cases, a controlled vocabulary will consist of only a small set of terms such as those that would be needed to describe *Content type* or *Intellectual Level***.** If the number of terms is small (less than a hundred terms, for example), a simple list will suffice and there is no need to organize them further.

In the case where the controlled vocabulary contains many hundreds or thousands of terms, it ***may*** be helpful to organize them according to facets**.** This type of facet analysis results in identifying the basic broad categories of the vocabulary and grouping the terms under them. In a taxonomy, for example, facets might be used to determine separate hierarchies for a Web navigation scheme.

The Art and Architecture Thesaurus® (AAT) is an example of a controlled vocabulary using facets. The following seven facets are used to categorize the entire vocabulary:

- Associated Concepts Facet
- Physical Attributes Facet
- Styles and Periods Facet
- Agents Facet
- Activities Facet
- Materials Facet
- Objects Facet

See section 9.3.4.6 for an example of a faceted display.

5.3.5 Using Warrant to Select Terms

The process of selecting terms for inclusion in controlled vocabularies involves consulting various sources of words and phrases as well as criteria based on:

- the natural language used to describe content objects (literary warrant),
- the language of users (user warrant), and
- the needs and priorities of the organization (organizational warrant).

5.3.5.1 Literary Warrant

Assessing literary warrant involves reviewing the primary or secondary content objects that the vocabulary will be used to index as well as consulting reference sources such as dictionaries or textbooks and existing vocabularies for the content domain. The word or phrases chosen ***should*** match as closely as possible the prevailing usage in the domain's literature.

5.3.5.2 Organizational Warrant

Determining organizational warrant requires identifying the form or forms of terms that are preferred by the organization or organizations that will use the controlled vocabulary.

5.3.5.3 User Warrant

User warrant is generally reflected by the use of terms in requests for information on the concept or from searches on the term by users of an information storage and retrieval system.

5.3.6 Controlled Vocabulary Impact on Information Retrieval

Information retrieval effectiveness is traditionally measured by two parameters: recall and precision. (See Definitions for an explanation of these measures.) Controlled vocabulary design can have a positive impact on both of these measures.

Recall can be improved through such controlled vocabulary methods as:

- Preferred terms and equivalence relationships for synonym control (see section 5.3.2)
- Preferred term form (see section 6.3)
- Associative (related term) relationships (see section 8.4)
- Classified and hierarchical relationships (see section 8.3)
- Postcoordination (see section 7.2.2.1)
- Concept mapping / clustering (see section 9.3.5)

Precision can be improved through such controlled vocabulary methods as:

- Parenthetical qualifiers to control ambiguity (see section 6.2.1)
- Broader and narrower term hierarchical relationships (see section 8.3)
- Compound terms (see section 7)
- Precoordination (see section 7.2.2.2)

5.4 Structure

Controlled vocabularies are structured to enable displaying the different types of relationships among the terms they contain. There are four different types of controlled vocabularies, determined by their increasingly complex structure. These are:

- List
- Synonym ring

- Taxonomy
- Thesaurus

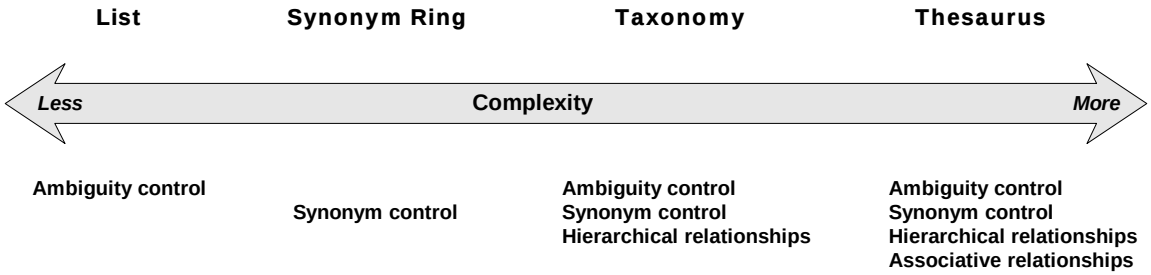


Figure 5: Increasing structural complexity among controlled vocabularies

This figure shows the increasingly complex structure of controlled vocabularies dictated by the requirements of the types of relationships each ***must*** accommodate. It also shows that the more complex vocabularies (taxonomies, thesauri) include the simpler structures (lists, synonym rings). For example, a Thesaurus includes explicit devices for controlling synonyms, arranging hierarchies, and creating associative relationships while a List is a simple set of terms containing no relationships of any kind.

Sections 5.4.1 through 5.4.4 describe and provide examples of each of the vocabulary types. Refer to Appendix B for a comparison of the four types.

5.4.1 List

A list (also sometimes called a pick list) is a limited set of terms arranged as a simple alphabetical list or in some other logically evident way. Lists are used to describe aspects of content objects or entities that have a limited number of possibilities. Examples include geography (e.g., country, state, city), language (e.g., English, French, Swedish), or format (e.g., text, image, sound).

Example 1: Simple alphabetical list

Alabama
Alaska
Arkansas
California
Connecticut
Delaware

Example 2: Simple logical list

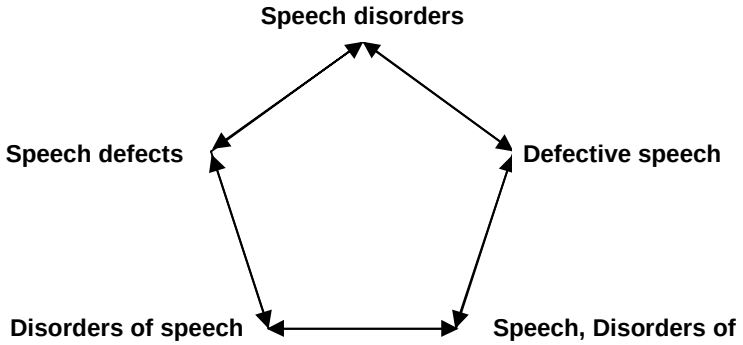
Mercury
Venus
Earth
Mars
Jupiter
Saturn
Uranus
Neptune
Pluto

5.4.2 Synonym Ring

While a synonym ring is considered to be a type of controlled vocabulary, it plays a somewhat different role than the other types covered by this Standard. Synonym rings cannot be used during the indexing process. Rather, they are used only during retrieval. Use of synonym rings ensures that a concept that can be described by multiple synonymous or equivalent terms will be retrieved if any one of the terms is used in a search.

A synonym ring, therefore, is a set of terms that are considered equivalent for the purposes of retrieval. Synonym rings usually occur as sets of flat lists. A synonym ring allows users to access all content objects or database entries containing any one of the terms. Synonym rings are generally used in the interface in an electronic information system, and provide access to content that is represented in natural, uncontrolled language.

Example 3: Synonym ring for speech disorders



5.4.3 Taxonomy

A taxonomy is a controlled vocabulary consisting of preferred terms, all of which are connected in a hierarchy or polyhierarchy.

Example 4: Taxonomy hierarchy

chemistry
physical chemistry
electrochemistry
magnetohydrodyamics

5.4.4 Thesaurus

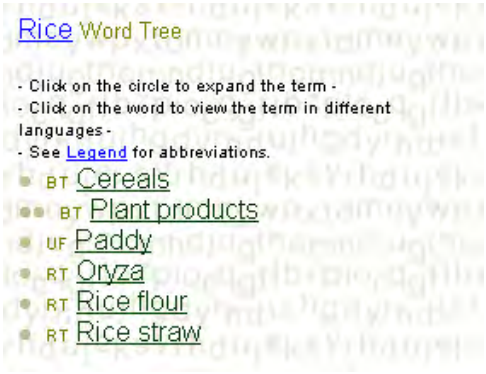
A thesaurus is a controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators. Relationship indicators ***should*** be employed reciprocally.

Example 5: Print thesaurus entry

ABSORPTION
The retention and conversion into another form of energy of rays, waves, or particles by a substance.
UF ABSORPTIVE PROPERTIES
BT **SORPTION**
NT **BIOLOGICAL ABSORPTION**
RESONANCE ABSORPTION
TWO PHOTON ABSORPTION
X RAY ABSORPTION ANALYSIS

Source: DTIC Thesaurus [2]

Example 6: Online thesaurus entry



Source: AGROVOC [1]

5.5 Controlled Vocabulary Metadata and Metadata Schemas

Metadata, data about data, has become increasingly important to discovery and retrieval of electronic information, especially over the Internet. Various metadata element sets (collections of predefined properties for describing a resource) have been created to support particular communities or information domains. These element sets specify semantic rules for each element, which can include the recommended or required use of a controlled vocabulary for the values assigned to an element. *The Dublin Core Element Set* [24], is the most well known and generalized element set.

Metadata schemas extend metadata element sets by adding rules for syntax—form and structure—and in many cases encoding rules. The use of metadata schemas can facilitate interoperability between systems, including the interoperability of controlled vocabularies.

Metadata can be used with controlled vocabularies in several ways:

1. Using a controlled vocabulary as the source for allowed terms for a particular metadata element.

   Most existing metadata element sets support subject-related metadata through a "Keyword" or "Subject" field. Often the element set, or a particular community's implementation of the element set, will indicate a controlled vocabulary that ***should*** or ***must*** be used for a particular metadata element. In this instance, the controlled vocabulary is being used to select descriptive metadata about the content resource.

2. Using metadata to describe a controlled vocabulary as a whole for resource discovery.

   This use of metadata is no different than describing any other type of resource. Many existing metadata element sets and schemas could be utilized to describe a controlled vocabulary resource.

3. Using metadata and a metadata schema to represent the entire contents of the controlled vocabulary.

   This use of metadata is generally designed to facilitate searching or exchange of the controlled vocabulary. It requires at a minimum: a metadata element set to describe the concepts, terms, and relationships; a set of relationship definitions; and a metadata schema for representing relationships among terms.

There are a number of initiatives underway to address the third type of metadata. A list of these is provided in the Bibliography. At this time, none of the initiatives have advanced to the stage where a best practice standard recommendation can be made.

Controlled vocabulary developers ***should*** evaluate these metadata initiatives for relevance to their community and information domain prior to developing any new metadata element sets or schema.

6 Term Choice, Scope, and Form

This section covers the principles of term selection and the determination of the correct form of each term.

NOTE: In this section, the phrase rather than *is used to contrast preferred and non-preferred terms. Where two synonyms are juxtaposed without indication of a preference, they are separated by a slash.*

6.1 Choice of Terms

Selecting terms for inclusion in a controlled vocabulary is one of the most important factors in creating a product that has broad user acceptance. Indexers and searchers primarily, are affected by the choice of terms. Many issues need to be considered in selecting terms for a controlled vocabulary including:

- The information space or domain to which the vocabulary will be applied (section 11.1.1)
- Literary, user, and organizational warrant (section 5.3.5)
- Specificity or granularity of the terms (section 11.1.7)
- Relationship with other, related controlled vocabularies (section 10.9)

Matters of term selection are covered in detail in the sections referenced above.

6.2 Scope of Terms

The scope of terms is restricted to selected meanings within the domain of the controlled vocabulary. Each term ***should*** be formulated in such a way that it conveys the intended scope to any user of the controlled vocabulary. The rules defined in this section ***should*** be applied as the hierarchies are developed.

Terms whose meanings overlap in general usage and homographs (terms with identical spellings but different meanings) ***should*** be avoided as far as possible in the selection of terms. Literary warrant, i.e., the occurrence of terms in content objects, and other criteria ***may***, however, justify the use of such terms. Section 6.2.1 covers methods of disambiguating homographs. Section 6.2.2 explains the use of scope notes for terms whose meanings overlap.

6.2.1 Homographs

The use of homographs as terms in a controlled vocabulary sometimes requires clarification of their meaning through a qualifier (a "gloss" in linguistic terminology). Such a gloss specifies the domain of meaning to which the term belongs. A qualifier is a word or words used with a term to make the specific meaning unambiguous. The qualifier, which is enclosed in parentheses, is part of the term.

Use of qualifiers ***should*** be avoided whenever possible because of the problems that parentheticals can cause in filing and in retrieval. See section 9.2.6 for more on filing and sorting.

Rules for using qualifiers:

a) A compound term ***should*** be used instead of a single-word term with a parenthetical qualifier, if usage permits, i.e., if the compound occurs in natural language.

Example 7: Compound terms rather than qualifiers

phonograph records *rather than* records (phonograph)
religious tolerance *rather than* tolerance (religion)

NOTE: In both of these examples, compound terms are natural expressions that can be selected for controlled vocabularies in the fields of music and religion, respectively.

b) A qualifier ***should*** be added to each homograph, even when one is used in the primary sense of the domain and the second in a different sense. For example, cranes (lifting equipment) should be the term in an engineering controlled vocabulary that also includes cranes (birds).

c) A homograph's parenthetical qualifier ***may*** be left out when a term is used in only one of its meanings within a given controlled vocabulary domain and the meaning is obvious to users within that domain. However, if the vocabulary will be used across domains, e.g. in a multidisciplinary retrieval system, the qualifier ***should*** be included as it facilitates cross-database searching and mapping of terms in disparate domains.

Example 8: Qualifiers indicating domain

developing (photography)
organism (philosophy)
translation (genetics)

d) Parenthetical qualifiers ***should not*** be used to represent compound concepts, e.g., cookery (zucchini) or pipes (plastic). Plastic is used in the latter example to indicate a type of pipe rather than to disambiguate the word "pipe". Appropriate uses of qualifiers with the term pipes would be: pipes (musical instruments) and pipes (smoking implements). (See section 7 for guidelines on formulating compound terms.)

e) The qualifier itself may be a term, often a broader term than the one being qualified. It ***should*** be as brief as possible, ideally consisting of one word, but ***should not*** be a homograph.

f) Qualifiers ***should*** be standardized within a given controlled vocabulary to the extent possible, e.g., biology and bioscience ***should not*** both be used as qualifiers.

g) A qualifier is not a scope note. However, a qualified term ***may*** have a scope note appended to it. For guidelines on scope notes, see section 6.2.2.

h) Qualifiers ***should*** also be added to entry terms when their meaning is ambiguous.

Example 9: Qualifiers to distinguish ambiguous homographs

cranes (birds)
cranes (lifting equipment)

mercury (metal)
Mercury (planet)
Mercury (Roman deity)

seals (animals)
seals (law)
seals (numismatics)

socialization (economics)
socialization (social psychology)

6.2.2 Scope Notes

A scope note is used:

- to restrict or expand the application of a term,
- to distinguish between terms that have overlapping meanings in natural language, or
- to provide other advice on term usage to either the indexer or the searcher.

A scope note ***should*** state the chosen meaning of a term. It ***may*** also indicate other meanings that are recognized in natural language, but which have been deliberately excluded from the controlled vocabulary.

Unlike a parenthetical qualifier, a scope note is not part of a term. While qualifiers are generally added only to homographs, a scope note (SN) can be supplied for any term.

Example 10: Scope Note

illuminations

SN Includes both the ornamental decoration and the illustrations in manuscripts as well as in some early printed books, if done by hand.

The scope of terms is indicated additionally through the semantic relationships represented in the controlled vocabulary (see section 8.1). Changes in the scope of terms ***should*** be recorded in History Notes (see section 6.2.3).

6.2.2.1 Reciprocal Scope Notes

When reference is made to other terms in a scope note, a reciprocal scope note ***should*** generally be provided for each term mentioned.

Example 11: Reciprocal Scope Notes

dogtrots

SN Passages sharing a roof common with the rest of a building, connecting two parts of a log house of the American folk art tradition. Distinguished from **breezeways** by its folk architecture tradition and log house context and its common roof.

breezeways

SN Roofed passages connecting two parts of a house or a house and a garage; common after 1930. Distinct from **dogtrots**, which occur in folk architecture log houses.

Even where the scope of only one of the terms requires clarification, it is useful to note in the term record for the second term that it has been cited in a scope note of a different term.

Example 12: Scope Note cross-reference

information science

X SN **library science**

The X indicates that there is a reference from the scope note of **library science** to **information science**. This reciprocal reference will ensure that when a change is made to one of the terms, or it is deleted, the effect on the other term will be considered.

6.2.3 History Notes

A History Note is used to track the development of terms over time. These notes provide important guidance for researchers who are interested in a topic covering many decades. It is especially important to indicate when and how a term has changed over time. If appropriate, the history note ***may*** also include the date discontinued, the term that succeeded the term, and/or the term that preceded it. History Notes are frequently marked by the abbreviation HN. (See also sections 11.3 and 11.3.2.2 for rules on when to use History Notes.)

Example 13: History Note

Palladian windows--[country or state]--[city]

Public Note

Large windows characteristic of neoclassic styles, divided by columns or piers resembling pilasters, into three lights, the middle one of which is usually wider than the others, and is sometimes arched.

History Note

Changed 1/1993 from non-postable term (Use **VENETIAN WINDOWS**).

Used For

Diocletian windows
Tripartite windows
Venetian windows

Broader Term

Windows

Hierarchy
<<<(3)**Architectural & site components**
<<(2)**Architectural elements**
<(1)**Windows**

Source: Thesaurus for Graphic Materials I Subject Terms (TGM I) [13]

6.3 Term Form

6.3.1 Single-Word vs. Multiword Terms

Each term included in a controlled vocabulary ***should*** represent a single concept (or unit of thought). A single concept is frequently expressed by a single-word term but in many cases a multiword term is required to represent the concept. There are extensive rules that govern the choice of single-word vs. multi-word terms. A multiword term for a single concept is a compound term. Because of the difficulty in defining "single concept," objective criteria for dealing with compound terms are provided in section 7.

Example 14: Terms representing single concepts

cathedrals
stained glass
flying buttresses

6.3.2 Types of Concepts

The concepts represented by terms ***may*** be grouped into general types. The compiler of a controlled vocabulary needs to be aware of these types, because the type of concept can affect some of the procedures used in controlled vocabulary construction; for example, the choice of singular or plural form (see section 6.5) and applying a test for the validity of a hierarchy (see section 8.3). The types of concepts shown in the following examples are not exhaustive.

Example 15: Terms for things and their physical parts

birds
carburetors
microforms
mountains
oil paintings
teddy bears

Example 16: Terms for materials

adhesives
mustard gas
oxygen
paints
water

Example 17: Terms for activities or processes

acidification
archery
painting
parenting
scuba diving
sewing

Example 18: Terms for events or occurrences

birthdays
civil wars
holidays
revolutions

Example 19: Terms for properties or states

consciousness
elasticity
personality
speed
texture

Example 20: Terms for disciplines or subject fields

anthropology
information science
organic chemistry
theology

Example 21: Terms for units of measurement

hertz
kilometers

6.3.3 Unique Entities

Unique entities, or “classes-of-one,” are usually expressed as proper nouns.

Example 22: Proper nouns as unique entities

Aristotle
Earth
Fourth of July
Library of Congress
Magna Carta
World Health Organization
Zimbabwe

6.4 Grammatical Forms of Terms

The following sections indicate the preferred grammatical forms of terms. Entry terms from non-preferred grammatical forms ***may*** be provided.

6.4.1 Nouns and Noun Phrases

The grammatical form of a term ***should*** be a noun or noun phrase. For an exception to this recommendation, see section 6.4.2.

Example 23: Nouns or noun phrases

accelerator mass spectrometry
drawings
liver
space shuttles

6.4.1.1 Verbal Nouns

Verbs expressed as infinitives (without "to") or participles ***should not*** be used alone as terms. Activities ***should*** be represented by nouns or gerunds.

Example 24: Verbal nouns

catalysis *rather than* catalyze
cookery *rather than* cook
freezing *rather than* freeze
reading *rather than* read
distillation *rather than* distill

6.4.1.2 Noun Phrases

Noun phrases are compound terms that ***may*** be established as terms if they represent a single concept (see section 7). Noun phrases occur in two forms:

- premodified, or adjectival; and
- postmodified, or prepositional.

6.4.1.2.1 Premodified Noun Phrases

Adjectival noun phrases (premodified) are the preferred form.

Example 25: Adjectival noun phrases

African American women authors
basal metabolism
cold fusion
historical drama
public television
rapid transit

6.4.1.2.2 Postmodified Noun Phrases

Prepositional noun phrases (postmodified) are also allowed.

When possible, noun phrases ***should*** exclude prepositions; for example, use **carbohydrate metabolism** rather than **metabolism of carbohydrates**, and **children's hospitals** rather than **hospitals for children**. Terms in the form of prepositional noun phrases ***should*** be restricted to concepts that cannot be expressed in any other way, or that have become idiomatic.

Example 26: Prepositional noun phrases

accessories after the fact
burden of proof
cream of tartar
coats of arms
plaster of Paris
prisoners of war
sergeants-at-arms

6.4.2 Adjectives

Adjectives and adjectival phrases used alone ***may*** be established as terms in controlled vocabularies under certain special circumstances.

6.4.2.1 Single Adjectives

Single adjectives ***may*** be used in a "nominal" way; that is, the noun is obvious from the context or the adjective is used to describe an attribute of the content object other than topic, such as **color** or **size**.

Example 27: Single adjectives

small
medium
large

blue
green
red
yellow

6.4.2.2 Adjectives Used to Limit the Number of Compound Terms

As an alternative to the creation of multiple compound terms, adjectives ***may*** appear as separate terms when designed to be precoordinated in indexing or postcoordinated in searching. They ***should*** generally not be assigned as indexing terms in isolation. Given the possibility of false coordination in searching (e.g., the linking of an adjective with the wrong noun), adjectival terms ***should*** be used sparingly.

Example 28: Adjectives as terms in pre- and postcoordination

Adjectives used as terms	Precoordinated indexing terms
airborne	**airborne troops**
mobile	**mobile homes**
offshore	**offshore drilling**
portable	**portable heaters**
prestressed	**prestressed concrete**

Certain noun phrases ***may*** be used to modify other nouns, e.g., **high frequency** can modify the noun **waves**. The guidelines for adjectives ***may*** be applied to such noun phrases.

6.4.2.3 Adjectives Used as Cross-references

Adjectives ***may*** be used alone in general cross references to direct the user to or from a group of terms beginning with a corresponding noun, e.g., "**cardiac** . . . see also the terms beginning with **heart**." An example of a reference in the opposite direction (noun to adjective) is: "**France** see also the terms beginning with **French** (French art, French language, French literature, French wines)."

This guideline applies especially when the adjective and the noun to which it is related differ widely in their derivation and spelling, for example, **sea** and **marine**, or **ear** and **aural**. It works well in printed controlled vocabularies with a limited number of terms beginning with one of the terms. However, explicit references from all compound forms (e.g., **cardiac diseases** USE **heart diseases**) ***may*** be necessary in controlled vocabularies displayed on a screen, or in printed tools in which the compound terms beginning with a particular adjective span several columns or pages.

6.4.3 Adverbs

Single adverbs are rarely used. However, they ***may*** be needed in some domains such as music.

Example 29: Adverbs

legato
staccato

Adverbs such as "very" or "highly" ***should not*** be used alone as terms. A phrase beginning with such an adverb ***may*** be accepted as a term only when it has acquired a specialized meaning within a domain.

Example 30: Adverbial phrases

very high frequency
very large scale integration
very low density lipoproteins

6.4.4 Initial Articles

Initial articles in terms ***should*** be avoided except when required to convey essential information. The choice of whether to retain the initial article ***may*** also be dependent on the language of the controlled vocabulary and/or the language of the term. For example, the initial article is required for the term **El Salvador** but ***may*** be omitted for the term **The Big Easy.**

6.4.4.1 When to Remove Initial Articles

Remove the initial article when the term is clear without it. Otherwise, use a parenthetical qualifier.

Example 31: Removal of initial articles

arts *rather than* The arts
state (political entity) *rather than* The state
Winter's Tale *rather than* The Winter's Tale

6.4.4.2 When to Keep Initial Articles

If the initial article is an integral part of a proper name, and should be searchable, it ***should*** be included in the term in direct order. Otherwise, omit the article. Whether or not the initial article is included as an integral part of the name is dependent upon the language. In the following examples, English is the assumed language of the controlled vocabulary.

Example 32: Retention of initial articles

A Big Company [initialism: ABC]
El Niño
El Salvador

In cases where the article is included in direct order, a cross-reference from the element following the article ***should*** be provided, e.g., **Salvador** USE **El Salvador**.

6.5 Nouns

Nouns used as terms are divided into two categories:

- count nouns (see section 6.5.1) and
- noncount (mass) nouns (see section 6.5.2).

The guidelines for singular and plural based on these categories apply to the formulation of both terms and entry terms.

6.5.1 Count Nouns

Count nouns are names of objects or concepts that are subject to the question "How many?" but not "How much?". These ***should*** normally be expressed as plurals.

Example 33: Plural count nouns

books
chemical reactions
penguins
political parties
singers
vertebrates
windows

6.5.1.1 Exception to Plural Count Nouns

If in the domain of the controlled vocabulary there is literary or user warrant for the expression of count nouns in the singular, establishment of terms in that form is acceptable. For example, in the field of biomedicine, the names of parts of the body are generally formulated in the singular.

Example 34: Exceptions to plural count nouns in biomedicine

ear
eye
lung
nose
stomach

In a museum catalog, objects are typically treated as unique items, and terms are given in the singular.

Example 35: Exceptions to plural count nouns for art objects

chair
oil painting
tapestry

There are a few special circumstances where both the singular and plural forms may be needed, such as when the word forms are used to distinguish process from product.

Example 36: Use of both singular and plural forms

casting [process]
castings [product]

6.5.2 Mass Nouns

Mass (noncount) nouns are names of materials or substances that are subject to the question "How much?" but not "How many?". These ***should*** be expressed in the singular.

Example 37: Singular mass nouns

copper paint [but compare **paints**, below]
snow water

If the community of users served by a controlled vocabulary regards a given substance or material as a class with more than one member, the class ***should*** be expressed in the plural.

Example 38: Exceptions to singular mass nouns

plastics [i.e., various types of plastic]
paints

6.5.3 Other Types of Singular Nouns

6.5.3.1 Abstract Concepts

The names of abstract concepts, e.g., systems of belief, activities, emotions, properties, and disciplines, ***should*** be expressed in the singular, even though some of these terms are subject to the question "How much?".

Example 39: Singular abstract concepts

beliefs	**Judaism; Taoism**
activities	**digestion; distribution;**
or processes	**migration; welding, writing**
emotions	**anger; envy; love; pity**
properties	**conductivity; silence;**
or states	**worth;**
disciplines	**architecture; business; musicology; metallurgy**

6.5.3.2 Unique Entities

The names of unique entities, whether concrete or abstract, ***should*** be expressed in the singular.

Example 40: Singular unique entities

Shangri-La
Statue of Liberty

6.5.4 Coexistence of Singular and Plural Forms

Where the singular and plural forms of a term represent different concepts, separate terms for each ***should*** be entered in the controlled vocabulary as appropriate. The distinction ***should*** be indicated by a qualifier.

Example 41: Coexistence of singular and plural forms using qualifiers

bridge (game)
bridges (dentistry)
bridges (structures)

damage (injury)
damages (law)

wood (material)
woods (forested areas)

6.6 Selecting the Preferred Form

The authority for the form selected ***should*** be recorded in the term record (see section 11.1.4). The variant forms ***should*** be included as entry terms. The source(s) of variant forms ***may*** also be recorded.

6.6.1 Usage

Terms ***should*** reflect the usage of people familiar with the domain of the controlled vocabulary as reflected in literary, organizational, and user warrant (see section 5.3.5). Neutral terms ***should*** be selected, e.g., **developing nations** rather than **underdeveloped countries**. (See also section 6.6.4.)

6.6.1.1 Usage – Literary Warrant

Words and phrases drawn from the literature of the field ***should*** determine the formulation of terms. When two or more variants have literary warrant, the most frequently used term ***should*** be selected as the term. Guidance in the selection of terms ***may*** be found in reference works of the domain, such as dictionaries, glossaries, encyclopedias, and authoritative treatises of the field. Other controlled vocabularies, abstracting and indexing services, and subject heading lists ***may*** be useful as sources for selection of preferred forms. Opinions of subject experts regarding the preferred form of terms ***may*** be sought.

6.6.1.2 Usage – Organizational Warrant

The usage of terms within the organization(s) using the vocabulary can be an aid in deciding which term is a preferred term and which equivalent terms will become non-preferred entry terms.

6.6.1.3 Usage – User Warrant

User warrant can be utilized to enhance completeness of the vocabulary by organizing candidate terms into broad categories based on user preferences. Users ***should*** be asked to review drafts of the vocabulary to add missing terms, identify terms that are incorrect or obsolete, create more useful term forms, and validate relationships among terms.

6.6.2 Spelling

6.6.2.1 Spelling – Warrant

The most widely accepted spelling of words, based on warrant, ***should*** be adopted. If variant spellings exist and are commonly recognized, each ***should*** be entered in the controlled vocabulary, and a cross-reference ***should*** be made from the non-preferred to the preferred form.

Example 42: Using warrant to determine spelling

Romania *rather than* Roumania
theater *rather than* theatre [for American English]

6.6.2.2 Spelling – Authorities

Spelling ***should*** follow the practice of well established dictionaries or glossaries. If a choice between spellings is made for dialectal reasons (for example, between American and British English), the choice ***should*** be adhered to consistently throughout the controlled vocabulary; exceptions ***may*** be made for proper names.

Example 43: Spelling based on dialectical authority

NOTE: This example assumes the authority is American English rather than British English.

catalogs *rather than* catalogues
color *rather than* colour
labor *rather than* labour *but* **Labour Party**

6.6.3 Abbreviations, Initialisms, and Acronyms

This section deals with the choice of full vs. abbreviated forms of terms. Capitalization is covered in section 6.7.1.

6.6.3.1 Preference for Abbreviation

Abbreviations and acronyms ***should*** be selected as terms only when they have become so well established that the full form of the term or proper name is rarely used. Cross-references ***should*** be made from the full forms.

Example 44: Abbreviations and acronyms

AIDS *rather than* acquired immune (disease) deficiency syndrome
DNA *rather than* deoxyribonucleic acid
lasers *rather than* light amplification by stimulated emission of radiation
UNESCO *rather than* United Nations Educational, Scientific, and Cultural Organization
VTOL *rather than* vertical takeoff and landing aircraft

6.6.3.2 Preference for Full Form

6.6.3.2.1 General Use

The full form of terms ***should*** be selected as terms when the abbreviated form is not widely used and generally understood. Cross references ***should*** be made from the abbreviated forms.

Example 45: Preference for full form of terms over acronyms

automated teller machines *rather than* ATMs
driving while intoxicated *rather than* DWI
prisoners of war *rather than* POWs

6.6.3.2.2 Ambiguity

Many acronyms and abbreviations stand for more than one word or phrase; the full form of the term ***should*** therefore be selected as the term in preference to the abbreviated form, even when the abbreviation has only one value in the domain of the controlled vocabulary.

Example 46: Full form of terms to eliminate ambiguity

artificial insemination *rather than* AI
artificial intelligence *rather than* AI

6.6.4 Neologisms, Slang, and Jargon

Neologisms, slang, and jargon terms often cover new concepts originating within a particular specialty, subculture, or social group. Such terms are generally not included in standard dictionaries. When no widely accepted alternative exists, the neologism, slang, or jargon term ***should*** be accepted as a term.

Example 47: Use of slang terms

kludge
yuppies

6.6.4.1 Provisional Terms

A neologism, slang, or jargon term ***may*** be labeled provisional (see section 11.1.6), and ***may*** be elevated to full term status as the term becomes accepted into the language.

Example 48: Use of provisional slang terms

burnout

6.6.4.2 Cross-references from Slang Terms

A slang term that is an alternative to an existing and well-established term ***should*** be admitted as an entry term to the term.

Example 49: Cross reference from slang terms

psychiatrists *rather than* shrinks
shrinks USE **psychiatrists**

6.6.5 Popular and Scientific Names

If a popular and a scientific name refer to the same concept, the form most likely to be sought by the users of the controlled vocabulary ***should*** be chosen as the term, and a cross-reference provided from the non-preferred term. For example, **penguins *should*** be chosen as the term in a nontechnical controlled vocabulary with a cross reference from the scientific equivalent, **Sphenisciformes.** However, **Sphenisciformes *should*** be selected as the preferred term in a zoological controlled vocabulary with a cross-reference from the popular name, **penguins**.

6.6.6 Loanwords, Translations of Loanwords, and Foreign-Language Equivalents

6.6.6.1 Loanwords

Loanwords are terms borrowed from other languages that have become naturalized in the borrowing language. If such terms are well established, they ***should*** be admitted into the controlled vocabulary. Diacritics ***should*** be included if required by the orthographic authorities for the borrowing language.

Example 50: Loanwords from other languages

coup d'état
gestalt
habeas corpus
pas de deux
weltanschauung

6.6.6.2 Translations of Loanwords

Occasionally a loanword and a commonly accepted translation coexist in the language of the controlled vocabulary. The loanword ***should*** be preferred if it is more widely accepted in the domain of the controlled vocabulary.

Example 51: Translation of loanwords

amicus curiae *rather than* friend of the court [in a legal controlled vocabulary]

If the translation becomes well established, however, it ***should*** be selected as the term. In all cases where a concept can be expressed by both a loanword and a translated equivalent, a cross-reference ***should*** be made from the non-preferred term.

Example 52: Use of loanword translation as preferred term

braking radiation *rather than* bremsstrahlung
pen name *rather than* nom de plume

6.6.6.3 Foreign-Language Equivalents

Foreign-language terms, i.e., terms which have not become naturalized in the language of the controlled vocabulary, ***should*** be linked to terms in the preferred language in cases where the foreign

terms are likely to be sought by users. The language chosen for the term ***should*** be that which the user would likely expect, with a cross reference from the equivalent term in the other language. (For guidelines on multilingual controlled vocabularies, see ISO 5964 [30].)

Example 53: Foreign language equivalents

Shavuot / Pentecost [Hebrew / English]

6.6.7 Proper Names

Variant forms of proper names of persons, institutions, organizations, and places, as well as titles, if uncontrolled, create problems for searchers. Frequency of need for proper name access points generally determines which of the following practices is adopted. Proper names of persons, institutions, organizations, places, and titles ***may*** be:

a) controlled by inclusion in a controlled vocabulary of subject (topical) terms;

b) controlled through a separate name authority file; or

c) left uncontrolled.

In options b) or c), proper names ***may*** be assigned to content objects as identifiers, differentiated from topical terms. Different methods ***may*** be adopted for controlling various types of names. For example, geographic names could be included in a controlled vocabulary of topical terms, and personal names established in an authority file.

Reasons for merging the files of name headings and topical terms are (a) that the borderline between them is not sharp, and (b) that it is often desirable to link a class with its instances hierarchically, e.g., **museums** and **Louvre** (see section 8.3.2).

When proper names are included in a controlled vocabulary, the form of the name ***should*** be selected in accordance with a recognized code of cataloging practice, such as the *Anglo-American Cataloguing Rules [20]*.

6.6.7.1 Place Names

Names of countries and geographic regions frequently vary from language to language. Variant terms referring to the same place also occur within a single language community as illustrated in the two following examples:

Example 54: Places with both an official and a popular name in common use

Netherlands / Holland

Example 55: Places with an Anglicized and vernacular name that coexist

Cambodia / Kampuchea
Israel / Yisrael

The form most familiar to the users of the controlled vocabulary ***should*** be designated as the entry term, and cross-references ***should*** be provided from the variants. Other things being equal, preference ***should*** be given to the official rather than the popular name. The short form of the official name ***should*** be preferred. Standard authorities, such as the *U.S. Geographic Names Information System [28]*, ***should*** be consulted for the official forms.

Example 56: Cross references from popular place names

Arabian Peninsula *rather than* Arabia
Netherlands *rather than* Holland
Philippines *rather than* Republic of the Philippines

Trade names are proper nouns and the rules in this section apply to their use.

6.7 Capitalization and Non-alphabetic Characters

6.7.1 Capitalization

It is recommended that predominantly lowercase characters be used for terms in controlled vocabularies. (The display of terms in indexes is outside the scope of this Standard. Note that many existing controlled vocabularies deviate from this recommendation and use initial capitals in the display of terms.)

Capitals ***should*** be used only for the initial letter(s) of proper names, trade names, and for those components of taxonomic names, such as genus, which are conventionally capitalized. Capitals ***should*** be used for all the letters of initialisms or where featured in unusual positions in product or corporate names. Because lowercase letters can occur in unusual positions in proper names, using a combination of capitals and lowercase letters in controlled vocabularies indicates to the user the correct orthography of a term in natural language and serves to distinguish common nouns from similar proper names.

Example 57: Capitalization of proper and trade names

dBASE IV
DNA
information systems
Information Systems Corp.
NewsBank

6.7.2 Non-alphabetic Characters

To simplify filing and searching, the use of symbols and punctuation marks in terms and entry terms ***should*** be minimized. Established orthographic authorities ***should*** be used to determine when such characters are essential.

6.7.2.1 Parentheses

To eliminate problems in filing and searching, parentheses ***should*** be avoided in controlled vocabularies whenever possible. Parentheses ***should*** be used only to enclose qualifiers (see section 6.2.1) and trademark indicators, or when they constitute part of a term.

Example 58: Use of parentheses with qualifiers

Paris (France)
Amazon (TM)

6.7.2.2 Hyphens

To eliminate problems in filing and searching, hyphens generally ***should not*** be used in controlled vocabularies. Hyphens ***should*** be retained in topical terms when dropping them would lead to ambiguity. Hyphens ***should*** be retained where they occur as part of abbreviations, trademarks, chemical names, or proper nouns.

Example 59: Removal of hyphens

high temperature testing
nonfiction

Example 60: Retention of hyphens

MS-DOS
n-body problem
Newton-John, Olivia
p-benzoquinone
Saur, Karl-Otto
Stratford-upon-Avon
un-ionized particles

*NOTE: The two hyphens retained for **n-body** and **un-ionized** are examples of topical terms where dropping them would lead to ambiguity. The remaining terms retain the hyphen as part of their official name.*

6.7.2.3 Apostrophes

To eliminate problems in filing and searching, apostrophes generally ***should not*** be used in controlled vocabularies. Two exceptions are described in the following subsections.

6.7.2.3.1 Possessive Case

Apostrophes indicating the possessive case, whether in the singular or plural, ***should*** be retained. (For the choice of singular or plural terms, see section 6.5.) The singular form is most frequently found in eponyms. Standard orthographic authorities ***should*** be consulted for the position of the apostrophe in common nouns.

Example 61: Singular and plural use of possessive apostrophes

Singular	Plural
Boyle's law	**artists' colonies**
Pascal's triangle	**Presidents' Day**
	women's rights

*NOTE: For medical eponyms, the use of the possessive form ('s) is becoming progressively less common, e.g. **Down syndrome** instead of Down's syndrome..*

6.7.2.3.2 Proper Names

Apostrophes that are part of proper names ***must*** be retained.

Example 62: Apostrophes in proper names

O'Dwyer's tubes

6.7.2.4 Diacritical Marks

Diacritical marks ***may*** be used if they are required for proper names or by the accepted standards of a discipline. (See section 6.6.6.1 for the use of diacritics in loanwords and section 6.7.3 for the use of diacritics in romanization.)

Example 63: Use of diacritical marks

El Niño
Guillain-Barré syndrome
Mössbauer effect
résumés

6.7.2.5 Other Symbols and Punctuation Marks

Symbols, such as the ampersand, and punctuation marks ***should not*** be used except in trademarks and proper names because they create filing and searching problems.

Example 64: Use of symbols in proper names

AT&T
PL/C

6.7.3 Romanization

Commonly accepted spellings for terms or proper names from languages written in non-roman scripts, as found in authoritative reference sources, ***should*** be used in preference to systematic romanization, i.e., applying a table to convert the characters in a non-roman script into roman ones. A cross-reference ***should*** be provided from the systematic romanization to the established spelling.

Example 65: Romanization of terms

Chebyshev polynomials *rather than* Tchebyshev polynomials
or Čebyšev polynomials

Koran *rather than* Qur´an

When romanizing (transliterating or transcribing), the procedure described in a relevant standard or a nationally accepted library scheme ***should*** be followed as far as possible. (Library schemes used in the U.S. have been collected in ALA-LC Romanization Tables [25].) If a choice exists, a romanization system that uses few or no diacritical marks ***should*** be selected.

7 Compound Terms

7.1 General

Compound or multiword terms in natural language are considered lexemes, i.e., words bound together as lexical units. **New Jersey** is a good example. Dictionaries differ in their policies regarding the inclusion of various categories of compound terms, and so they are not always authorities in this regard for controlled vocabulary designers. For this reason, guidelines on the handling of compound terms in controlled vocabularies have been developed.

To be acceptable as a term, a compound term ***should*** express a single concept or unit of thought, capable of being arranged in a genus-species relationship within a hierarchy or tree structure. **Ships and boats** is an example of a compound term that is acceptable because both **ships** and **boats** are members of a broader concept, **vessels.**

7.2 Purpose of Guidelines on Compound Terms

As a general principle of vocabulary control, terms ***should*** represent single concepts, expressed by a single word or by a multiword term unless it is a precoordinated term. See section 7.2.1 for more information on this topic.

The establishment of procedures for dealing consistently with compound terms is one of the most difficult areas in the field of controlled vocabulary construction and indexing. The guidelines below are intended to:

- aid in achieving intra-indexer, inter-indexer, and inter-organizational consistency;
- avoid over-complexity of the indexing language;
- achieve a controlled vocabulary whose structure is based upon principles that are amenable to logical explanation; and
- enhance the ease and precision of searching.

7.2.1 Precoordinated Terms

A precoordinated term combines two or more concepts to denote a specific concept and its attributes more precisely.

Rules ***may*** be developed for combining concepts to form precoordinate indexing terms. While the use of precoordinated terms can increase the size of a controlled vocabulary, the inclusion of compound terms can simplify the indexing, browse, and navigation processes. The rules for developing precoordinated indexing terms are generally outside of the scope of this Standard. These guidelines are available in sources such as the Library of Congress *Subject Cataloging Manual: Subject Headings* [32].

Precoordinated terms are used to:

- provide a method of locating highly specific information by browse and navigation rather than by searching the component terms separately;
- group closely related terms together (alphabetically) that might be spread throughout the controlled vocabulary otherwise; and
- describe complex concepts that are very frequently used in the domain of reference.

Example 66: Precoordinated terms

aid to families with dependent children
avian hypersensitivity pneumonitis
carrier sense multiple access
New England—Genealogy—Handbooks, Manuals, etc.
United States—History—Civil War, 1861-1865

7.2.2 Retrieval Considerations

7.2.2.1 Postcoordinate Retrieval

Terms can be combined at the time of search and retrieval using Boolean or other operators to express complex concepts. This is called postcoordinate retrieval. Postcoordinate retrieval is extremely useful for searching very large and complex databases. Use of postcoordinate retrieval techniques helps to keep the controlled vocabulary to a manageable size because terms can be combined as needed to represent the desired specific concepts at the time of search and retrieval.

Example 67: Combining terms during retrieval to express complex concepts

liver AND **anatomy**
New York AND **subway**

7.2.2.2 Precoordinate Retrieval

Individual terms representing concepts ***may*** be precoordinated into semantically linked, heading subheading combinations. These terms can be used during the indexing and searching processes. They are also very useful in browsing and navigation, especially by users who are not totally familiar with a controlled vocabulary and its structure.

Example 68: Complex precoordinated terms

Furniture—California—San Francisco—History—20th Century
Liver—Blood Vessels—Diseases—Congresses

7.3 Factors to be Considered When Establishing Compound Terms

The factors listed below ***should*** be used to decide which multiword terms ***should*** be split into separate terms and which ***should*** be retained in compound form.

These procedures need not be applied rigidly. Administrative policies or system constraints ***may*** govern which of these practices are adopted by a given organization. However, section 10 shows how different practices can affect interoperability.

a) *Literary warrant*. A compound term might be employed so frequently within the literature of the domain covered by the controlled vocabulary that splitting the term into its components would be unacceptable to users who consider it a lexeme.

b) *Regulating the number of terms in the controlled vocabulary*. Compound terms increase the number of terms in a controlled vocabulary.

c) *Print versus information retrieval systems*. If a controlled vocabulary is used for printed and online resources as well as website navigation, the requirements of users of all three formats ***should*** be taken into account.

   - Printed indexes ***may*** require precoordinated headings for compound concepts, so that the number of entries per term can readily be browsed by a searcher.
   - Computerized retrieval systems often benefit from indexing of compound concepts using separate terms. These can then be combined as needed by the searcher (postcoordination).
   - Web navigation schemes often benefit from the use of precoordinated terms to guide the user to the specific content required by showing a clear path through the hierarchy.

d) *Avoiding false hits in retrieval.* Compound terms often eliminate certain kinds of false hits (content retrieved by unwanted combinations of terms) in retrieval when the same words in a different sequence have a different meaning. For example, if the concept **library science** is represented by the separate terms **library** and **science**, a search on this combination will also retrieve **science library**.

   In addition, the use of compound terms can solve the famous **copper coating on lead pipes** problem. If these two concepts are indexed using the single terms:

   copper
   lead
   coating
   pipes

   there is no way to prevent retrieval by searches on **copper pipes** AND **lead coating**.

e) The nature of terminology in a given field. This **may** call for special criteria to regulate the treatment of compound terms.

7.4 Elements of Compound Terms

Elements of a compound term are distinguished by their logical roles or functions. These are relevant to the criteria for determining when a compound term ***should*** be created (see section 7.5).

The majority of compound terms, including both pre- and postmodified noun phrases, have one of two possible components:

a) The *focus* (also known as the head noun): the noun component that identifies the broader class of things or events to which the term as a whole refers.

   Example 69: Focus of a compound term

   concrete *in the compound term* **reinforced concrete**
   glass *in the compound term* **stained glass**

b) The *modifier* (also known as the difference) is the part of a compound term that refers to a characteristic, or logical difference, which narrows the denotation of the focus by specifying a subclass of the broader concept represented by the focus.

Example 70: Modifier of a compound term

reinforced [which specifies a subclass, i.e., a type, of concrete in reinforced concrete]
stained [which specifies a subclass of glass in stained glass]

7.5 Criteria for Establishing Compound Terms

Compound or multiword terms ***should*** be established as terms in the following circumstances:

a) Splitting the parts would lead to ambiguity or a loss of meaning.

Example 71: Compound terms to avoid ambiguity if split

data dictionaries
plant food
pocket knives

b) One component of the term is not relevant to the scope of the controlled vocabulary or is too vague to exist as an independent term.

Example 72: Compound terms with nondistinctive elements

NOTE: nondistinctive element underlined

composite drawings
first aid
stone cutters

c) The modifier in the term has lost its original meaning, so the meaning of the compound term as a whole is not the sum of the meaning of its parts.

Example 73: Compound terms with distinct meanings

deck chairs
trade winds

d) The modifier suggests a resemblance, as in a metaphor, to an unrelated thing or event.

Example 74: Compound terms with metaphorical modifiers

butterfly valves
tree structures

e) The term contains an adjective that does not define a subclass of the focus, and is not actually a member of that class at all.

Example 75: Compound terms with adjectival modifiers

artificial limbs
paper flowers
rubber ducks
tin soldiers

f) The term is a proper name, or includes proper nouns or proper adjectives.

Example 76: Compound proper names

Freudian slips
Hudson's Bay Company
New Jersey
United Nations

g) The compound term has become so familiar in common use, or in the field covered by the controlled vocabulary, that it is considered for practical purposes to represent a single concept.

Example 77: Compound terms in common use

data processing
gross domestic product

7.6 Criteria for Determining When Compound Terms Should be Split

7.6.1 Factors to be Considered

It is more difficult to specify exact criteria for recognizing the kinds of compound terms that ***should*** be split into separate components, each of which is then entered as a noun or noun phrase in the controlled vocabulary.

The following recommendations are based upon general criteria, but they ***should not*** be regarded as mandatory instructions to be applied rigidly in all circumstances. As noted above, such decisions are often based on the nature of terminology in a given field, which ***may*** call for special treatment of compound terms, or on the necessity for controlling the size of the vocabulary.

7.6.2 Hierarchical Structure

The following recommendations are offered as a means of achieving consistent treatment of compound terms that ***should*** be split. They also ensure a correct hierarchical structure in which each term fits into a logical conceptual framework. Each of these recommendations is accompanied by a converse condition that can be used to identify classes of terms that usually cannot be split without risk of ambiguity.

7.6.2.1 Focus and Modifier

A compound term ***should*** be split when its focus refers to a property or part, and its modifier represents the whole or possessor of that property or part.

Example 78: Splitting compound terms when focus represents the "part"

aircraft + **engines** *rather than* aircraft engines
hospitals + **personnel** *rather than* hospital personnel
soils + **acidity** *rather than* soil acidity

Conversely, a compound term ***should not*** be split when the focus term refers to a whole and the modifier is a term for its part or property.

Example 79: Retaining compound terms when focus represents the "whole"

acid soils
skilled personnel

7.6.2.2 Transitive Action

A compound term ***should*** be split if it consists of a term representing a transitive action modified by a term for the object on which the action is performed.

Example 80: Splitting compound terms when focus represents a transitive action

offices [object] + **management** [action] *rather than* office management
textiles [object] + **printing** [action] *rather than* textile printing

Conversely, a compound term ***should not*** be split when its component term for a thing or material is modified by a term for the action carried out upon it.

Example 81: Retaining compound terms when focus represents the object of a transitive action

cast steel *rather than* cast [action] + steel [object]
printed textiles *rather than* printed [action] + textiles [object]

7.6.2.3 Intransitive Action

A compound term ***should*** be split if it consists of a term for an intransitive action modified by a term for the performer (agent) of the action.

Example 82: Splitting compound terms when focus represents an intransitive action

birds [agent] + **migration** [action] *rather than* bird migration
fire [agent] + **flame retardants** [counteragent] *rather than* fire flame retardants

Conversely, a compound term ***should not*** be split when the term for an object is modified by a term for the intransitive action in which the object is involved.

Example 83: Retaining compound terms when focus represents the object of an intransitive action

dancing shoes (dancing [action] + shoes [object])

7.7 Node Labels

A compound term ***should not*** be created solely for the reason that it forms a logical level in a hierarchy and would serve to group a set of narrower terms. For this purpose node labels ***should*** be used instead (see section 8.3.5).

Phrases ***may*** be used as node labels if the term forms a logical level in the hierarchy.

Example 84: Compound phrases as node labels in a hierarchy

cookware
[pots and pans]
frying pans
sauce pans
baking dishes

7.8 Order of Words in Compound Terms

Noun phrases ***should*** be entered in a controlled vocabulary in natural language order, i.e., direct order. For further guidance on the form of premodified and postmodified noun phrases, see section 6.4.1.2.

Example 85: Natural language order of compound terms

birds of prey
oral surgery
secondary schools

7.8.1 Cross-references from Inversions

The inverted form of a noun phrase ***may*** be included as a cross-reference to the preferred term in natural language order.

Example 86: Cross references from noun phrase inversions

surgery, oral USE **oral surgery**
schools, secondary USE **secondary schools**

See also the discussion of permuted displays in section 9.2.1.1.

8 Relationships

There are three types of relationships used in controlled vocabularies:

a) Equivalency (see section 8.2)

b) Hierarchy (see section 8.3)

c) Association (see section 8.4)

8.1 Semantic Linking

The relationships among terms in a controlled vocabulary are indicated by semantic linking. Semantic linking encompasses various techniques and conventions for indicating the relationships among terms.

Formats for displaying semantic relationships are described in section 9; this section is concerned with defining and illustrating the relationships themselves. Table 1 illustrates the basic types of relationships and provides some simple examples. Each of these types is covered in detail in sections 8.2 through 8.4.

Table 1: Selected semantic relationships among terms

Relationship Type	Example
Equivalency	
Synonymy	UN / United Nations
Lexical variants	pediatrics / paediatrics
Near synonymy	sea water / salt water smoothness / roughness
Hierarchy	
Generic or IsA	birds / parrots
Instance or IsA	sea / Mediterranean Sea
Whole / Part	brain / brain stem
Associative	
Cause / Effect	accident / injury
Process / Agent	velocity measurement / speedometer
Process / Counter-agent	fire / flame retardant
Action / Product	writing / publication
Action / Property	communication / communication skills
Action / Target	teaching / student
Concept or Object / Property	steel alloy / corrosion resistance
Concept or Object/ Origins	water / well
Concept or Object / Measurement Unit or Mechanism	chronometer / minute
Raw material / Product	grapes / wine
Discipline or Field / Object or Practitioner	neonatology / infant

8.1.1 Indicating Relationships Among Terms

A basic property of relationships in controlled vocabularies is that they are reciprocal; that is, each relationship indicated between **Term A** and **Term B *must*** have a corresponding relationship from **Term B** to **Term A**. This rule ***must*** be observed for all types of relationships.

The conventional abbreviations for relationship indicators are used in the examples below. Additional abbreviations for specialized purposes are found in the following sections. (A complete list of abbreviations used in this Standard appears in section 4.2.)

The relationship indicators are always paired operators. Some indicators are symmetric while others are asymmetric as illustrated below:

- Related Term (RT) is symmetric:

  If **Term A** RT **Term B**, then **Term B** RT **Term A**

- Preferred Term (Equivalency) – USE and UF are asymmetric:

  If **Term A** USE **Term B**, then **Term B** UF **Term A**

- Hierarchical Relationships – BT and NT are asymmetric:

  If **Term A** BT **Term B** then **Term B** NT **Term A**.

8.2 Equivalence Relationships

When the same concept can be expressed by two or more terms, one of these is selected as the preferred term.

The relationship between preferred and non-preferred terms is an equivalence relationship in which each term is regarded as referring to the same concept. The preferred term in effect substitutes for other terms expressing equivalent or nearly equivalent concepts. A cross-reference to the preferred term ***should*** be made from any "equivalent" entry term.

The equivalence relationship is expressed by the following conventions:

- U or USE, which leads from a non-preferred (entry) term to the preferred term, and
- UF or USED FOR, the reciprocal relationship, which leads from the preferred entry term to the non-preferred term(s).

Example 87: USE and UF relationships

Aves USE **birds**
birds UF Aves

outline USE **shape**
shape UF outline

These relationship indicators are the equivalents of *see* and *x* (see from), respectively, in many traditional subject heading lists.

The equivalence relationship covers five basic types:

a) synonyms (see section 8.2.1)
b) lexical variants (see section 8.2.2)
c) near-synonyms (see section 8.2.3)
d) generic posting (see section 8.2.4)
e) cross reference to elements of compound terms (see section 8.2.5)

8.2.1 Synonyms

Synonyms are terms whose meanings are regarded as the same or nearly the same in a wide range of contexts. True synonyms are rare in natural language. Although the terms are interchangeable in many circumstances, usage can vary as a result of such factors as level of formality, professional vs. lay context, or pejorative vs. neutral vs. complimentary connotation. The examples listed below are illustrative of the various classes of synonyms that may be encountered in practice. A slash is used to separate the synonyms, indicating that the preferred term has not been selected.

Example 88: Synonyms of different linguistic origin

cats / felines
freedom / liberty
sodium / natrium
sweat / perspiration

Example 89: Popular and scientific name synonyms

aspirin / acetylsalicylic acid
gulls / Laridae
salt / sodium chloride

Example 90: Generic and trade name synonyms

petroleum jelly / Vaseline
photocopies / Xeroxes
refrigerators / Frigidaires
tissues / Kleenex

Example 91: Variant names for emergent concepts

hovercraft / air cushion vehicles

Example 92: Current or favored terms replacing outdated or deprecated terms

poliomyelitis / infantile paralysis
developing countries / underdeveloped countries

Example 93: Slang or jargon synonyms

helicopters / whirlybirds
psychiatrists / shrinks

Example 94: Dialectical variants

elevators / lifts
subways / undergrounds

In these and other cases, preferred terms ***should*** be selected to serve the needs of the majority of users, bearing in mind the criteria enumerated in sections 6 and 7. For the sake of predictability, these criteria ***should*** be applied consistently throughout the controlled vocabulary. If, for example, American spelling is preferred to British spelling, or scientific names are preferred to popular names, this decision ***should*** be explained in the introduction to the controlled vocabulary and ***should*** be applied consistently in the formulation of terms.

8.2.2 Lexical Variants

Lexical variants differ from synonyms in that synonyms are different terms for the same concept, while lexical variants are different word forms for the same expression. These forms ***may*** derive from spelling or grammatical variation or from abbreviated formats.

Example 95: Lexical variants (direct versus inverted order, orthographic variants, stem variants, and irregular plurals)

radar antennas / antennas, radar
Romania / Rumania / Roumania
ground water / ground-water / groundwater
online / on-line
pediatrics / paediatrics
mice / mouse

Example 96: Full name and abbreviation variants

International Federation for Documentation / FID
pi mesons / pions
polyvinyl chloride / PVC

8.2.3 Near-Synonyms

Near-synonyms are terms whose meanings are generally regarded as different, but which are treated as equivalents for the purposes of a controlled vocabulary. The extent to which terms are treated as near-synonyms depends in large measure upon the domain covered by the controlled vocabulary and its size. Near-synonyms may include antonyms or represent points on a continuum.

Example 97: Near-synonyms

sea water / salt water	[variant terms]
meteors / meteorites / meteoroids	[points on a continuum]
smoothness / roughness	[antonyms]

For each of these sets of near synonyms, a vocabulary developer might decide to designate one of the terms as the preferred term with the understanding that it will retrieve all content described by the other terms as well.

NOTE: Antonyms can also be treated as related terms, rather than equivalent terms. See section 8.4.

As a general rule, terms ***should*** be treated as near-synonyms only in subject areas that are peripheral to the domain of the controlled vocabulary. When concepts can be distinguished in the controlled vocabulary domain with sufficient precision to justify their representation as separate terms, they ***should*** be individually defined and retained. If two concepts cannot be consistently and reliably differentiated from each other, however, a term for one concept ***should*** be selected as the preferred term and a USE reference made from the other.

8.2.4 Generic Posting

Generic posting is a technique in which the name of a class and the names of its members are treated as equivalents, with the broader class name functioning as the preferred term.

Example 98: Generic posting for class members

waxes	UF plant waxes
plant waxes	USE **waxes**

furniture
UF beds
UF chairs
UF desks
UF tables

beds USE **furniture**
chairs USE **furniture**
desks USE **furniture**
tables USE **furniture**

If employed, this technique ***should*** be limited to the peripheral areas of a subject field, or used when the number of documents on the members of a class does not warrant its being split into subclasses. This practice places limits on the specificity of the controlled vocabulary, and ***should*** be used with discretion. The narrower terms are useful entry terms.

8.2.5 Cross-references to Elements of Compound Terms

A USE....AND reference ***may*** be made from a compound term to its components in cases where postcoordination of the individual term components is used to represent the desired concept. Care should be taken to ensure that it is clear to indexers and searchers that multiple terms must be used.

Example 99: Cross references to compound term elements

coal mining USE **coal** AND **mining**
ferromagnetic films USE **ferromagnetic materials** AND **films**

USED FOR . . . AND . . . is a reference that ***may*** be used for the reciprocal of the reference from the compound term, as UF alone can suggest that one of the two components could be used alone in searching for the concept.

Example 100: Cross references from compound term elements

coal USED FOR coal AND mining
mining USED FOR coal AND mining

8.3 Hierarchical Relationships

The use of hierarchical relationships is the primary feature that distinguishes a taxonomy or thesaurus from other, simple forms of controlled vocabularies such as lists and synonym rings.

Hierarchical relationships are based on degrees or levels of superordination and subordination, where the superordinate term represents a class or a whole, and subordinate terms refer to its members or parts. Reciprocity ***should*** be expressed by the following relationship indicators:

- BT (Broader Term), a label for the superordinate (parent) term
- NT (Narrower Term), a label for the subordinate (child) term

Example 101: Hierarchical relationship notation (BT and NT)

mammals
BT **vertebrates**

vertebrates
NT **mammals**

In the flat format controlled vocabulary (see section 9.3.1.2) BT and NT indicate one level broader and one level narrower, respectively. There are other types of displays that indicate multiple levels of hierarchy (see section 9).

Hierarchical relationships ***may*** also be indicated by systematic presentations such as tree structures or graphic displays (see section 9.3.4).

Hierarchical relationships cover three logically different and mutually exclusive situations:

a) the generic relationship (see section 8.3.1);

b) the instance relationship (see section 8.3.2); and

c) the whole-part relationship (see section 8.3.3).

Each of these types of relationships ***may*** apply to a single term. Special codes ***may*** be used to distinguish them.

Each of the relationships ***should*** lead to hierarchies that are amenable to a logical test (see Figure 6 and Figure 7 below) through reference to the basic types of concept represented by the terms, for example, those listed in section 6.3.

Every subordinate term ***should*** refer to the same basic kind of concept as its superordinate term; that is, both the broader and the narrower term ***should*** represent a thing, an action, a property, etc. For example:

- **anatomy** (a discipline) and **central nervous system** (a body part that can be an object of study of that discipline) represent different kinds of concepts; therefore, these terms cannot be related hierarchically;
- **central nervous system** and **brain** both represent body parts; these terms can therefore be related hierarchically.

8.3.1 Generic Relationships

This relationship identifies the link between a class and its members or species. This type of relationship is often called “IsA.” A simple way to apply the test for validity described above is to formulate the statement “[narrower term] *is a* [broader term].” This relationship is also amenable to a logical "all-and-some" test as shown in Figure 6:

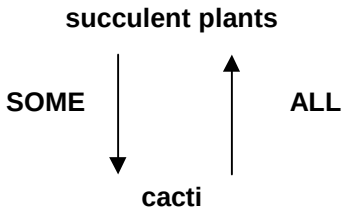
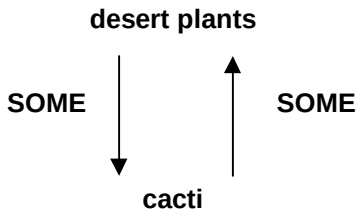


Figure 6: All and some relationship – passes test

Figure 6 illustrates that some members of the class **succulent plants** are known as **cacti** and that all **cacti**, by definition and regardless of context, are **succulent plants**. Figure 7 illustrates a case where the "all-and-some" test is not met. While some members of the class **desert plants** are known as **cacti**, some, but not all, cacti are desert plants. These terms ***should*** therefore be assigned to different hierarchies in the controlled vocabulary, and both terms ***should*** be assigned to the same content object when indexing a work on "cacti as desert plants.."

Figure 7: All and some relationship – does not pass test



8.3.1.1 Codes for the Generic Relationship

The generic nature of a relationship ***may*** be identified by the BT/NT coding described above, or if more refined coding is desired for the various types of hierarchical relationship, by the following abbreviations:

- BTG = Broader term (generic)
- NTG = Narrower term (generic)

Example 102: Generic relationship notation (BTG and NTG)

rats
BTG **rodents**

rodents
NTG **rats**

8.3.2 Instance Relationships

This relationship identifies the link between a general category of things or events, expressed by a common noun, and an individual instance of that category, often a proper name. This type of relationship is also known as an "IsA" relationship.

Example 103: Instance relationships

mountain regions
Alps
Himalayas

In the above example, the **Alps** and the **Himalayas** are assigned to subordinate positions in a hierarchy, yet they are neither kinds nor parts of **mountain regions**, but represent specific examples or instances.

8.3.2.1 Codes for the Instance Relationship

The hierarchical instance relationship ***may*** be indicated by the following abbreviations:

- BTI = Broader term (instance)
- NTI = Narrower term (instance)

Example 104: Instance relationship notation (BTI and NTI)

fairy tales
NTI **Cinderella**
NTI **Rumpelstiltskin**

8.3.3 Whole-Part Relationships

This relationship covers situations in which one concept is inherently included in another, regardless of context, so that the terms can be organized into logical hierarchies, with the whole treated as a broader term. This relationship can be applied to several types of term; the three types enumerated below are not intended to be exhaustive. In the following examples, parts are indicated through indentation.

Example 105: Whole-part relationships – systems and organs of the body

nervous system
  central nervous system
    brain
    spinal cord

Example 106: Whole-part relationships – geographic locations

Canada
  Ontario
    Ottawa
    Toronto

Example 107: Whole-part relationships – hierarchical organizational, corporate, social, or political structures

armies
  military divisions
    battalions
      regiments

8.3.3.1 Codes for the Whole-Part Relationship

The hierarchical whole-part relationship ***may*** be indicated by the following abbreviations:

- BTP = Broader term (partitive)
- NTP = Narrower term (partitive)

Example 108: Whole-part relationship notation (BTP and NTP)

central nervous system
  BTP **nervous system**

nervous system
  NTP **central nervous system**

8.3.3.2 Parts of Multiple Wholes

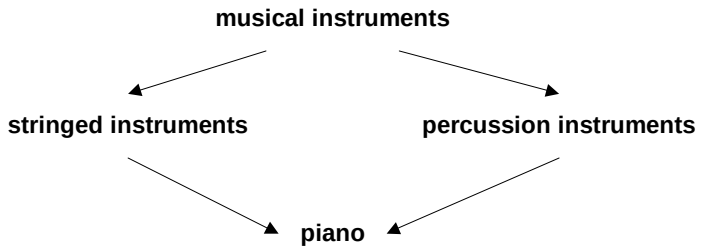
When a whole-part relationship is not exclusive to a pair of terms, i.e., the part can belong to multiple wholes, the name of the whole and its part(s) ***should not*** have a hierarchical relationship. Rather, they ***should*** be linked associatively rather than hierarchically in the controlled vocabulary. Carburetors, for example, are parts of machines other than cars. Therefore, the appropriate relationship in this instance is **cars** RT **carburetors**.

8.3.4 Polyhierarchical Relationships

Some concepts belong, on logical grounds, to more than one category. They are then said to possess polyhierarchical relationships.

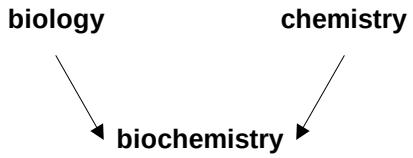
In the following example, the term **pianos** is assigned to subordinate positions on the basis of its generic relationship to two broader terms—in other words, **pianos** would be an NT to both **stringed instruments** and **percussion instruments**.

Example 109: Polyhierarchy based on generic relationship



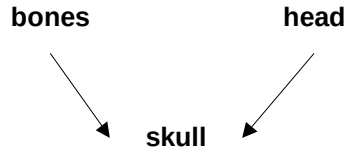
In other cases, polyhierarchical links ***may*** be based upon whole-part relationships, as shown in the following example.

Example 110: Polyhierarchy based on whole-part relationship



In some cases, polyhierarchical links can be based on logically different relationships, as shown in the following example.

Example 111: Polyhierarchy based on multiple types of relationships



In the example above, the link between **bones** and **skull** is based upon the generic relationship (the skull is a kind of bone), whereas the link between **head** and **skull** is based on the hierarchical whole-part relationship (the skull is part of the head). These relationships ***may*** be represented using the BTG and BTP notation.

Example 112: Multiple type polyhierarchical relationship notation (BTG and BTP)

skull
BTG **bones**
BTP **head**

8.3.5 Node Labels in Hierarchies

When terms are arranged in hierarchies (such as tree structures) in a taxonomy, thesaurus, or Web navigation layout, node labels ***may*** be used to show the principles of division among a set of sibling terms (terms that share a broader term).

Although their function is similar to that of broader terms, node labels are not terms, and ***must not*** be used as indexing terms. They are often typographically distinguished from terms, e.g., through the use of italics and/or enclosure in angle brackets.

Example 113: Node labels in hierarchies

cars
by motive power
diesel cars
electric cars
by purpose
racing cars
sports cars

8.4 Associative Relationships

This relationship covers associations between terms that are neither equivalent nor hierarchical, yet the terms are semantically or conceptually associated to such an extent that the link between them ***should*** be made explicit in the controlled vocabulary, on the grounds that it ***may*** suggest additional terms for use in indexing or retrieval. The most common associative relationship used in thesauri is symmetrical and is generally indicated by the abbreviation RT (related term), but more semantically refined associations may also be developed to capture both symmetric and asymmetric associations. (See also section 8.4.4.)

Example 114: Associative relationship notation (RT)

cells
RT **cytology**

cytology
RT **cells**

The associative relationship is the most difficult one to define, yet it is important to make explicit the nature of the relationship between terms linked in this way and to avoid subjective judgments as much as possible; otherwise, RT references could be established inconsistently.

As a general guideline, whenever one term is used, the other ***should*** always be implied within the common frames of reference shared by the users of the controlled vocabulary. Moreover, one of the terms is often a necessary component in any explanation or definition of the other; the term **cells**, for example, forms a necessary part of the definition of **cytology**.

Either of the following types of terms can be linked by the associative relationship:

a) those belonging to the same hierarchy (see section 8.4.1)

b) those belonging to different hierarchies (see section 8.4.2)

8.4.1 Relationships Between Terms Belonging to the Same Hierarchy

Relationships are needed for terms belonging to the same hierarchy in various special situations, primarily to guide the user in locating the desired term.

8.4.1.1 Relationships Between Overlapping Sibling Terms

Related Term (RT) references are required for sibling terms with overlapping meanings, such as **ships** and **boats**, where each of the terms can be precisely defined (so they do not form an equivalence set), yet they are sometimes used loosely and almost interchangeably. The user interested in one ***should*** be reminded of the other.

In the systematic section of a controlled vocabulary containing organized hierarchies (see section 9.3.4), the method of display will bring the Related Term references together. The relationship ***should*** be indicated explicitly, however, in an alphabetical listing and in the alphabetic section of a hierarchical controlled vocabulary.

Example 115: Overlapping sibling terms hierarchical display

boats
BT **vehicles**
RT **ships**

ships
BT **vehicles**
RT **boats**

Example 116: Overlapping sibling terms alphabetical display

boats
SEE ALSO **ships**

ships
SEE ALSO **boats**

8.4.1.2 Relationships Between Mutually Exclusive Sibling Terms

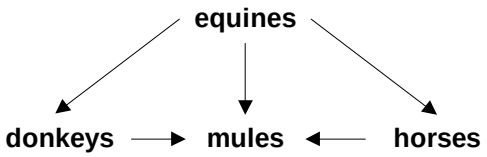
It is not necessary to interrelate all sibling terms. For example, there is no need to associate terms such as **roses** and **daffodils**, which share the broader term **flowers**, because the meaning of the terms does not overlap, i.e., they are mutually exclusive.

8.4.1.3 Derivational Relationships

Concepts linked by a familial or derivational relationship (i.e., one of the concepts is derived from the other) also require Related Term (RT) references. This guideline applies to the relationship between such terms as **parents** and **children** or to the relationship of **mules** to **donkeys** and **horses**.

In the example diagram that follows, **donkeys**, **horses**, and **mules** are all subclasses of **equines**. In the alphabetic example, Related Terms are provided between **mules** and **donkeys** as well as **mules** and **horses**. **Horses** and **donkeys** are not linked by Related Terms because they do not share a derivational relationship.

Example 117: Derivational relationship graphic display



Example 118: Derivational relationship hierarchical display

donkeys
BT **equines**
RT **mules**
equines
NT **donkeys**
NT **horses**
NT **mules**
horses
BT **equines**
RT **mules**
mules
BT **equines**
RT **donkeys**
RT **horses**

8.4.2 Relationships Between Terms Belonging to Different Hierarchies

It is possible to establish many grounds for associating terms belonging to different hierarchies. Related Term references are often made between etymologically related terms, i.e., terms that contain the same root, but which do not represent the same kind of thing. The following are some representative examples of typical relational situations. For guidance on coding the precise nature of a relationship, see section 8.4.4. The following examples illustrate the types of associative relationships listed in Table 1 in section 8.1.

Example 119: Process / Agent associative relationships

temperature control
RT **thermostats**
thermostats
RT **temperature control**

hunting
RT **hunters**
hunters
RT **hunting**

Example 120: Process / Counteragent associative relationships

fire
RT **flame retardants**
flame retardants
RT **fire**

inflammation
RT **anti-inflammatory agents**
anti-inflammatory agents
RT **inflammation**

Example 121: Action / Property associative relationships

environmental cleanup
RT **pollution**
pollution
RT **environmental cleanup**

polling
RT **public opinion**
public opinion
RT **polling**

Example 122: Action / Product associative relationships

weaving
RT **cloth**
cloth
RT **weaving**

lacrimation
RT **tears**
tears
RT **lacrimation**

Example 123:Action / Target associative relationships

harvesting
RT crops
crops
RT **harvesting**

binding
RT **books**
books
RT **binding**

Example 124: Cause / Effect associative relationships

death
RT **bereavement**
bereavement
RT **death**

pathogens
RT **infections**
infections
RT **pathogens**

Example 125: Concept or Object / Property associative relationships

poisons
RT **toxicity**
toxicity
RT **poisons**

liquids
RT **surface tension**
surface tension
RT **liquids**

Example 126: Concept or Object / Origins associative relationships

beluga caviar
RT **Caspian Sea**
Caspian Sea
RT **beluga caviar**

Socratic method
RT **Greek civilization**
Greek civilization
RT **Socratic method**

Example 127: Concept or Object / Units or Mechanisms of Measurement associative relationships

electric current
RT **amperes**
amperes
RT **electric current**

temperature
RT **thermometers**
thermometers
RT **temperature**

Example 128: Raw Material / Product associative relationship

wheat
RT **flour**
flour
RT **wheat**

graphite
RT **lubricant**
lubricant
RT **graphite**

Example 129: Discipline or Field / Object or Practitioner associative relationships

mathematics
RT **mathematicians**
mathematicians
RT **mathematics**

neurology
RT **nervous system**
nervous system
RT **neurology**

botany
RT **plants**
plants
RT **botany**

Additionally, an associative relationship ***may*** be used for phrases where the noun is not a true broader term. For example, a **rubber duck** is not a **duck**.

Example 130: Noun is not a true broader term associative relationships

ducks
RT **rubber ducks**
rubber ducks
RT **ducks**

fishes
RT **fossil fishes**
fossil fishes
RT **fishes**

Antonyms ***may*** also be treated as associative relationships. (For an alternative approach where antonyms are treated as equivalents, see section 8.2.3.)

Example 131: Antonyms as associative relationships

short
RT **tall**
height
RT **depth**

8.4.3 Node Labels for Related Terms

In order to bring closely related concepts together in the alphabetical array under a given term, related terms ***may*** be divided into categories that do not form part of a logical hierarchy. These related terms ***should*** then be identified by a node label.

Example 132: Node labels for related terms

books
RT
[operations]
binding
printing

In this example, [operations] functions as a node label that describes a category or facet to which the related terms belong. Node labels used with narrower terms generally describe a characteristic or

sub-division. For more information on the use of node labels to indicate principles of division for logical hierarchies, see section 8.3.5.

8.4.4 Specifying Types of Related Term References

In certain controlled vocabularies, it ***may*** be considered desirable to refine Related Term references in order to make the nature of the relationships explicit.

Codes for such relationship indicators and their reciprocals ***may*** be developed locally. These local codes ***should*** be clearly explained and illustrated in the introduction or documentation of the published or machine-readable controlled vocabulary.

9 Displaying Controlled Vocabularies

The way in which a controlled vocabulary is presented affects the user's willingness and ability to make use of it. This section defines requirements and recommendations for effective display of controlled vocabularies.

Note that in some applications, portions of a controlled vocabulary ***may*** be displayed in connection with the content objects they reference. However, this display is governed by the software used to generate the search results and is not covered by this Standard.

9.1 General Considerations

9.1.1 Elements to Address

The vocabulary developer ***should*** address the following elements of the vocabulary display:

- **Presentation**
  Presentation decisions include how to represent the term relationships, whether and how to make typography distinctions, capitalization, and filing rules. These are discussed in section 9.2.
- **Type of Display**
  A large variety of display types can be used, from simple alphabetical listings to complex graphical displays. Often multiple structures are presented to the user. The most common structures and their advantages and disadvantages are discussed in section 9.3. Appendix C is a summary table of all the display types including the advantages and disadvantages of each.
- **Format**
  Controlled vocabularies ***may*** be delivered in print or electronic formats or both. Web and hyperlinking technology allows additional display capabilities not available in print form. Section 9.4 discusses format-specific considerations.
- **Documentation**
  All controlled vocabularies ***should*** provide user documentation that describes how to use the vocabulary. Section 9.5 discusses what to include in the documentation.

9.1.2 User Categories

The design of displays for controlled vocabularies ***should*** take into account the needs of each anticipated class of user:

a) **Controlled vocabulary maintenance personnel** – These users have expertise in indexing and controlled vocabulary construction and are likely to be experts in the subject domain of the controlled vocabulary.

They must have access to all views of a controlled vocabulary and complete information

about each term, with the ability to edit and manipulate term records, cross-references, classification notation, and hierarchies. They require "housekeeping" displays not needed by end users of a controlled vocabulary. (See section 11.3 for more information on maintenance procedures for controlled vocabularies.)

b) **Indexers and expert searchers** – These users have expertise in indexing, online information retrieval, use of controlled vocabularies, or all of these.

Indexers are likely to have expertise in the subject domain of the controlled vocabulary while expert searchers may or may not have such expertise. These sophisticated users require the ability to search and view cross-references, definitions, and notes for terms as well as various levels of the classification or hierarchies.

Postings data are especially important for searchers. Sophisticated controlled vocabulary displays and terminology are appropriate for these users.

c) **End users** – These users are not likely to be experienced in the jargon and complexities of online information retrieval or the conventions of controlled vocabulary notation. They may have expertise in the subject field and understand its terminology.

The types of displays available to expert searchers can be useful to end users as well, when designed with their needs in mind. End users can benefit from on-screen instructions in addition to any printed documentation that exist.

Controlled vocabulary developers ***may*** want to produce different versions of the vocabulary:

- a basic list of terms, references, and relationships designed for the end user or occasional searcher, and
- a more complete version designed for the indexer and the expert searcher, which may include detailed scope notes, indexing instructions, information on term history, and postings data.

9.2 Presentation

9.2.1 Displaying the Equivalence Relationship

USE references from non-preferred terms ***should*** be incorporated into the main listing of a controlled vocabulary rather than being relegated to an auxiliary "access vocabulary" or separate list of entry terms.

- USE (U) equivalence relationships ***should*** be created to lead users from entry vocabulary terms to the preferred term.
- USED FOR (UF) equivalence relationships ***should*** be created to show users what synonyms and near-synonyms are covered by the preferred term.

Example 133: Equivalence relationship display

ACRYLONITRILE POLYMERS
UF ACRILAN
BT **POLYMERS**
NT **ACRYLIC RESINS**

ACRYLONITRILE RUBBER
USE **NITRILE RUBBER**

ACTH
USE **ADRENOCORTICOTROPIC HORMONE**

ACTINIDE SERIES

Elements having atomic numbers 89 through 103. For the series of nuclides resulting from the decay of U-235, see RADIOACTIVE SERIES.

BT METALS
NT ACTINIUM
AMERICUM
BERKELIUM

Source: DTIC Thesaurus [2]

9.2.1.1 USE References and Permuted Displays

Some controlled vocabularies include an auxiliary permuted or rotated list that gives access to every word in terms and USE references (see examples in section 9.3.2). The permuted display ***should not*** be used as a substitute for the inclusion of useful inversions as USE references in the main alphabetic list.

9.2.1.2 USE References for Inverted Forms

USE references ***should*** be created from the inverted form of a term to the preferred term in direct order to provide assistance to the user in locating the appropriate term, especially when the controlled vocabulary appears as a printed product. Most controlled vocabularies that are available online provide search access to the vocabulary by keyword. Consequently, very few include inverted forms in the main alphabetical listing as they are unnecessary.

Example 134: USE reference display

abnormalities, cardiovascular
USE **cardiovascular abnormalities**

A USE reference from the inverted form of the term **information science** is, in contrast, not essential and would, moreover, be misleading as it might be interpreted as **science information**.

An automatically generated permuted display of the controlled vocabulary will display the term **information science** under **science** as well as under **information**.

9.2.1.3 USE References in Hierarchical Relationships

The type of hierarchical relationship display chosen (see section 9.3.4) affects decisions on the necessity of USE references.

If the flat format of display is being used, i.e. only one level of narrower terms (see section 9.3.1.2), and any of the narrower terms have further levels of narrower terms, these would not be displayed under the top level broader term. In this case, USE references ***should*** be made from the inverted form of these compound terms.

Example 135: USE references for multiple levels of narrower terms

Hierarchy for broadest term:
libraries
NT **academic libraries**
public libraries
special libraries

Separate hierarchy listing for one of the narrower terms:
academic libraries
NT **college libraries**
university libraries

Each of these narrower terms would need USE references from the inverted form of the term to the direct form.

libraries, college
USE **college libraries**

libraries, university
USE **university libraries**

Moreover, in cases where a term has numerous broader term, narrower term, and related term relationships, USE references from the inverted form of the term can make it easier to locate the preferred term and all its relationships.

If hierarchical information is displayed only by a tree structure for the controlled vocabulary (see section 9.3.4.2), it is helpful to include USE references from the inverted form of certain terms to make it easier to locate the preferred term.

If all narrower terms are included in the alphabetic display and the inverted USE references would be alphabetically adjacent to the hierarchy's broader term, the inverted USE references for the narrower terms ***may not*** be needed.

Example 136: Unnecessary USE references

Hierarchy:
libraries
NT **academic libraries**
public libraries
special libraries

USE references for the following terms ***may not*** be needed:
libraries, academic
libraries, public
libraries, special

9.2.2 Displaying Hierarchical and Associative Relationships

9.2.2.1 Term Detail Presentation

When hierarchical and associative relationships have been established for a vocabulary, each term display ***should*** include some or all of the following information about the term, preferably in this order:

a) Scope notes
b) USE references
c) USED FOR (UF) references
d) Broader Term (BT) references to terms that are one level broader
e) Narrower Term (NT) references to terms that are one level narrower
f) Related Term (RT) references to terms that have an associative relationship with the term

9.2.2.2 Narrower Terms

If there are multiple Narrower Terms, they are generally arranged alphabetically. They ***may*** also be arranged logically if another order would be more intuitive for the user: Indentation or other cues to the hierarchy are essential if more than one level of hierarchy is presented.

Example 137: Alphabetical arrangement of Narrower Terms

flowers
NT **daffodils**
iris
pansies
peonies
tulips

Example 138: Logical arrangement of Narrower Terms

planets
NT **Mercury**
Venus
Earth
Mars
Jupiter
Saturn
Uranus
Neptune
Pluto

9.2.2.3 Node Labels

Node labels ***may*** also be used to group both narrower and related terms in categories. Node labels in hierarchies, which are not to be used in indexing or searching, ***should*** be distinguished from terms, e.g., by placing them in square brackets.

Example 139: Node label display

flowers
NT
[flowers by plant type]
annuals
marigolds
petunias
zinnias

perennials
dianthus
coreopsis
peonies
[flowers by flowering season]
spring flowers
crocuses
daffodils
autumn flowers
chrysanthemums
physalis

9.2.3 Indentation

Indentation ***should*** be used as an additional visual cue for hierarchical and associative relationship layouts. For examples of indentation in multilevel hierarchical displays, see Section 9.3.4.1.

Example 140: Use of indentation with relationship display

controlled vocabularies
BT **indexing languages**
NT **subject heading lists**
thesauri

preferred terms
UF descriptors
BTP **thesauri**

entry terms
SN the lead-in terms of a controlled vocabulary
UF use references
BTP **thesauri**

thesauri
BT **controlled vocabularies**
NT
[by display format]
alphabetical thesauri
faceted thesauri
hierarchical thesauri
[by number of languages]
monolingual thesauri
multilingual thesauri
NTP **descriptors**
entry terms
relationship indicators
RT **authority files**
computer-aided indexing
dictionaries
glossaries
nomenclature

9.2.4 Typography

Terms, non-preferred terms, relationship indicators, and textual notes ***should*** be typographically distinguished. Such distinctions are more important for certain types of displays, such as hierarchical, or for certain formats, particularly print. Additional typographic distinctions ***may*** be needed for electronic and Web formats to indicate hyperlinks. (See section 9.4.3.4.)

Suggested typographic specifications are:

- lightface or italics for non-preferred terms and notes,
- all capitals for relationship indicators such as USE, and
- boldface for terms.

Example 141: Use of typographics in term display

teenagers USE **adolescents**

9.2.5 Capitals and Lowercase Letters

The controlled vocabulary ***should*** serve as an orthographic authority in addition to noting preferred terminology. A combination of capitals and lowercase letters ***should*** therefore be used in controlled vocabulary terms as dictated by usage. See section 6.7.1 for specific guidelines on capitalization.

9.2.6 Filing and Sorting

This section provides general rules for the filing and sorting of alphabetic characters and numerals in controlled vocabularies. Resources such as the *ALA Filing Rules [21]*, *Library of Congress Filing Rules [31]*, and *British Standard Alphabetical Arrangement and the Filing Order of Numerals and Symbols* (BS 1749) [26] contain rules and guidelines for handling specific filing problems. These filing codes are, however, not compatible with each other in certain details, and only one of them ***should*** be chosen as an authority.

In electronic systems, sorting rules and handling of non-alphabetic characters ***may*** be pre-established by the software. Some systems provide options for developers to select the sorting rules. Where such options exist, the recommendations in this Standard ***should*** be selected.

9.2.6.1 Alphabetic Characters

Terms consisting of letters ***should*** be filed word-by-word rather than letter-by-letter.

In word-by-word filing, a space is significant. This filing principle is called “nothing before something.” It keeps together terms that begin with the same word.

Example 142: Comparison of word-by-word and letter-by-letter filing

Word-by-word filing	**Letter-by-letter filing**
gas USE **gases**	gas USE **gases**
gas coolants	**gas coolants**
gas welding	**gases**
gases	**gaskets**
gaskets	**gas welding**

Another advantage of word-by-word filing is that the space is a significant character in computer programs that are used to sort terms and to print the controlled vocabulary.

A disadvantage of word-by-word filing is that it separates compound words from those consisting of two words, e.g., **folk song**, **folk tune**, **folklore**, **folktale** [shown in sorted order]**.** Creating USE references from the variant spellings ***may*** compensate for this problem.

9.2.6.2 Numerals

Computer systems differ in the way numbers are filed. Some systems file numerals at the beginning of the alphabet and some at the end. Either position is acceptable, provided that the filing position of numerals is explained in the introduction to the controlled vocabulary.

Applying the principles of ascending order and word-by-word filing (see section 9.2.6.1) to numerals will cause **19** to file before **2** and **2004** before **30**. A special sort routine (now fairly commonplace) is required to ensure that computer-generated sort order is based on arithmetical value, e.g., **1**, **1.5**, **10**, **151**, **1000**.

Special sorting instructions are usually inserted by humans if numbers are to be ignored, e.g., in chemical terms, or filed as spelled out, e.g., **21 [twenty-one] gun salute**.

Roman numerals also usually require special sorting instructions in order to ensure that they will file properly, e.g. **XXIII psalm** before **LII psalm**.

9.2.6.3 Non-Alphanumeric Characters

Punctuation marks ***should*** generally be ignored in filing entry terms and terms.

9.2.6.3.1 Commas

Commas in inverted USE references ***should*** be ignored when filing or sorting.

Example 143: Sorting with commas ignored

cat breeds
cat, Siamese USE **Siamese cats**
catalogs

9.2.6.3.2 Parentheses

- Parentheses within terms ***should*** be ignored in filing.
  For example, **NAD(P)H-Flavin Oxidoreductase** would be treated as if it were spelled **NADPH-Flavin Oxidoreductase**.
- Parentheses around qualifiers ***should*** be treated as a special character in filing.
  It is desirable for the user to see homographic terms juxtaposed to assist in the selection of the desired term. Ignoring the parentheses around qualifiers would lead to the dispersal of homographic terms among compound terms. The desired sequence is illustrated in the following example.

Example 144: Sorting of qualifiers with parentheses

Earth (planet)
earth (soil) USE **soil**
earth science

9.2.6.4 Relation of Character Set to Search Commands

Non-alphabetic characters used in machine-readable controlled vocabularies ***should not*** conflict with special characters used in search commands. For example, parentheses used in qualifiers ***should not*** be interpreted as nesting indicators in a search statement. If the potential for ambiguity exists, a different character ***should*** be substituted, e.g., square brackets ***may*** enclose qualifiers.

9.3 Types of Displays

Given the variety of controlled vocabulary display types available, it is not possible to recommend a single type. Multiple display types are frequently available to the user, especially when using electronic formats. The number of display types used in print formats is necessarily limited by space considerations.

The most commonly used display types are discussed below. Appendix C describes the characteristics and uses of each of the display types for assistance in deciding which one(s) to use.

9.3.1 Alphabetical Displays

9.3.1.1 Alphabetical Listing

An alphabetical listing is the most basic type of vocabulary display. It ***may*** contain both preferred terms and entry terms with their respective USE and USED FOR references. In the machine-readable controlled vocabulary, spelling variations, common misspellings, and common typographical errors ***may*** be stored as USE references to enhance retrieval, but need not be displayed to users. Controlled vocabularies linked to databases sometimes display the number of postings for assigned terms. The postings data in a controlled vocabulary display ***should not*** include the occurrences in the full text that match terms from the controlled vocabulary.

Example 145: Alphabetical listing display

Early Admission
Early Adolescence
U **Early Adolescents**
Early Adolescents
UF Early Adolescence; Young Adolescents
Early Childhood (1966 1980)
U **Young Children**
Early Childhood Education
Early Detection
U **Identification**
Early Diagnosis
U **Identification**
Early Experience
UF Preschool Experience

Source: Thesaurus of ERIC Descriptors [14]

9.3.1.2 Flat Format Displays

The flat format is the most commonly used controlled vocabulary display format. It consists of all the terms arranged in alphabetical order, including their term details, and one level of BT/NT hierarchy.

In some computer systems, each term in a hierarchical display is assigned a line number, which the user can reference in expanding a search. Node labels ***may*** also be assigned line numbers so they can be used to facilitate including narrower terms in a search strategy.

NOTE: A flat format display can also be considered a type of hierarchical display. See section 9.3.4 for additional hierarchical displays.

Example 146: Flat format display

Whale watching
RT Whales

Whalers (Persons)
RT Whales

Whales
SN: Aquatic mammals of the order Cetacea.
UF Cetaceans (NPT)
BT Marine mammals
NT Baleen whales
NT Fossil whales
NT Toothed whales
RT Whale oil
RT Whale watching
RT Whalers (Persons)
RT Whaling
SPEQ Ballenas

Whaling
RT Whales

Source: Thomson Gale Master Thesaurus, Synaptica® software by Synapse Corporation [15]

9.3.2 Permuted Displays

A permuted or rotated display of the terms lists each term multiple times in the alphabetic sequence of the controlled vocabulary for each of the words in the term. For example, a term such as **very high frequency radiation** would appear four times—each entry appearing in its proper place in the alphabetic sequence. The two types of permuted displays are Key Word In Context (or KWIC) and Key Word Out of Context (or KWOC), as shown in the following examples.

Example 147: KWIC Index permuted display

	frequency
very high	**frequency radiation**
	frequent flier miles
	high altitude
very	**high frequency**
very	**high frequency radiation**
	high hat
	radiation
very high frequency	**radiation**
	radiation therapy
high	**velocity**
	very high frequency radiation
	very large scale integration
pressure	**vessels**

Example 148: KWOC Index permuted display

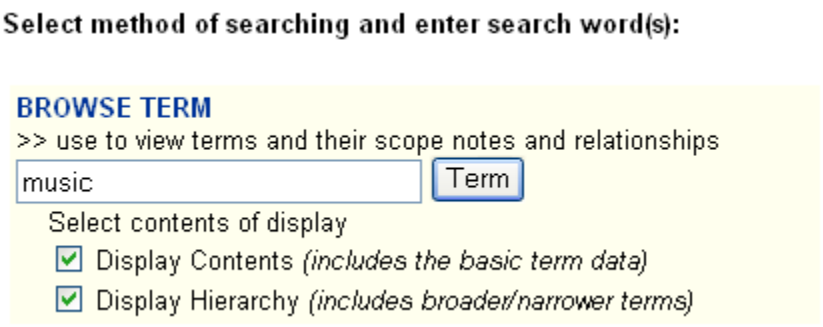
frequency	frequency
	very high frequency
	very high frequency radiation
frequent	frequent flier miles
high	high altitude
	high hat
	high velocity
	very high frequency
	very high frequency radiation
radiation	radiation
	radiation therapy
	very high frequency radiation
very	very high frequency
	very high frequency radiation
	very large scale integration
vessel	pressure vessel

9.3.3 Term Detail Displays

The complete details for each term ***should*** be presented in at least one of the display types offered. In print formats, this is typically incorporated into the hierarchical display. In electronic formats, users ***should*** be able to select a term from any display type and see an expanded view of the detail for that term, either fully or in part. Web implementations of controlled vocabularies ***should*** include a hyperlink from the term, wherever it appears, to the full term detail display.

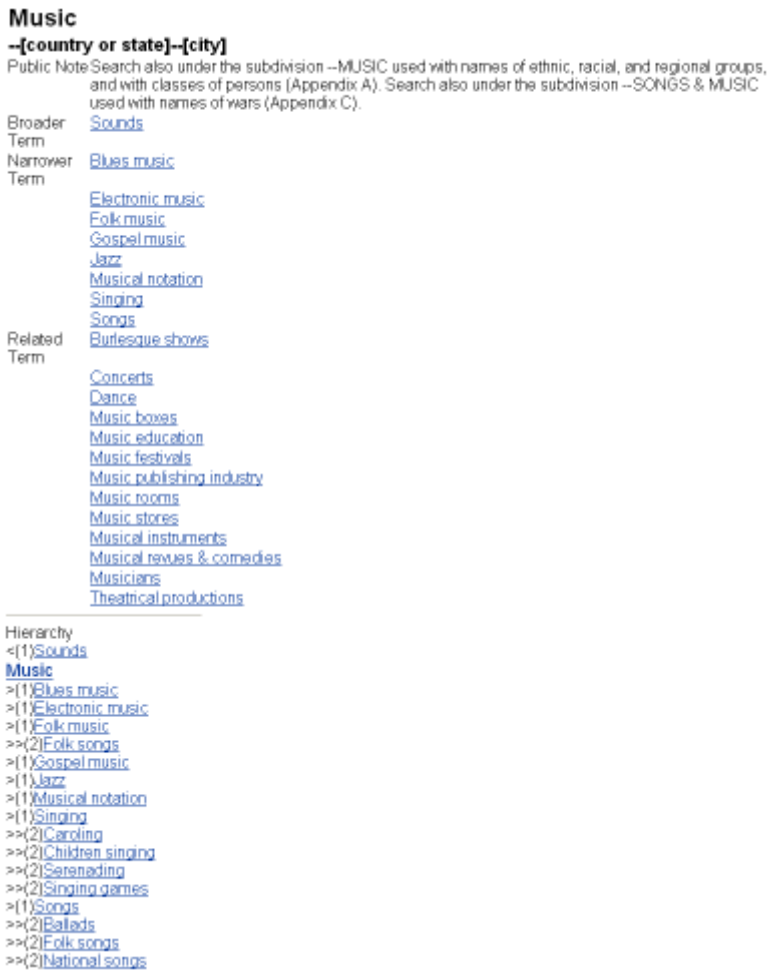
Electronic access to term details ***may*** offer users the option of calling up a term's history, scope note, or definition, separately from its relationships. Example 149 illustrates a thesaurus search screen that provides users such an option. Example 150 displays the full term detail results of the Example 149 search query. Example 151 depicts a term detail display from a different thesaurus.

Example 149: Search option for a term's contents and/or its hierarchy



Source: Thesaurus for Graphic Materials (TGM I) [13]

Example 150: Results of search query displaying both term contents and hierarchy



Source: Thesaurus for Graphic Materials (TGM I) [13]

Example 151: Term detail display

MeSH Heading	Kidney
Tree Number	A05.810.453
Annotation	kidney tissue or cells in cultures: Manual 18.6.15, 18.7.2, 26.23+; fetal or embryonic kidney tissue or cells in culture: do not use /embryol with KIDNEY; /blood supply: consider also RENAL CIRCULATION; /radionuclide: consider RADIOISOTOPE RENOGRAPHY, a specific kidney function test (see note there); /surg: consider NEPHRECTOMY; cyst = KIDNEY, CYSTIC but KIDNEY, POLYCYSTIC is also available; inflammation = NEPHRITIS; Goldblatt kidney = HYPERTENSION, GOLDBLATT see HYPERTENSION, RENOVASCULAR; nephrostomy: consider NEPHROSTOMY, PERCUTANEOUS
Scope Note	Body organ that filters blood for the secretion of URINE and that regulates ion concentrations.
Entry Term	Kidneys
See Also	Diuresis
See Also	Nephrectomy
See Also	Renal Circulation
Consider Also	consider also terms at GLOMERUL-, NEPHR-, PYEL-, and RENAL
Allowable Qualifiers	AB AH BS CH CY DE EM EN GD IM IN IR ME MI PA PH PP PS RA RE RI SE SU UL US VI
History Note	/transplantation was KIDNEY TRANSPLANTATION 1963-65; was KIDNEY/transplantation 1966-89
Entry Combination	transplantation:Kidney Transplantation
Unique ID	D007668

MeSH Tree Structures

Urogenital System [A05]
- Urinary Tract [A05.810]
  - Bladder [A05.810.161]
  - ▶ Kidney [A05.810.453]
    - Kidney Cortex [A05.810.453.324]
    - Kidney Medulla [A05.810.453.466]
    - Kidney Pelvis [A05.810.453.537] +
    - Nephrons [A05.810.453.736] +
  - Ureter [A05.810.776]
  - Urethra [A05.810.876]

Source: MeSH [7]

9.3.4 Hierarchical Displays

Taxonomies, thesauri, and any vocabularies with established relationships between terms ***should*** include a hierarchical display that illustrates the relationships. There are a number of variations in how hierarchical displays can be structured.

a) Multilevel Hierarchies

b) Tree Structure

c) Top Term Structure

d) Two-Way Hierarchical Structure

e) Broad Categories

f) Faceted

Sections 9.3.4.1 through 9.3.4.6 describe the most commonly used forms for displaying hierarchical relationships among terms. See also section 9.3.1.2 on Flat Format Displays.

9.3.4.1 Multilevel Hierarchical Displays

In a multilevel hierarchical display format, all levels of the broader and narrower terms related to a given term are immediately visible. This is in contrast with the flat format described above, in which only one level of broader or narrower terms is displayed and the user is required to navigate from term to term one level at a time to discover the full hierarchy.

In both multilevel formats described in the following sections, the sibling terms (the related terms at the same level) of the given term are only displayed by viewing the broader term.

Because the entire hierarchy is repeated over and over again, these display formats are far more space-consuming than the flat format and hence less desirable for a printed controlled vocabulary. However, if the hierarchies are not very deep, e.g., only three or four levels, the extra space required could be justifiable.

9.3.4.1.1 Multilevel Broader and Narrower Terms Hierarchical Displays

This type of multilevel display employs special notation, such as BT1, BT2 (one level broader, two levels broader) and NT1, NT2 (one level narrower, two levels narrower), to show the full hierarchy for each term. In some cases, a symbol notation is used to designate the levels.

Example 152: Multilevel hierarchy display using numerical level notation

Organisms
NT1 Eukaryotes
NT2 Animals
NT3 Aquatic animals
NT4 Aquatic mammals
NT5 Marine mammals
NT6 Whales
NT7 Baleen whales
NT8 Blue whale
NT8 Gray whale
NT8 Humpback whale
NT8 Minke whale
NT8 Right whales
NT9 Bowhead whale
NT9 Northern right whale
NT9 Southern right whale
NT7 Fossil whales
NT7 Toothed whales
NT8 Beaked whales
NT8 Beluga (Whale)
NT8 Dolphins (Mammals)
NT9 Bottlenosed dolphins
NT9 Killer whales
NT9 Pilot whales
NT9 River dolphins
NT8 Narwhal
NT8 Porpoises
NT8 Sperm whale

Source: Thomson Gale Master Thesaurus, Synaptica® software by Synapse Corporation [15]

Example 153: Multilevel hierarchy using symbol level notation

Processed products
• NT **Fermented products**
•• NT **Alcoholic beverages**
••• NT **Beers**
••• NT **Ciders**
••• NT **Liqueurs**
••• NT **Spirits**
••• NT **Wines**
•••• NT **Champagne and sparkling wines**
•••• NT **White wines**
•••• NT **Rose wines**

NOTE: This display uses • • • • to indicate the various levels.

Source: AGROVOC [1]

9.3.4.1.2 Generic Structure Hierarchical Displays

Multiple levels of hierarchy ***may*** be indicated without broader term / narrower term (BT/NT) notation by using the abbreviation GS (generic structure) with indentation and punctuation marks, such as periods, as cues to the levels of hierarchy.

Example 154: Multilevel hierarchy with generic structure notation

clinostats
(added July 2000)
DEF Devices for producing vector–averaged gravitational environments which mimic microgravity
UF *random positioning machines*
GS **simulators**
. environment simulators
. . space simulators
. . . clinostats
RT **bioreactors**
centrifuges
clinorotation
gravitational effects
gravitational physiology
microgravity
rotating environments
space environment simulation
tissue engineering
weightlessness simulation

Source: NASA Thesaurus [9]

9.3.4.2 Tree Structure

In a tree structure, each term is assigned a classification notation or line number. This leads the user from the alphabetic display to the full hierarchical display (sometimes called *systematic display* or *classified display*), which is placed in a separate sequence. In the flat format or multilevel structure, hierarchical relationships ***may*** be built from the bottom up; in a tree structure, they are often built from the top down. The power of a tree structure can be exploited more fully using electronic display technology. (See section 9.4.2 and 9.4.3).

Example 155: Tree structure using classification notation

Respiratory Tract Diseases	**[C08]**
Bronchial Diseases	**[C08.127]**
>>> Asthma	**[C08.127.108]**
Asthma, Exercise-Induced	**[C08.127.108.110]**
Status Asthmaticus	**[C08.127.108.880]**

Source: Medical Subject Headings (MeSH) [7]

If a hierarchical classification scheme is applied to a tree structure, its notation ***must*** be carefully developed so that it will be hospitable to interpolation at any level. Computer-generated or humanly-assigned line numbers can be revised when terms are added, but the notation is not expressive, i.e., it does not reflect the levels of hierarchy.

Pure notation, consisting of either letters or numbers, has a smaller base than mixed or alphanumeric notation and is therefore less hospitable to new terms. Mixed notation can, however, create problems for novice users, who can have difficulty locating and understanding the codes.

9.3.4.3 Top Term Structure Displays

In a top term structure, the alphabetic display includes all the relationships found in the flat format, with the addition of a relationship indicator for the top term (TT) of the hierarchy.

The top term leads the user to a separate sequence of the controlled vocabulary in which top terms are arranged alphabetically, with each top term followed by all its narrower terms, arranged hierarchically on various levels.

Example 156: Top Term Structure

Top Term Cluster: **INTELLECTUAL PROPERTY**

Intellectual property
- **Copyright**
- **Copyright infringement**
- **Copyright royalties**
- **Gray market imports**
- **Intangible property**
- **Intellectual property**
- **Inventions**
- **Inventors**
- **Patent agreements**
- **Patent law and legislation**
- **Patent licenses**
- **Patents**
- **Product counterfeiting**
- **Trademark agreements**
- **Trademarks**

Source: Legislative Indexing Vocabulary Thesaurus [5]

9.3.4.4 Two-Way Hierarchical Structure Displays

A two-way hierarchical display option is generally appended to a flat format controlled vocabulary. Each term is an access point, and all levels of broader and narrower terms are displayed, generally without notation and with indentation as a cue to hierarchy. This format is logically equivalent to the generic structure, i.e., one can view all of the broader and narrower terms of a term in a single display.

Example 157: Two way hierarchical structure

Poultry

Used For	**Fowls**	Flat format display
Broader Term	**Birds**	
Narrower Term	**Chickens**	
Related Term	**Ducks**	
	Geese	
	Poultry houses	
	Poultry industry	
	Turkeys	

Hierarchy
<<(2)**Animals**
<(1)**Birds**
Poultry
>(1)**Chickens**
>>(2)**Roosters**

Two-way hierarchy display

Source: Thesaurus for Graphic Materials (TGM I) [13]

9.3.4.5 Broad Categories Displays

The alphabetic arrays of some controlled vocabularies include numbers that identify the broad category to which each term belongs. The notation for the category ***may*** be searchable online or used for purposes of selective dissemination of information. Such controlled vocabularies generally feature a separate section that displays the terms for each numbered category or subcategory in an alphabetic sequence, undifferentiated in terms of hierarchical levels.

Example 158: Broad categories (numbered / classified)÷

01 AVIATION TECHNOLOGY
01 Aerodynamics
.
.
.
03 Aircraft
03 01 Helicopters
03 02 Bombers
03 04 Patrol and Reconnaissance Aircraft
03 05 Transport Aircraft
03 06 Training Aircraft
.
.
.
04 Flight Control and Instrumentation

Source: DTIC Thesaurus [2]

9.3.4.6 Faceted Display

Some controlled vocabularies provide a display of the terms organized according to the broad categories or facets to which the term belongs. Facets ***may*** have a hierarchical arrangement as well so that narrower facets are arranged within broader categories.

Example 159: Faceted Display

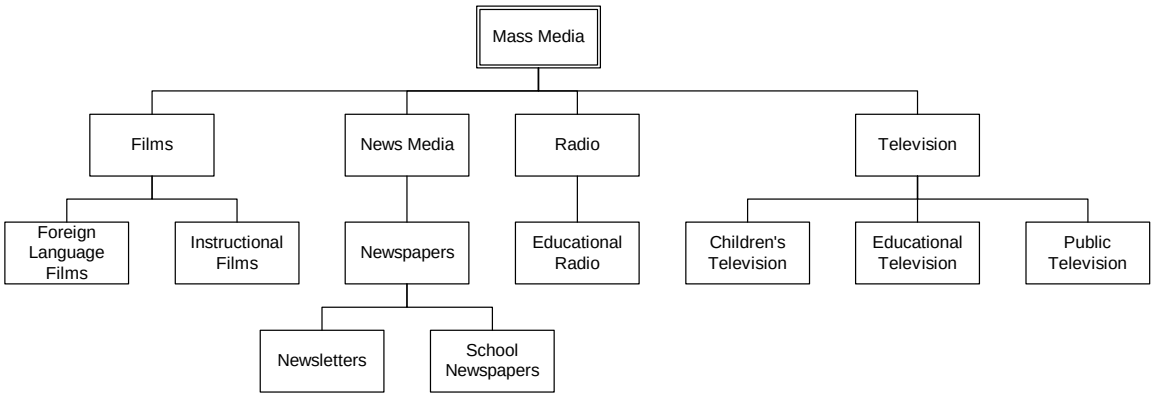
Source: Art & Architecture Thesaurus® [3]

9.3.5 Graphic Displays

Research has shown that graphic displays can communicate relationships among concepts more effectively than linear displays to some users. Graphic displays are logically equivalent to a tree structure, but usually do not have a notation. The decision to include a graphic display ***should*** take into account the domain and search habits of users. Graphic displays can be more effective in an interactive computer environment where terms are hyperlinked to the term details or a conventional flat format display.

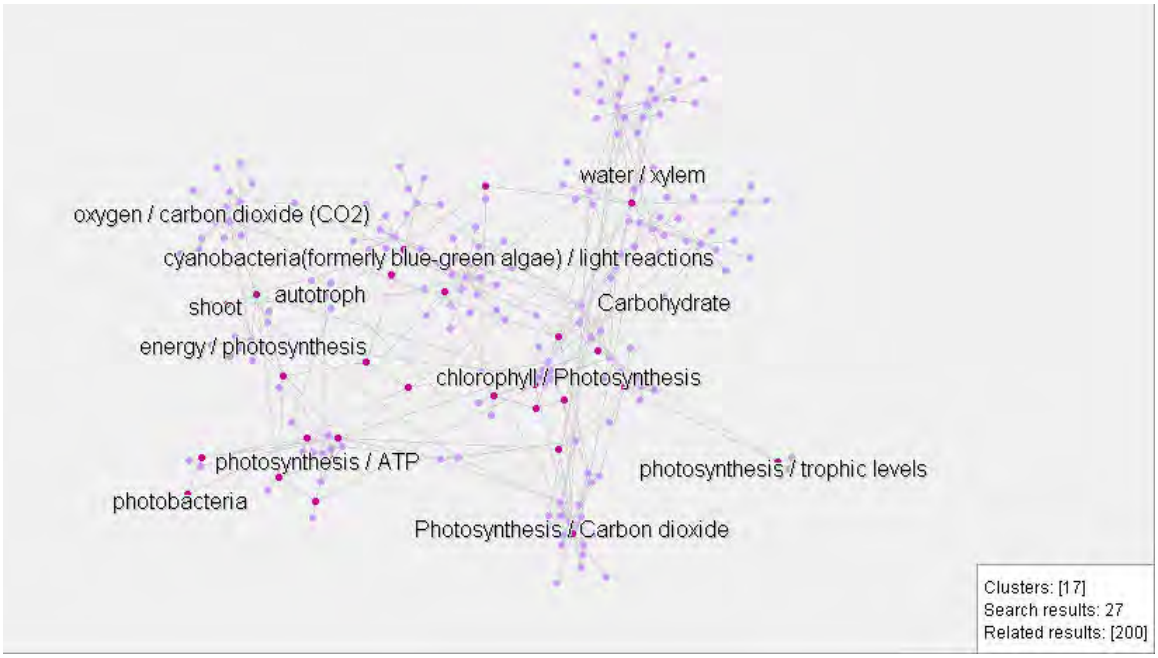
There are many commercially available products that provide graphic displays of terms and term relationships. Some products are able to generate concept maps using the terms in a controlled vocabulary or even words in natural language using contextual and semantic information to determine the mapping. (See the Bibliography for more information on graphic displays and concept maps.) Concept maps that appear in conjunction with printed vocabularies are static. In an electronic system, a graphic display is generated in real time based on an underlying controlled vocabulary and/or a set of full text resources or other content objects.

Example 160: Static concept map



NOTE: This graphic display was created with terms and hierarchies from the ERIC Thesaurus [14]. Descending lines represent narrower terms.

Example 161: Concept map clusters generated in real-time



Source: Photosynthesis Concept Map, Xrefer [11]

9.4 Display Formats – Physical Form

Many controlled vocabularies are now available in electronic format instead of or in addition to the traditional printed format. Electronic display options provide significant enhancements over the printed format but they also require additional design considerations. Electronic displays have been further enhanced with the development of Web and hyperlinking technologies. This section discusses the issues to be addressed in making format decisions and identifies special considerations for particular formats.

A controlled vocabulary is generally used in the context of an information storage and retrieval system. The choice of display formats will depend upon the domain, the searching behavior of users in that domain, and the overall design of the information system of which the controlled vocabulary is a component.

Controlled vocabularies ***may*** be produced in multiple formats including:

a) A traditional print format

b) An electronic file, usually PDF, of the print document.

c) An online, interactive version that is searchable and browsable. This can be part of an associated information retrieval system or a Web browser application.

d) A stand-alone software application that can be installed on the end user's personal computer.

9.4.1 Print Format – Special Considerations

9.4.1.1 Minimizing Double Lookups

A major principle in the design of a printed controlled vocabulary is the minimization of double lookups, i.e., the need to consult more than one sequence of terms to find the desired entries. The flat format of hierarchical display, a very frequently used style, requires one or more lookups to move through an entire hierarchical structure because only one level of narrower terms is included for each

term in the display. In order to minimize multiple lookups and to save time for the vocabulary user, many printed controlled vocabularies now include the complete hierarchical structure with each term. The chief disadvantage of this approach is, of course, that it requires much more space.

9.4.1.2 Juxtaposition of Terms

Juxtaposition of terms plays a role in the creation of USE references for a printed controlled vocabulary (see section 9.2.1). For example, an entry term that would immediately precede or follow the term to which it leads ***may*** be suppressed. Some singular and plural forms of terms can be widely separated in a printed alphabetic list, however, and a USE reference from one to the other would then be warranted, even if the policy regarding preference for singular or plural is stated in the documentation of the controlled vocabulary (see section 9.5).

Example 162: USE references for alphabetically separated terms

NOTE: In this example, several terms occur between the singular and plural forms of cat so a reference from one to the other is desirable, especially because the singular form is the first word of a compound term.

cat USE **cats**
cat fleas
catalogs
catharsis
cats

Including terms that occur between the position of a potential entry term and its corresponding term, or the interfiling of terms with uncontrolled terms in an index, ***may*** create a need for additional USE references.

Similar considerations apply to establishing related term references in a printed controlled vocabulary. Most printed controlled vocabularies are produced from machine-readable files that also generate a screen display. Because a screen display holds less information than a printed page, a more generous entry vocabulary is required for screen displays than for printed versions of a controlled vocabulary. The reader's eye takes in the larger panorama of the printed page, thus making it easier to locate terms. A possible approach is to include all potentially useful references in the machine-readable file and suppress unnecessary ones in the printed product.

9.4.1.3 Running Heads

Whenever a printed controlled vocabulary includes multiple sequences, e.g., an alphabetic list, a rotated list, and a hierarchical display, each page ***should*** feature a running head to identify the sequence. Running heads are useful even if the sequences are in separate volumes.

9.4.2 Screen Format – Special Considerations

This section discusses special considerations for controlled vocabularies displayed on a computer screen, regardless of the computer application. Section 9.4.3 discusses additional considerations for applications designed for display in Web browsers or utilizing Web-related capabilities such as hyperlinking.

9.4.2.1 User Interface Design

Viewing information on a screen differs from viewing printed information. With a screen display:

- it is harder to browse and remember one's context;
- the screen is more difficult and tiring to view than printed media;
- the available screen window "page" size can make it difficult to grasp information that is perfectly comprehensible in printed form; and
- the display format could be simplified and streamlined to assist users unfamiliar with the entire notion of a controlled vocabulary.

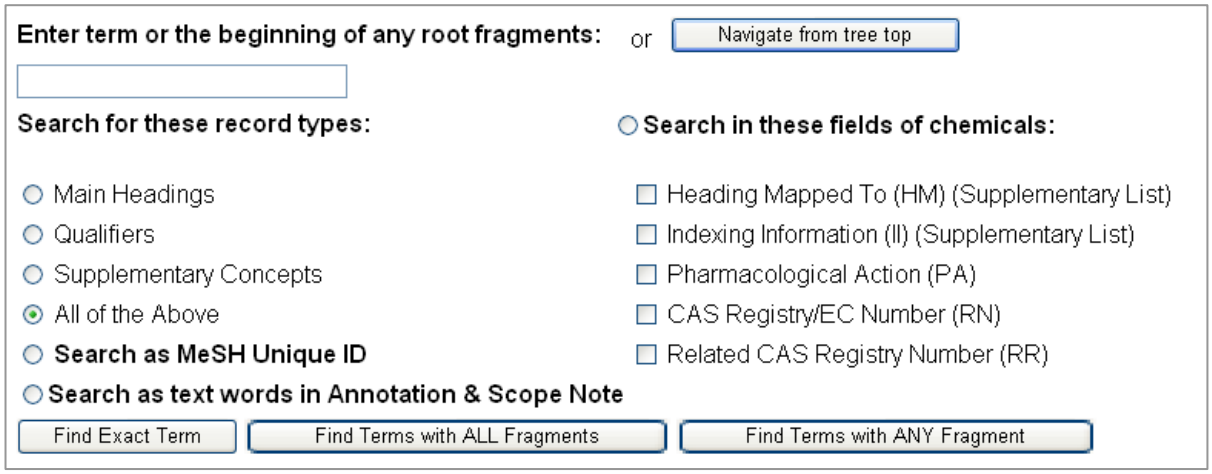
A great deal of research has been performed and is continuing concerning user search behavior and cognitive learning styles in the context of interactive search systems, library catalogs, and hypermedia. In addition to existing standards in computer-human interaction, working groups are producing draft standards for user interfaces, screen layouts, menus, windows, webpages, and other interface formats. Designing for accessibility for people with disabilities is also a growing research and standardization area. Usability and accessibility standards ***should*** be applied rigorously to all controlled vocabulary display designs. See the Bibliography for sources of information on interface design and accessibility.

9.4.2.2 Keyword Searching of Controlled Vocabularies

A user first approaches a controlled vocabulary in order to locate desired terms for use in subject searching or indexing. Electronic controlled vocabularies ***should*** provide keyword searching of all of the terms in the vocabulary. This search serves the same purpose as the permuted and/or rotated indexes that are commonly used in print formats.

Most keyword search options provide several alternatives for users in what parts of the controlled vocabulary are searched, as illustrated by Example 163.

Example 163: Keyword and other search access to the controlled vocabulary



Source: Medical Subject Headings (MeSH) Browser [7]

Keyword searches ***should*** retrieve all occurrences of the term, especially in compound terms.

Example 164: Display resulting from a keyword search of a controlled vocabulary

NOTE: This is a list of terms that is returned by a keyword search on the word corn:

Broom corn (millet)
Broom corn (sorghum)
Corn (maize)
Corn (oil)
Corn (soft wheat)
Corn (triticum)
Corn (zea)
Corn Belt (USA)
Corn cob mix
Corn flour
Corn salad
Corn starch
Dent corn
Flint corn
High fructose corn syrup
Indian corn
Kafir corn
Soft corn
Sweet corn
Waxy corn

Source: AGROVOC [1]

9.4.2.3 Term Detail Display

Users ***should*** have the option of viewing a term's history, scope note, or definition, as well as all term relationships—equivalence, hierarchical, and associative—plus any specialized relationships created for the controlled vocabulary. These recommendations are especially important for screen displays of controlled vocabularies because it is very difficult to get an overview of the structure of the vocabulary from the small view that is generally available.

9.4.2.4 Hierarchy Level Display

It ***should*** be possible to display the hierarchies for controlled vocabularies at various levels. For example, if the hierarchy of a full controlled vocabulary has five levels, it ***should*** be possible to display an outline of the first three levels, as shown in the following example:

Example 165: Display of selected hierarchy levels

Sounds
>(1)Music
>>(2)Blues music
>>(2)Electronic music
>>(2)Folk music
>>>(3)Folk songs
>>(2)Gospel music
>>(2)Jazz
>>(2)Musical notation
>>(2)Singing
>>>(3)Caroling
>>>(3)Children singing
>>>(3)Serenading
>>>(3)Singing games
>>(2)Songs
>>>(3)Ballads
>>>(3)Folk songs
>>>(3)National songs
>(1)Whistling

Source: Thesaurus for Graphic Materials (TGM I) [13]

A tree structure with notation facilitates search expansion by enabling the straightforward inclusion of narrower terms. (See section 9.3.4.2).

9.4.2.5 Pick Lists

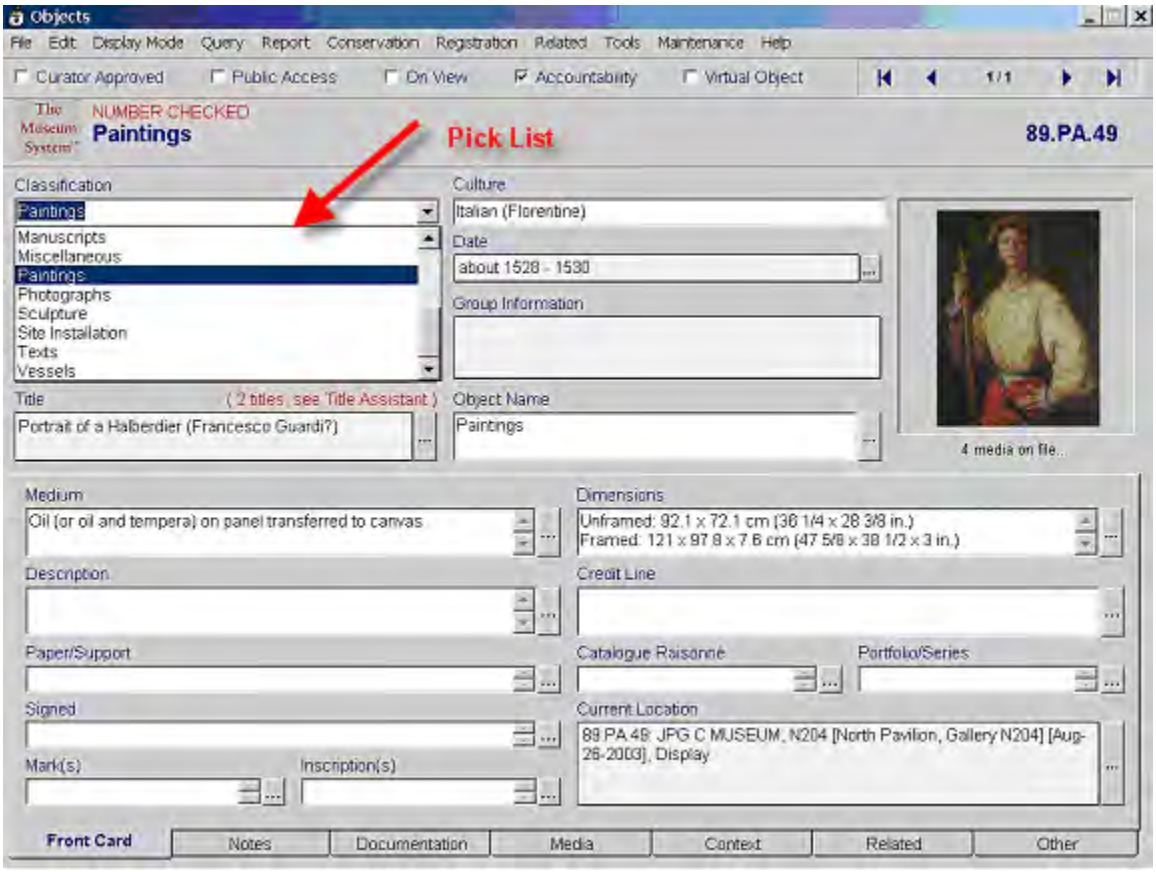
Some electronic implementations of controlled vocabularies use pick lists as a way to lead users to a small set of choices of terms for a given application. These are also called drop down menus. When the user clicks on an arrow associated with the list, the full list of choices will be displayed. This makes it very easy for users to select the appropriate terminology to represent various aspects of a content object either as it is being described or as part of the process of constructing a query.

Pick lists are frequently used to display a vocabulary that has been sub-divided into facets, simple lists of terms that are appropriate for various metadata elements such as:

- Formats (e.g. full text, PDF, CD-ROM, print)
- Languages
- States
- Countries
- Intellectual level (e.g. K-12, news, scholarly journal)

Each of these pick lists ***may*** comprise an entire controlled vocabulary in itself or ***may*** be part of a larger vocabulary.

Example 166: Pick list selection



Source: The J. Paul Getty Museum's implementation of The Museum System software by Gallery Systems [12]

9.4.3 Web Format – Special Considerations

This section discusses the special considerations for controlled vocabularies displayed using Web browsers and Web-related technologies. The screen format considerations discussed in section 9.4.2 would also apply to Web format vocabularies.

9.4.3.1 Path Hierarchy Display

Web displays often include a simplified hierarchy "path" on each display screen. This enables the user to place the current display in the hierarchy context and, when hyperlinked, to navigate back up the hierarchy, and even jump multiple levels.

Example 167: Simplified path hierarchy Web display

Books > Subjects > History > Europe > Monaco

Source: Amazon.com

Path hierarchical displays can also be used as an alternative method for viewing a vocabulary's hierarchies in context.

Example 168: Path hierarchy in context

Organisms
Organisms > Eukaryotes
Organisms > Eukaryotes > Animals
Organisms > Eukaryotes > Animals > Aquatic animals
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Blue whale
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Gray whale
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Humpback whale
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Minke whale
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Right whales
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Right whales > Bowhead whale
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Right whales > Northern right whale
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Baleen whales > Right whales > Southern right whale
Organisms > Eukaryotes > Animals > Aquatic animals > Aquatic mammals > Marine mammals > Whales > Fossil whales

Source: Thomson Gale Master Thesaurus, Synaptica® software by Synapse Corporation [15]

9.4.3.2 Web Navigation Techniques

Use of hyperlinks and other Web navigation tools can make it much easier to move around a controlled vocabulary, enabling users to:

- move easily around the hierarchies, navigating from level to level to find the desired level of specificity;
- link to and view definitions, scope notes, and history notes only when needed;
- move easily from one type of display to another (e.g. from an alphabetical listing to a hierarchical structure, or from a graphical concept map to a term detail display); and
- link to the full record for each term in the controlled vocabulary from within any display.

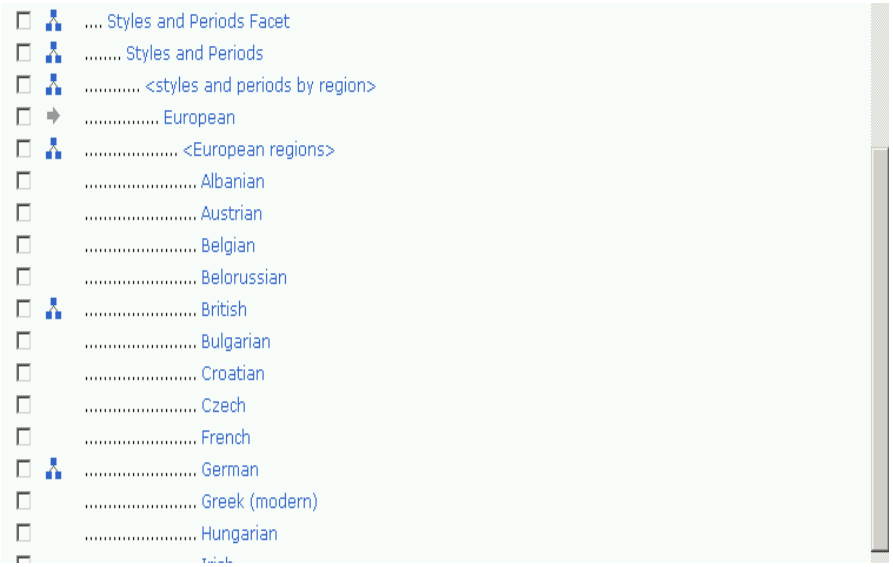
9.4.3.3 Browsing

Controlled vocabularies that have been implemented for Web access ***should*** allow the user to browse through an alphabetical listing of the terms. The user ***may*** move through the listing by:

- using the scroll slider bar on the right-hand edge of the window;
- using the up and down navigation arrows **(↑ ↓)**;
- using the page up **(PgUp**) and page down **(PgDn**) keys; or
- clicking on a plus sign **(+)** or equivalent symbol next to a term to expand or collapse hierarchy levels.

Example 169: Web Browse Display

NOTE: The tree structure symbol in this display works like the plus sign. Clicking on the symbol expands the hierarchy display.



Source: Art & Architecture Thesaurus® [3]

Some browse environments ask the user to select the letter of the alphabet corresponding to the first letter of the desired term. When the user clicks on the letter, the Browse function presents a section of the controlled vocabulary beginning with the desired letter.

Example 170: Browsing by letter of the alphabet

*NOTE: Here is a typical list of terms that would be displayed by selecting **A** from the list below.*

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Achievement
Achievement Comparison
Achievement Gains
Achievement Incentives
Achievement Level
Achievement Losses
Achievement Motivation
Achievement Need
Achievement Prediction
Achievement Rating
Achievement Tests

Source: Thesaurus of ERIC Descriptors [14]

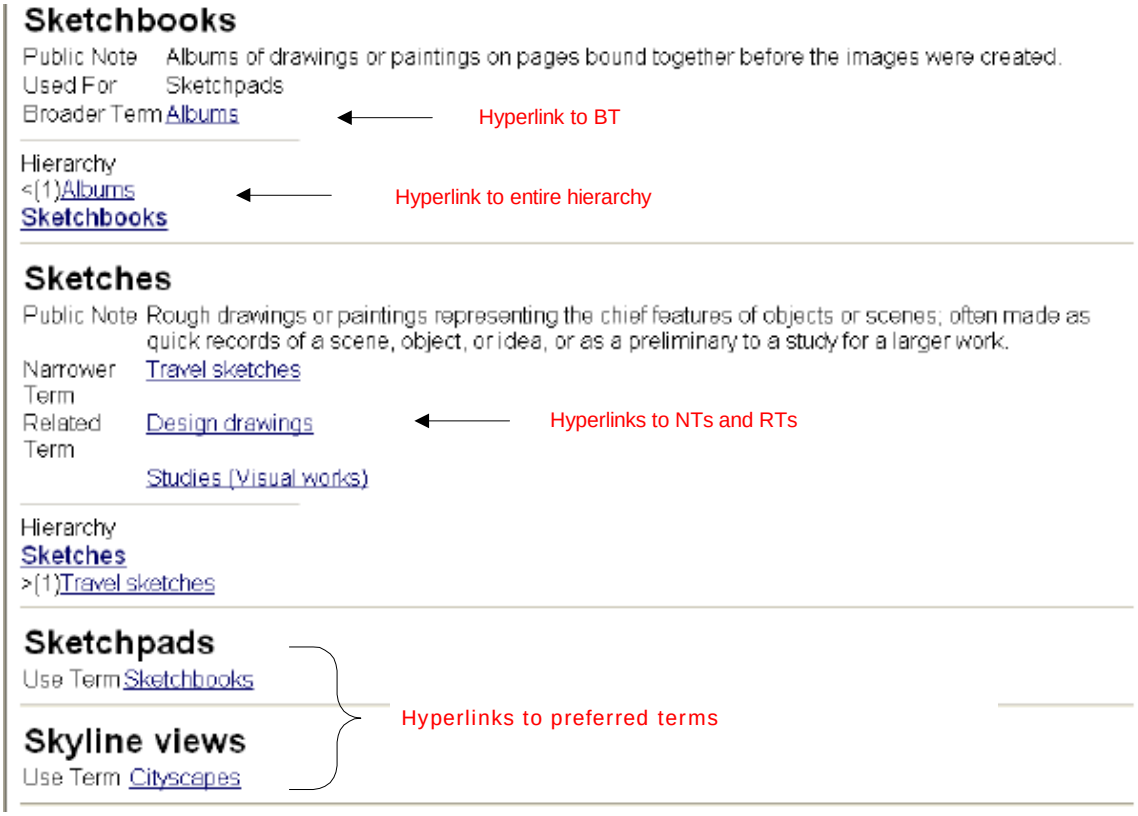
9.4.3.4 Hyperlink Navigation

A major benefit for all types of users of Web implementations of controlled vocabularies is the ability to use hyperlinks to move to various parts of the controlled vocabulary. A hyperlink is an HTML convention that enables a user to click on a pre-established link to navigate to a different part of a document, another content object, and/or to other websites or parts of websites as required. A typical

Web-enabled controlled vocabulary might include links from a term to the following portions of a controlled vocabulary:

- Narrower and broader terms
- Related terms
- Individual term records
- Scope notes
- History notes
- Facets
- Tree structures
- Classification code or structure

Example 171: Hyperlinks to other displays



Source: Thesaurus for Graphic Materials (TGM I) [13]

9.5 Documentation

Thorough descriptions of the controlled vocabulary and its use ***should*** be documented. Separate documentation ***may*** be desirable for vocabulary maintainers, indexers, and searchers.

With print controlled vocabularies, this documentation ***should*** be part of the introduction. If the controlled vocabulary is available online, the user documentation ***should*** also be available online. The documentation for electronic systems is frequently available as a PDF document that can be downloaded and printed by users for future reference. In software applications, the documentation ***may*** be available as in context online help.

All controlled vocabularies ***should*** include comprehensive supporting documentation that clearly states:

- the **purpose** of the controlled vocabulary;
- the **scope**, i.e., the subject field(s) covered, with core and fringe areas separately identified;
- the **meaning** of all conventions, abbreviations, and any punctuation marks used in nonstandard ways;
- the **rules and authorities** adopted in selecting the preferred forms of terms and in establishing their relationships;
- whether the controlled vocabulary **complies with a national or international standard** for controlled vocabulary construction;
- the **filing rules** employed, citing an appropriate standard or guideline when used;
- the **total number of terms**, with separate totals for terms and entry terms;

  NOTE: The total number of terms indicates the size of the indexing/searching vocabulary. The ratio of entry terms to total terms indicates how accessible the controlled vocabulary is to end users, i.e. how easily a user can find the preferred entry term for a concept.

- the **date** on which the controlled vocabulary was last updated;
- a statement on the **updating policy**;
- the **contact information** (e.g., name and address) of the responsible organization to which comments and suggestions can be sent; and
- any special online navigation conventions or **special options**.

10 Interoperability

10.1 The Need for Interoperability

Two conflicting interests drive the development and use of controlled vocabularies. On the one hand, communities develop controlled vocabularies specific to their concepts, terminology, and needs as a means both of accessing data that is important to them and of codifying the content specific to their domain. On the other hand, searchers and those who support them want to use a single search to find resources in databases serving different domains and accessed by different controlled vocabularies, across which there may be no consensus regarding concepts, terminology, and knowledge organization.

Interoperability is the goal of taking vocabularies which in most cases were intended to stand alone and relating them to each other in sufficient detail to permit searches drawn from one vocabulary to be effective in another; resources described with one vocabulary to be described successfully with another; and vocabularies tied to particular domains to be drawn together in larger, more comprehensive conceptual and terminological maps. The needs of different controlled vocabulary users are summarized in Table 2.

Table 2: Interoperability needs

Who	Need
Searchers	**Metasearching** of multiple content resources using the searcher's preferred query vocabulary
Indexers	**Indexing** of content in a domain using the controlled vocabulary from another domain
Content Producers and Distributors	**Merging of two or more databases** that have been indexed using different controlled vocabularies
User Communities Indexers Content Producers	**Merging of two or more controlled vocabularies** to form a new controlled vocabulary that will encompass all the concepts and terms contained in the originals
Multilingual user communities Global organizations	**Multiple language** searching, indexing, and retrieval

Each of these needs presents special problems for the controlled vocabulary designer. Sections 10.2 through 10.7 review some of the problems posed by these requirements for interoperability. Sections 10.8, 10.9, and Appendix D discuss some of the methods that have been tried to address interoperability challenges.

10.2 Factors Affecting Interoperability

Some classic problems can arise when an attempt is made to use a single controlled vocabulary across multiple domains or systems, or when merging multiple controlled vocabularies. Success in meeting the goals of interoperability is conditional on several factors:

a) **The similarity of the content subject matter in the different domains**
This is especially important when using a single vocabulary to index or search across multiple content resources.

b) **Different controlled vocabularies used to index content from similar domains**
This situation arises when different controlled vocabularies are used to represent the content of one or more domains. For example, both *Medical Subject Headings* [7] and *LC Subject Headings* [33] are used to index the same medical texts. Because the terms included in the various controlled vocabularies can vary in format and specificity and the hierarchical structures can be different—even if they adhere to the Standard—retrieval results can be unpredictable across the different domains.

c) **The degree of specificity or granularity of the controlled vocabularies used to index different content domains or databases**
Because of inherent differences in the specificity of the terms used to index content, searching a more specific or granular term selected from one controlled vocabulary may not return the desired results in another database because the indexing was done at a broader level, e.g., *Education Index* [27] uses the term **Religious schools**, whereas another index might use **Private schools** to represent that concept.

d) **Differences in how synonyms and near-synonyms are handled**
Different classes of users may find that the available terms do not meet their needs; for

example, there may not be a sufficiently rich entry vocabulary. Similarly, users may not be able to locate the needed terms because there aren't enough cross references of various sorts included in the vocabulary.

e) **The search methodologies expected by the databases being searched**
Libraries have addressed this problem through the use of Z39.50 [23], a common search protocol, to perform searches across multiple Online Public Access Catalogs (OPACs). However, even with such a common protocol, the interpretation of search terms and the search results that are returned could be entirely different depending on whether a particular database expects precoordinated searching or postcoordinated searching. For example, one controlled vocabulary might have been developed to assist Web navigation while another was constructed to provide terms for postcoordinated queries using Boolean operators.

f) **The literary, organizational, and user warrants used in developing the vocabulary**
If the warrants for each database / system are sufficiently different, there could be little commonality among the terms across the vocabularies or there could be different meanings for the same term without distinguishing qualifiers (see section 6.2.1).

g) **The intended purpose of the databases or systems**
When purposes or intended audiences are very different, users might be disappointed with their search results. For example, the results obtained from a database of study aids for K-12 students would be very different from a search of a high-end scientific/technical database such as *PubMed [36]*.

10.3 Multilingual Controlled Vocabularies

NOTE: For the purposes of this Standard, multilingual controlled vocabularies are included as a special case of interoperability. For all other issues related to multilingual vocabularies consult ISO 5964, Guidelines for the establishment and development of multilingual thesauri.

In many cases, the term lists that are combined to form a multilingual vocabulary have been developed independently by groups in each country or language area that will participate in the multilingual controlled vocabulary. Consequently, the work to develop and maintain multilingual controlled vocabularies has many issues in common with other interoperability efforts.

The earliest controlled vocabularies to be developed were almost always monolingual or intended to serve the users of a particular language, with perhaps a few "loaner" words borrowed from other languages as required. As internet access has become global and countries such as Canada have mandated a bilingual environment, it has become imperative to provide controlled vocabularies that include terms and their equivalents for users of different languages. Further, disciplines such as music and art that are highly international in nature require terms selected from multiple languages to represent the content of the domain accurately.

Issues of specificity and cultural context play a significant role in the selection of terms and the creation of various types of relationships in multilingual controlled vocabularies. In addition, the mapping process (that is, the translation of one term into an equivalent term in another language or the establishment of correspondences between controlled terms in lists from two different languages) requires term level, syntactic level, and semantic level linking to ensure that the desired concept is conveyed into the target language as accurately as possible. Lastly, maintenance of the relationships between terms in different controlled vocabularies have to be accounted for, especially when different agencies are responsible for expanding and developing each vocabulary. The example below illustrates an English-French-Spanish multilingual vocabulary in the area of education.

Example 172: English / French / Spanish multilingual vocabulary (English view)

Abstracting
MT 5.35 Documentary information processing
FR Rédaction de résumés
SP Elaboración de resúmenes análiticos
BT1 Documentary analysis
BT2 Information processing
RT Abstracts

Source: UNESCO Thesaurus [16]

10.4 Searching

Searching across systems and domains where multiple controlled vocabularies are in use can affect both search precision and recall. (See section 5.3.6 for a discussion of controlled vocabulary impact on search precision and recall.)

To make different controlled vocabularies compatible for searching, much work has focused on concept and/or term mapping using human intellectual work as well as machine-aided processes. The work has generally taken one of three different approaches:

- **Combined controlled vocabularies** – Individual controlled vocabularies, each developed for a particular content domain, have been integrated to form a larger controlled vocabulary that is applicable to all of the specific domains. The *Unified Medical Language System* (UMLS) [18] is an example of this approach.
- **Micro or specialized controlled vocabularies** – A generic controlled vocabulary structure is defined and agreed upon. This superstructure then is used as the guideline and template for creating multiple, highly specific controlled vocabularies to meet the needs of various content domains. The Library of Congress has used this approach for some of its domain-specific controlled vocabularies, such as the *Thesaurus for Graphic Materials (TGM) [13]* and the *Thesaurus for the Global Legal Information Network [38].*
- **Switching languages** – Various mappings (see section 10.9.2) are employed to translate the terms contained in one controlled vocabulary into equivalent terms in another. The Renardus Service [37], which uses the Dewey Decimal Classification as a switching language, is an example of this approach.

With any of these approaches, it is important to take into account the mode of search access for each system. If, for example, one system relies exclusively on precoordinated terms and the associated browse / navigate methods of locating content while another assumes postcoordinated searching, the results will be unpredictable at best.

10.5 Indexing

Assigning terms to represent concepts contained in content objects frequently requires consulting a controlled vocabulary. Indexers or others who have become familiar with a particular controlled vocabulary ***may*** elect to use it for another content domain. This process can work quite effectively if the new domains share several aspects with the original domain for which the vocabulary was developed. A single controlled vocabulary ***may*** be used to index content from another domain if all the conditions described in section 10.2 are satisfied.

10.6 Merging Databases

Merging databases that have been indexed using different controlled vocabularies can present nearly insurmountable problems. It is often the case that there are compelling reasons for merging databases regardless of the problems encountered. Occasionally, the organization will recognize the

issues created by multiple vocabularies and develop an entirely new controlled vocabulary to meet the needs of the databases to be merged. The content will then be re-indexed using the new terms. Obviously, this is a very expensive approach, but it could be the only realistic solution.

10.7 Merging Controlled Vocabularies

Merging controlled vocabularies is often required when there is a need to merge databases as described above. However, merging controlled vocabularies could also be undertaken to simplify the task of indexers and others who use one or more controlled vocabularies to index multiple content collections.

If the controlled vocabularies to be merged are very different in their specificity, a lot of work could be required to create new hierarchies and other relationships among terms to make the resulting vocabulary structurally consistent.

Lexicographers are frequently needed to address issues of semantics and form, specificity, and usability, especially when the same or similar terms are used with different meanings in the vocabularies being merged or when different terms are used to represent identical content.

10.8 Achieving Interoperability

Although most controlled vocabularies developed in recent years comply with either the ISO 5964 standard or the ANSI/NISO Z39.19 standard, differences and variability in terminology and syntax still cause problems in cross-domain or cross-system interoperability.

The need for interoperability arose soon after the first controlled vocabularies were developed. Considerable work, both practical and scholarly, has been done to develop methods that will enable controlled vocabularies to be used across multiple databases and systems and to enable them to be shared by indexers and searchers. All are still evolving. Further discussion of these methods can be found in Appendix D.

10.9 Storage and Maintenance of Relationships among Terms in Multiple Controlled Vocabularies

Once the various relationships among terms across multiple controlled vocabularies are identified, some provision ***must*** be made for retaining and maintaining these relationships for future use. Several options are being used to store this information and to make it available in support of interoperability initiatives.

10.9.1 Authority Records

Special fields in authority formats ***may*** be used to store information about the relationships among terms in different controlled vocabularies.

For example, Northwestern University [35] has an ongoing effort that maps corresponding *Library of Congress Subject Headings (LCSH) [33]* and *Medical Subject Headings (MeSH) [7]* and stores the mapping data into linking entry fields in the LCSH and MeSH authority records.

10.9.2 Vocabulary Mapping

This option requires designating one controlled vocabulary as the master with others as subsidiaries. The goal is to map the terminologies of the various controlled vocabularies to be included against a common classification scheme. Two key features are needed to enable development of such a mapping:

1. The vocabularies overlap with each other in terms of coverage.
2. The terms are ordered by means of hierarchical structures and other types of relationships that are shared by all. This enables them to be mapped against a shared classification scheme.

Example 173: Vocabulary mapping between ERIC and LCSH

ERIC Thesaurus Term	LCSH Term
Echolocation	Echolocation (Physiology)
Crack	Crack (Drug)
Radiology	Radiology, Medical
Rh factors	Rh factor

Source: Vizine-Goetz, et al; Vocabulary Mapping for Terminology Services [41]

Mappings can also be used for special situations, such as multilingual controlled vocabularies as shown in the following example:

Example 174: Multilingual vocabulary mapping

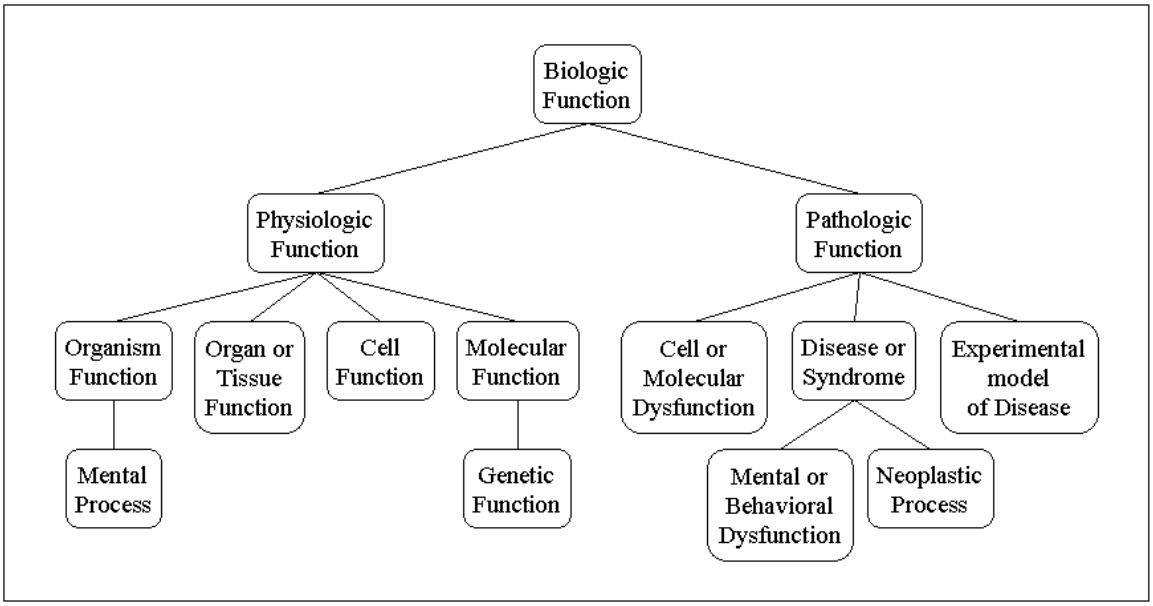
	English (LCSH)	Deutsch (SWD)	Français (RAMEAU)
☐	All terrain cycling	Mountainbiking **or** Geländeradsport	Vélo tout terrain **or** Vélo tout terrain
☐	All terrain cycling -- Training	Training **and** Geländeradsport	Entraînement **and** Vélo tout terrain
☐	Cycling	Radsport **or** Radfahren	Cyclisme
☐	Cycling accidents	Sportverletzung **and** Radsport	Cyclistes -- Lésions et blessures
☐	Cycling for women	Frauensport **and** Radsport	Cyclisme féminin
☐	Cycling -- Law and legislation	Sportrecht **and** Radsport	Droit **and** Cyclisme
☐	Cycling -- Records	Rekord **and** Radsport	Records **and** Cyclisme
☐	Cycling -- Safety measures	Unfallverhütung **and** Radsport	Mesures de sécurité **and** Cyclisme
☐	Cycling -- Training	Training **and** Radsport	Entraînement **and** Cyclisme

Source: Multilingual Access to Subjects (mac) [6]

10.9.3 Semantic Network

A semantic network ***may*** be used to cluster terms from one or many controlled vocabularies according to some underlying organizational structure that is common to all of the candidate controlled vocabularies. This is usually achieved by grouping the terms according to their relationship to some conceptual scheme. Example 175 illustrates the semantic network for the *Biologic Function* hierarchy from the UMLS.

Example 175: Semantic network for clustered terms from a combined vocabulary



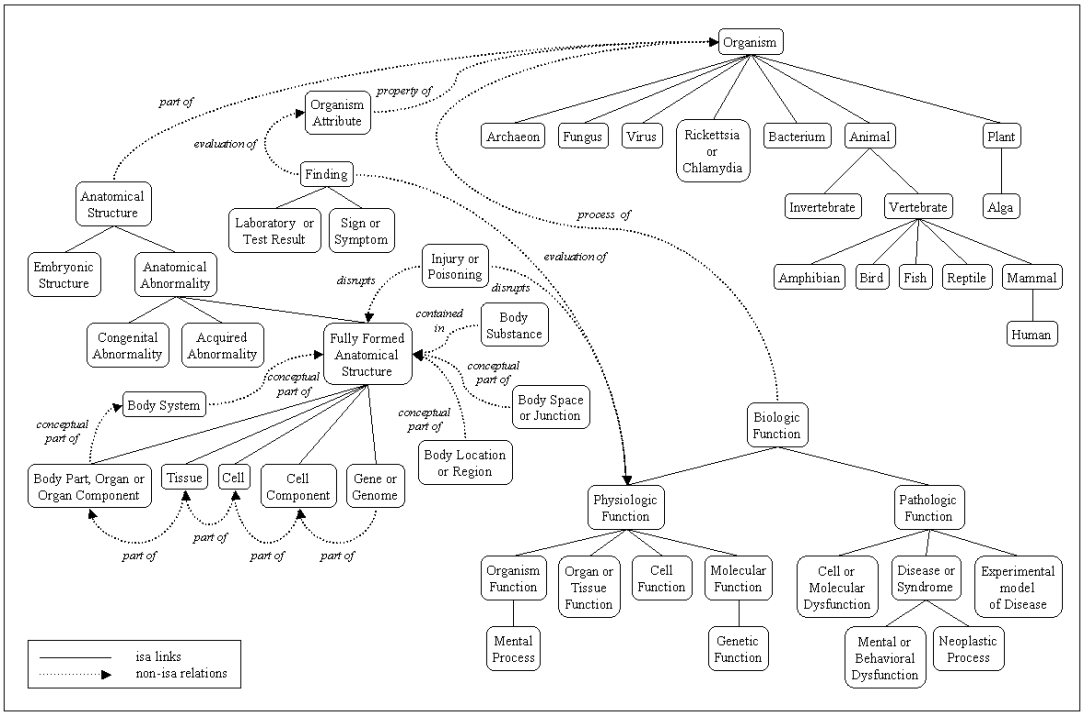
NOTE: This diagram is a portion of the UMLS semantic network.

Source: UMLS Semantic Network, Figure 1 [17]

The semantic type "Biologic Function" has two children—"Physiologic Function" and "Pathologic Function"—and each of these in turn has several children and grandchildren. Each child in the hierarchy is linked to its parent by the "IsA" link.

The following example illustrates both the hierarchical relationships and the semantic relationships among the concepts in the semantic map. This diagram is also extracted from the documentation provided by the UMLS for the semantic network.

Example 176: Semantic network showing types of relationships among concepts



Note: This diagram is a portion of the UMLS semantic network.

Source: UMLS Semantic Network, Figure 3 [17]

In establishing a semantic network, it is critical that the **Types** of concepts and **Types** of relationships be defined. The UMLS network, for example, has defined 135 semantic types and 54 relationships.

Sometimes the semantic network is derived from the controlled vocabularies themselves and in other cases it is derived from the target databases. In the cases where a semantic network is derived from the underlying database, one result could be generation of synonym rings for use during retrieval. When these are available, searchers usually have the option of selecting the desired semantic closeness of the terms to be used in searching. For example, the choices might be:

- exact synonyms
- near synonyms
- closely related terms
- loosely related terms
- antonyms

10.9.4 Lexical Database

A lexical database ***may*** be used to associate terms from multiple controlled vocabularies into clusters of related concepts. Various types of relationships ***may*** be accommodated including synonyms, antonyms, hierarchical relationships of various sorts, IsA, and HasA. WordNet™ is an example of a lexical database. Example 177 is the result of a search on the term *bridge* in the WordNet database:

Example 177: Multiple meanings for the same noun in a lexical database

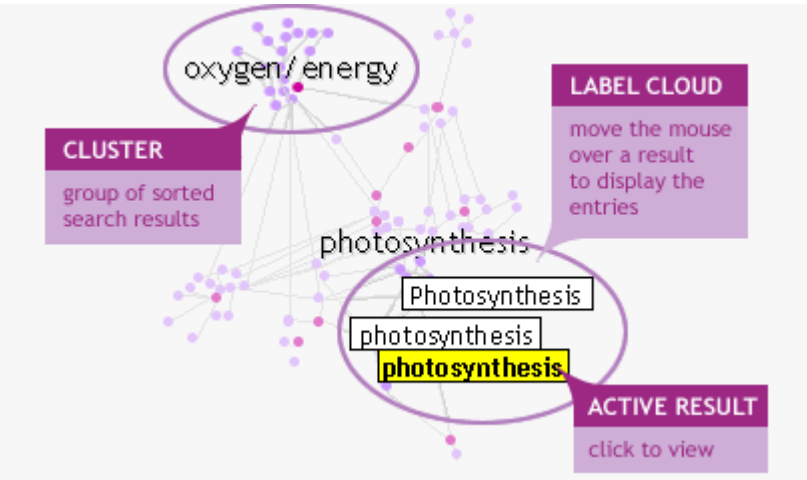
1. **bridge, span** -- (a structure that allows people or vehicles to cross an obstacle such as a river or canal or railway etc.)
2. **bridge, bridge circuit** -- (a circuit consisting of two branches (4 arms arranged in a diamond configuration) across which a meter is connected)
3. **bridge** -- (something resembling a bridge in form or function; "his letters provided a bridge across the centuries")
4. **bridge** -- (the hard ridge that forms the upper part of the nose; "her glasses left marks on the bridge of her nose")
5. **bridge** -- (any of various card games based on whist for four players)
6. **bridge** -- (a wooden support that holds the strings up)
7. **bridge, bridgework** -- (a denture anchored to teeth on either side of missing teeth)
8. **bridge, nosepiece** -- (the link between two lenses; rests on nose)
9. **bridge, bridge deck** -- (an upper deck where a ship is steered and the captain stands)

Source: WordNet [19]

Once a sense is selected, the user can then view the term's synonyms, coordinate terms, kind-of, part-of, has-part, domain, familiarity, etc.

Other lexical databases such as Xrefer use a graphical approach to displaying the relationships among concepts. Example 178 is a sample of Xrefer's concept map of the term *photosynthesis*:

Example 178: Lexical database term mapping



NOTE: This view includes superimposed callout boxes explaining what the user sees and how to use the mouse to view underlying details and content.

Source: Xrefer Concept Map [11]

11 Construction, Testing, Maintenance, and Management Systems

11.1 Construction

Construction of controlled vocabularies is a time-consuming, labor-intensive process, especially if the domain to be covered is broad and the terminology in use is rich and complex. The work involved is justified because the use of controlled vocabularies helps to ensure consistency in indexing and promotes more satisfactory retrieval.

11.1.1 Avoid Duplicating Existing Vocabularies

Before embarking on the substantial work of creating a new controlled vocabulary, the sponsoring organization ***should*** ascertain whether an existing controlled vocabulary covers the same or an overlapping domain of knowledge. (See the Bibliography for listings of controlled vocabularies and Web portals about controlled vocabularies.) Complete duplication of subject coverage is rare, but access to one or more controlled vocabularies in related fields can frequently serve as a useful starting point. For example, a controlled vocabulary developer in the field of malpractice law should begin by consulting existing controlled vocabularies in medicine and law, which will surely contain many of the terms and relationships required for the more specialized, interdisciplinary controlled vocabulary.

11.1.2 Determine the Structure and Display Formats

If possible, the structure of the controlled vocabulary (e.g., flat, generic structure, hierarchical display, or graphic display) and display format (print, online, web-enabled) ***should*** be decided before terms are collected and considered as candidates for inclusion.

As shown in section 9, the display format of the controlled vocabulary affects the types of cross-references and relationship indicators that are provided. For example, if a tree structure with classification notation is adopted, the controlled vocabulary might not include BT/ NT references. The format ***should*** therefore be selected before relationships among terms are constructed. (Although it is possible to convert one controlled vocabulary display format to another, one ***must*** begin with consistent coding of relationships in a single system.)

11.1.3 Construction Methods

Three initial approaches to controlled vocabulary construction are recommended.

11.1.3.1 The Committee Approach

Experts in the subject domain of the controlled vocabulary draw up a list of the key terms in the field and indicate the relationships among them, with assistance from experts in controlled vocabulary design. Lists of terms ***may*** be taken from various sources or submitted by various users / communities. Two main methods of creating controlled vocabularies by committee have been used.

a) **Top Down** – The broadest terms are identified first and then narrower terms are selected to reach the desired level of specificity. The necessary hierarchical structures and relationships are created as the work proceeds.

b) **Bottom Up** – This case frequently occurs when lists of terms have been derived from a corpus of content objects and are then to be incorporated in a controlled vocabulary. As in the case above, the necessary hierarchical structures and relationships are created as the work proceeds, but starting from the terms having the narrowest scope and moving to the more generic ones.

If a new controlled vocabulary is being created, the "top down" approach is preferred. Once a controlled vocabulary is in place, the "bottom up" approach is most often used to add new terms to cover new concepts.

11.1.3.2 The Empirical Approach

The empirical approach has two basic methods:

a) **The deductive method** – Terms are extracted from content objects (by humans or computers; see section 11.1.3.4 for a discussion of machine assistance), optionally during a preliminary stage of indexing, but no attempt is made to control the vocabulary, nor to determine relationships between terms, until a sufficient number of terms has been collected. All terms are then reviewed by a group of experts, preferably consisting of both information and subject specialists. They should first identify terms that represent the broadest classes, and then allocate remaining terms to these classes on the basis of their logical relationships, so that the hierarchies tend to be established on a broader-to-narrower basis. Vocabulary

control should be developed at the stage where hierarchies and other relationships are established, following the principles described in sections 7 and 8.

b) **The inductive method** – New terms are selected for potential inclusion in the controlled vocabulary as they are encountered in content objects. Vocabulary control is applied from the outset. If the vocabulary being constructed is to have some sort of hierarchical arrangements, each term, as it is admitted, is designated as a member of one or more broader classes that are constructed on an ad hoc basis at an early stage. The controlled vocabulary is therefore established on a narrower-to-broader term basis. Controlled vocabulary construction is regarded from the outset as a continuous operation. Assistance from subject experts is strongly recommended; these experts may serve as members of a formal editorial board or committee.

11.1.3.3 Combination of Methods

In practice, more than one of these approaches is likely to be employed at one stage or another during the construction of a controlled vocabulary. For example, hierarchies and other relationships among terms that were first established inductively could later be examined from a deductive viewpoint. Both techniques are essentially empirical, and it ***should*** be accepted from the outset that some decisions regarding the terms and their interrelationships that were made during the early stages of compilation ***may*** have to be revised as further experience is gained. The compilers ***should*** check terms and hierarchies frequently to ensure consistent application of principles in such procedures as the construction of inter-term relationships and the splitting of compound terms (see sections 7 and 8).

11.1.3.4 Machine Assistance

It is assumed throughout this Standard that controlled vocabulary construction calls for intellectual decisions. Machine assistance can be employed, however, for term identification tasks such as the following:

a) **Identification of candidate terms** – Candidate terms may be identified automatically from machine-readable text, e.g., titles and/or abstracts. The number of potential terms should be reduced by use of stop words or a stop list (function words—articles, conjunctions, and prepositions such as "a", "and", and "for"—plus other words considered to be of no value for retrieval). Removing the stop words from consideration is often the first step in machine-aided indexing. The remaining terms should be matched against those already recorded in the controlled vocabulary. Unrecorded terms may be considered as candidates for inclusion.

b) **Registering frequency of term assignment** – In computerized indexing systems, the frequency with which a term has been used in indexing can be registered automatically. Terms with exceptionally high or low scores can be considered as candidates for modification or deletion.

c) **Recording terms from user queries** – Terms found in user queries that do not match one or more terms or entry terms *may* also be considered for inclusion, especially when a given term occurs in multiple queries. To ensure privacy, users ***should not*** be identified when tracking and collecting user queries. Terms that occur in natural language and are common ways of describing a concept are included in the controlled vocabulary as entry terms. An entry term is a word or phrase that is a commonly used name or label for a concept but is not the preferred term selected to represent the concept.

Example 179: Machine-assisted cross-reference to preferred term

acquired immune deficiency syndrome USE **AIDS**

NOTE: In this example, acquired immune deficiency syndrome *is an entry term.* ***AIDS*** *is the preferred term.*

11.1.4 Term Records

An individual record ***should*** be created for every term, and optionally for every entry term, as soon as it is admitted into a controlled vocabulary.

Records for entry terms ***may*** include source notes as well as the date of admission into the controlled vocabulary. For terms, the record ***may*** contain any or all of the following elements:

- term
- source(s) consulted for terms and entry terms.

  *NOTE: This field is especially important for neologisms or unfamiliar terms; it **may** include citations to published sources or the names of personal authorities consulted.*

- scope note
- USED FOR references – to indicate which synonyms, near synonyms, and other expressions are covered by the term.
- nondisplayable variations, e.g., common spelling errors (see section 6.6.2)
- broader terms
- narrower terms
- related terms
- locally established relationships
- category or classification number
- history note, including minimally the date added, as well as the record of changes, if any (see section 6.2.3)

See section 9.3.3 for examples of term records. Section 11.4.1 discusses field definition in controlled vocabulary management systems.

11.1.5 Term Verification

Before a term is admitted into a controlled vocabulary, it ***should*** be validated according to the rules in sections 6 and 7 concerning scope, form, and choice of terms. The compiler ***should*** also review relationships between each new term and other terms in the hierarchy to which it is assigned. The following types of authority within the subject domain ***should*** be checked before candidate terms are accepted for inclusion:

a) technical dictionaries, glossaries, scholarly monographs, reference texts, and encyclopedias;

b) existing controlled vocabularies; and

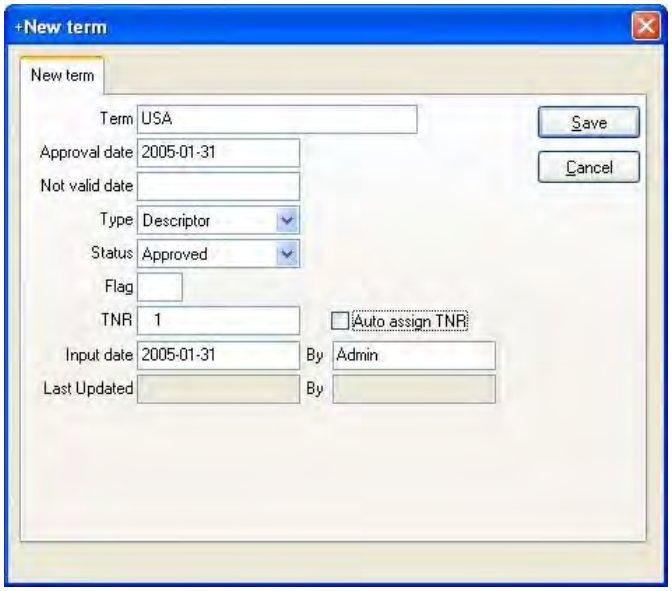
c) classification schemes.

In problematic cases, expert advice ***should*** be sought on the selection of a term from variant forms of terms.

11.1.6 Candidate Terms

Candidate terms (also called provisional terms) are proposed terms that have not gone through all acceptance procedures. These terms ***should*** be marked by a special symbol or phrase in the term record. As soon as a candidate term is approved as a term, the symbol or phrase ***must*** be deleted. In an online system in which the controlled vocabulary is linked to a single database, candidate terms are generally not displayed to users; in controlled vocabularies that are not linked in this way, candidate terms ***may*** be displayed.

Example 180: Editing Template for term entry and maintenance showing term status



Source: MultiTes product demonstration [8]

11.1.7 Levels of Specificity

The addition of highly specific terms is usually restricted to the core area of the subject field covered by a controlled vocabulary because the proliferation of such terms in fringe areas is likely to lead to a controlled vocabulary that is difficult to manage.

Although the cost of computer storage of a very large controlled vocabulary could be insignificant, the human cost of establishing relationships among numerous terms at the periphery of the domain is high. In an organization that deals with content objects covering more than one domain of knowledge, it ***may*** be necessary to develop a number of specialized controlled vocabularies, each linked to, and compatible with, a general controlled vocabulary that has a lower level of specificity, produced by the same organization. A model for this can be found in the United Nations *Macrothesaurus for Information Processing in the Field of Economic and Social Development [34]* or the UMLS *Metathesaurus [18]*. (See also Appendix D.3.)

11.1.8 Unassigned Terms

When hierarchies are being established in a controlled vocabulary, terms that have not yet been used in indexing are frequently admitted into the controlled vocabulary on the grounds that they are necessary to complete a hierarchy (for example, as broader terms), and that they have potential value as indexing terms in their own right.

11.2 Testing and Evaluation of Controlled Vocabularies

The two major reasons for testing and evaluating controlled vocabularies are:

1. To determine if the controlled vocabulary being used to describe content objects provides adequate search results (i.e. high relevance and recall) for most users, and
2. To determine if the controlled vocabulary matches users' expectations of the terms contained therein. For example, if the entry vocabulary is not sufficiently rich, users may not discover their desired terms.

11.2.1 Testing Methods

There is surprisingly little research or literature dealing specifically with the testing and evaluation of controlled vocabularies, which could be due in part to the difficulty of isolating the quality of a vocabulary from the performance of other elements of an information system such as content, technology, search engine, and interface. However, the following methods ***should*** be considered when faced with the challenge of measuring the quality or effectiveness of a controlled vocabulary:

- **Heuristic Evaluation** – An expert or panel of experts ***is*** asked to evaluate a controlled vocabulary. Such an evaluation can be informal and qualitative (e.g., each expert sharing their reactions and opinions) or formal and quantitative (e.g., all experts assigning scores according to a shared list of criteria).
- **Affinity Modeling** – A representative sample of users ***is*** asked to sort a collection of cards (i.e., card sorting) each marked with terms from the vocabulary into groups. The results are then analyzed against the existing term hierarchy. Users ***may*** also be invited to score equivalent and related terms according to level of similarity.
- **Usability Testing** – Holistic evaluation of the information system can yield insight into the efficacy of the controlled vocabulary. Consult the literature in the fields of human-computer interaction and information architecture for general usability guidelines and evaluation methods.

11.2.2 Evaluation Criteria

Soergel [39] has identified several considerations for use in evaluating controlled vocabularies. He suggests you begin by conducting a summary evaluation of the vocabulary's adequacy for the stated purpose, asking questions such as:

- Does the controlled vocabulary contain terms that are synonymous or equivalent without indicating the relationship?
- How well does the display reflect the conceptual analysis (e.g., sequence of concepts on some hierarchical level?
- Are the history notes provided where needed?
- Are all needed facets included?
- Are the terms up-to-date with accepted usage? (timeliness of terms)
- How complete are the entry term synonyms included in the alphabetical list?
- Form of terms:
  - Is term form consistent?
  - Does term form adhere to common usage?
- Are high-quality definitions, scope notes, etc. included?

In addition, there are three additional criteria applicable to vocabularies accessible electronically:

- Are cross-references active hyperlinks?
- Is there an online term index or search by keyword?
- Is the controlled vocabulary database searchable? How does the search work? What fields are searched?

Lastly, Soergel notes some points on procedure for evaluating a controlled vocabulary:

- What went into the construction of a thesaurus? Check sources used.
- Check procedures used in controlled vocabulary development.

- Examine the controlled vocabulary structure and content
- Use knowledge of controlled vocabulary structure for analysis of structure and internal consistency.
- Check against other thesauri and against encyclopedias, dictionaries, or other authoritative sources.
- During the evaluation, collect data on all the criteria in parallel. Keep notes according to the outline of criteria.
- Check how the controlled vocabulary works. Try using it for search requests and indexing content objects. In this case, knowledge of the subject areas is required for participants. Indexing experiments where several people index the same items can be very useful; disagreements can point out problems in the controlled vocabulary.

11.3 Maintenance

This section suggests procedures for adding, modifying, and deleting terms. The term record ***should*** note the date of each change and identify the individual responsible for it.

11.3.1 Updating the Vocabulary

Controlled vocabularies are reflections of language, and they are therefore dynamic instruments. Policies and procedures ***should*** be established for periodic review of terminology, establishment of new terms, and replacement of obsolete terms, especially in fields where the terminology changes rapidly. Controlled vocabulary editors ***should*** update the controlled vocabulary at intervals that will be determined by the frequency and volume of changes made, and by the method of distribution. Implementation of updates ***may*** need to be coordinated with updates of other information systems or controlled vocabularies.

Printed revisions will necessarily be produced less frequently than revisions to a computerized system. Electronic updates have the advantage of being accessible more quickly—instantaneously in some systems.

11.3.1.1 Addition of Terms

Whenever an appropriate term or combination of terms cannot be found in a controlled vocabulary, the indexer or searcher ***should*** nominate a new term as a candidate term (see section 11.1.6). The editor ***may*** want to use a print or electronic form for term nominations to ensure that all needed review information is submitted. Sample forms can be found in Appendix E.

New terms and those nominated for reconsideration ***should*** be reviewed by the editor of the controlled vocabulary, and preferably by an editorial board that functions in the same manner as the group responsible for the original edition. The following decisions ***should*** be made for each term:

a) Should this term be added to the controlled vocabulary, or is it already covered by an existing term or combination of terms? If it is covered, should an additional USE reference be added?

b) Are scope notes or other notes needed? (See section 6.2.2.)

c) If the term is to be added, what is the correct form, considering syntax, number, and spelling? (See section 6.3.)

d) How should the new term be interrelated with existing terms on a hierarchical and related-term basis? (See section 8.)

11.3.1.2 Modification of Existing Terms

Indexers and searchers ***should*** be able to propose modifications to existing terms or their relationships, explaining the rationale and supplying supporting documentation for the proposed changes. Like candidate term nominations, such proposals ***may*** be communicated electronically or via printed forms. Such proposed changes ***should*** be considered by the controlled vocabulary editor

and board, using the criteria for term selection in sections 6 and 7. If a term is modified, the date of the change ***should*** be recorded in the history note (see section 6.2.3), and a USE reference ***should*** be made from the old form to the new form. If the controlled vocabulary is used in an indexing system, the date on which an old term was last assigned ***should*** be included in the history note. If the relationships are modified, a record of the old ones ***should*** be maintained in the history note as well.

11.3.1.3 Deletion of Terms

Overused terms and terms that have been assigned very infrequently in indexing ***should*** be considered candidates for modification or deletion, as both kinds of term are generally ineffective in retrieval. In some cases, an overused term can be replaced by two or more terms of greater specificity, or these can be added to the term's hierarchy as narrower terms.

11.3.2 Vocabulary Updates and Database Records

11.3.2.1 Methods for Handling of Modified or Deleted Terms

Whenever a term is modified or deleted from a controlled vocabulary, the impact of the change on the ability to search previously indexed database records ***must*** be considered, unless the modified or deleted term has never been assigned. The problem of searching modified or deleted terms can be handled in several ways. The choice of method depends upon the size of the database, available human and financial resources, the nature and number of changes to be made, and whether automated procedures exist.

a) A deleted or modified term ***may*** be retained in the controlled vocabulary for retrieval or historical purposes only. If it is retained, it ***must*** be marked, e.g., “for retrieval purposes only,” and the date of its change in status ***must*** be recorded in the history note and displayed to users. If it is superseded by a newer term, a cross-reference to the new term ***should*** be provided. The superseded term ***should*** also be listed in the history note of the newer term, with the dates for which the former is valid.

b) One-for-one term changes, where one term simply replaces another, ***may*** be implemented in the database automatically if resources permit. A history note ***should*** be made in the term record to indicate that the term replacements were made.

c) More complicated term changes ***may*** require human intervention, e.g., where two or more specific terms are added to replace a more general term or become narrower terms to it. Deleting a term in the database without replacing it with another term could, over time, seriously reduce access to content objects if no replacement indexing is carried out. If it is discovered, however, that a term has been established and assigned in error, once the term has been deleted from database records, its record ***may*** be deleted from the controlled vocabulary as well.

   NOTE: In very large systems, the cost of such re-indexing could be prohibitive.

11.3.2.2 History of Changes

Each term ***should*** have a history note recording its date of entry (see section 11.1.4). The note ***should*** also track the history of modifications (see section 11.3.1.2), recording previous forms with reasons for the change. Where obsolete terms are retained in the controlled vocabulary but are not assigned in indexing, the date of replacement and reasons for it ***should*** be given. Where terms are deleted completely, it ***may*** be desirable to maintain a separate file of deleted terms, retaining a record of their scope and history in order to track such decisions in case a term from this group is later reconsidered. The following examples illustrate various types of history notes.

Example 181: Notation for a term's date of entry

NOTE: This vocabulary is a merger of two vocabularies. The dates indicate when each of the original vocabularies added the term.

HELIOS FACILITY

INIS: Jul 1981; ETDE: Jul 1979

(Large CO2 laser facility at Los Alamos for laser fusion experiments.)

Source: EDTE/INIS Joint Thesaurus [4]

Example 182: Notation of valid use dates for a term no longer in use

NOTE: The years in parentheses indicate the dates for which the former term was valid.

Action Programs (Community) (1966-1980)

USE **Community Action**

Source: Thesaurus of ERIC Descriptors [14]

Example 183: Notation for a term replacing an older term

FELDSPARS

(A group of abundant rock-forming minerals. From November 1976 till February 1997 ALBITE was a valid ETDE descriptor; from June 1977 till March 1996 MICROCLINE was a valid ETDE descriptor.)

UF *albite*

UF *microcline*

Source: EDTE/INIS Joint Thesaurus [4]

Example 184: Notations for a term with multiple historical changes

MeSH Heading	Fertilization in Vitro
Previous Indexing	Ectogenesis (1975-1979)
Previous Indexing	Fertilization (1966-1979)
Online Note	use FERTILIZATION IN VITRO to search TEST TUBE BABIES Apr 1979, use ECTOGENESIS 1978-Mar 79
History Note	79; TEST-TUBE FERTILIZATION was TEST TUBE BABIES see FERTILIZATION IN VITRO Apr-Dec 1979,was TEST TUBE BABIES see ECTOGENESIS 1978-Mar 79

Source: Medical Subject Headings (MeSH) [7]

Example 185: Notation for a term whose allowed use was modified

NOTE: This history notes indicates that from 1986-1990 the term was not authorized and a broader term was used for indexing.

MeSH Heading	Ferula
History Note	91(86)63; was see under PLANTS, TOXIC 1986-90; was heading 1963-85

Source: Medical Subject Headings (MeSH) [7]

11.3.3 Hyperlink Maintenance

When a controlled vocabulary utilizes hyperlinks in screen displays, e.g. in a Web-based system, routine checking and maintenance of the hyperlinks ***should*** be done. There are many automated tools available that can perform this function. Check the Internet for current software tools.

11.4 Management Systems

Because controlled vocabularies are generally used in the context of information storage and retrieval systems, the design or choice of a controlled vocabulary management system is necessarily dependent on and ***should*** be an integral part of (although not necessarily integrated into) the overall design of the information system of which it is a component.

Developers and maintainers of controlled vocabularies ***should*** have access to systems that can deal with the special needs of a controlled vocabulary. It ***should not*** be necessary to adapt the controlled vocabulary to an inadequate system. The system ***should*** be flexible enough to allow controlled vocabulary managers to take advantage of emerging technologies. The following are general requirements and recommendations for features of controlled vocabulary management software to be used by controlled vocabulary maintainers. See also the recommendations in section 9 regarding screen displays and Web features for users.

Here is a list of desirable features for controlled vocabulary management systems:

1. At a minimum support the basic relationships of U/UF, BT/NT, and RT
2. Support the use of Scope Notes and History Notes
3. Provide both hierarchical and alphabetical displays
4. Preferably be non-proprietary
5. Not be limited to a single user
6. Be hardware independent and operate in an open systems or standard operating system environment
7. Require little or no user training.
8. Provide detailed user documentation.

11.4.1 General System Considerations

Systems for managing controlled vocabulary construction and maintenance ***must*** have several basic capabilities in order to provide adequate support for the vocabulary and its users. The key features are listed below:

- **Typography** – The system ***should*** allow printing or screen display in both capitals and lowercase letters, and it ***should*** be able to handle any special characters required by the subject field or language of the controlled vocabulary. (See section 6.7.) Compliance with the *Unicode* [40] standard is recommended for all non-Roman characters.
- **Sorting** – The sorting functionality ***should*** comply with the filing and sorting guidelines presented in section 9.2.6.
- **Display types** – The system ***should*** be capable of handling alphabetical and hierarchical controlled vocabulary displays, where applicable. Additional display types (see section 9.3) ***may*** be warranted depending on user needs.
- **Field definitions** – The system ***should*** provide for the definition of fields other than those enumerated for term records (section 11.1.4) for purposes such as assigning codes from other systems to terms or providing term definitions as well as scope notes.

11.4.2 Term Records and Displays

The Term Record is the single place in a controlled vocabulary where all (or at least most of) the information about a term is collected in a single place for display to the user. The system ***should*** support the following for term records:

- **Term detail** – At a minimum, these records ***should*** contain the data and specify the relationships recommended in this Standard (see section 11.1.4).

  *NOTE: Individual controlled vocabularies **may** require additional data, such as term weights for retrieval, special relationships, or codes from standards such as the MARC 21 Authority Format.*

- **Term length** – There ***should not*** be limits on the number of characters in a term.
- **Preferred and entry terms** – The system ***should*** be able to distinguish between the non-preferred terms (USE references) that ***should*** be displayed to controlled vocabulary users and those that ***should not*** be displayed (e.g., close spelling variations and common typographical errors entered to enhance computerized retrieval).
- **Term relationships** – There ***should not*** be limits on the number or types of relationships (broader, narrower, associative, or equivalence) that can be established for a given term. It ***may*** also be useful to be able to define and name specialized relationship types which are more specific than USE/UF, and BT/NT or RT/RT, e.g., designating abbreviations, or expressing the exact nature of various types of narrower terms, such as the subsidiary of a company. See section 8 for a discussion of term relationships.
- **Hierarchies** – It ***should*** be possible to edit a term's position in a hierarchy, and to simultaneously display multiple hierarchies for a term where these exist.

11.4.3 Cross-References

When cross-reference links are created between terms, the system ***should***:

a) Check for the existence of a term chosen as a broader, narrower, related, or USE term, and check for the validity of references, so that:
   - no cross-references are made to unused "dead-end" terms
   - the only type of reference that leads from a non-preferred to a preferred term is a USE reference.

     NOTE: Non-preferred terms have no BT, NT, or RT references.

   - a term is not related to itself;
   - no conflicting references are made for the same term pairs (e.g., both a BT/NT and RT/RT reference);
   - duplicate cross-references are eliminated except in the case where an entry term links to more than one preferred term, for example:

     AMA USE American Medical Association
     AMA USE American Management Association

b) Create reciprocal references for BT/NT (or BTG/NTG and BTP/NTP), RT/RT, and USE/ UF pairs with the following conditions:
   - If the term referred to by a BT, NT, or RT reference already exists in the vocabulary, the system ***should*** create the reciprocal reference automatically. However, if the term does not exist, the system ***should*** first inquire whether the new term ought to be created.
   - If the reciprocal term of a UF reference does not exist, the system ***may*** create that term. See section 8.1.1 for logical formulations of the relationships.

Example 186: Creation of reciprocal references

editor creates:

cars USE **passenger cars**

system creates:

passenger cars UF cars

editor creates:

passenger cars BT **motor vehicles**

system creates:

motor vehicles NT **passenger cars**

c) Maintain these references reciprocally, so that when a term is modified or deleted, the references to it are changed accordingly in the records for all the terms related to it. For example, if the editor changes passenger cars to automobiles in the above example, the system ***should*** make the same change throughout the controlled vocabulary, e.g., **motor vehicles** NT **automobiles**.

11.4.4 Term Deletion

When a term is deleted, the system ***should*** prompt for verification of the deletion. The system ***must*** make all relevant reciprocal deletions so that all links to and from deleted terms are also deleted . The system ***should*** inform the editor when the deletion creates an orphan term (a term without associative or hierarchical relationships).

11.4.5 Candidate Terms

The system ***should*** include a mechanism for the entry and designation of candidate terms, and candidate relationships, which are not admitted to the controlled vocabulary until reviewed and approved by controlled vocabulary editors. The structure of candidate term records ***should*** be the same as that of term records (see section 11.1.4), because the former ***may*** be upgraded to the latter.

Candidate relationships ***should*** be reciprocated automatically. It ***should*** be possible to retrieve, display, and print out these candidate terms and relationships.

A merge or transfer function that allows for the comparison and merger of two similar terms can be useful in choosing among candidate terms entered by different indexers, or where revisions of a controlled vocabulary are necessary in a quickly changing environment. The preferred term receives the relationships and notes from the non-preferred term, and the non-preferred term becomes a USE reference.

It ***may*** be useful for the system to be able to capture terms used by database searchers, as these can be useful entry terms for the controlled vocabulary. The system ***may*** also provide a mechanism for searchers to propose candidate terms directly.

11.4.6 Error Checking

A good controlled vocabulary maintenance system ***should*** enable error checking by identifying:

- **Duplicate terms** – The system ***should*** be able to detect duplicate terms, whether entry terms or preferred terms. Duplicates detected by the system ***should*** be displayed to the controlled vocabulary editor, who will judge whether they are true duplicates or homographs requiring qualification.
- **Potential duplicates** – The system ***should*** be able to identify potential duplicates with typographic variations (capitals and lowercase letters) and qualifiers, e.g., mercury (metal) and Mercury (planet).
- **Structural inconsistencies** – The system ***should*** be able to check for discrepancies in term relationships, missing reciprocal relationships and inconsistent modification or deletion of terms

11.4.7 Searching and Browsing

The system ***should*** provide the following search and browse capabilities:

- Search by the complete term or a truncation of the term
- Search for non-preferred terms and retrieve references to preferred terms
- Search without case sensitivity, i.e. a search term in capitals ***should*** match a lowercase term or entry term and vice versa
- Browsing within hierarchical and alphabetical displays
- View a term in the context of its relationships and its complete term record from any display

NOTE: This capability is increasingly provided through the use of hyperlinking technology (see section 9.4.3.4).

11.4.8 Reports

Searchers, indexers, and vocabulary maintainers all need reports. Additionally, controlled vocabulary managers need specific “housekeeping” reports. These reports ***may*** be displayed on a screen or printed, but in either case, they ***should*** be clear and legible and ***should*** use space and paper as economically as possible. Controlled vocabulary reports ***may*** include, but are not limited to, the following:

a) Term listings, including:
   - terms added or modified since a certain date
   - terms that do or do not have a particular type of note or relationship (SN, UF, BT, NT, or RT)
   - orphan terms, with no hierarchical or associative relationships
   - top terms

b) Displays of term relationships, including:
   - all the relationships for an individual term
   - hierarchical displays
   - listings at various levels of a hierarchy
   - reports of special links created for a given controlled vocabulary
   - the entire hierarchical context for an individual term

c) Statistical reports on terms and their characteristics, including:
   - the number of preferred terms and all terms in the controlled vocabulary
   - the number of terms featuring a certain characteristic, for example, those that consist of three words or those that end in a certain string of characters, e.g., "itis"
   - the number of terms added, modified, deleted, or merged since a certain date
   - the number of BTs, NTs, UFs, RTs, SNs, or other value in the controlled vocabulary
   - the average number of characters per term and the size of the longest term
   - postings reports for controlled vocabularies linked to databases

11.4.9 Testing and Evaluating Systems for Use with Controlled Vocabularies

Virtually all controlled vocabularies developed today are created and maintained using some type of computer software. The software could be one module of an integrated library system (ILS) or it could be a stand-alone product developed by a vendor that specializes in this type of software.

Some studies have been done which compare the features and strengths and weaknesses of these software packages. Because they become out-of-date so quickly as a result of new releases, new versions, and new products, interested readers should check the Internet for the latest available information.

Appendix A
Summary of Standard Requirements / Recommendations

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

The table that follows summarizes the requirements and recommendations made in this Standard. The referenced section should be consulted for specific details and examples.

Topic	Section	Description
General – Purpose, Concepts, Principles, and Structure	**5**	
Purpose	**5.1**	The purpose of controlled vocabularies is to provide a means for organizing information.
Concepts	**5.2**	
Term	5.2.1	A term is defined to be one or more words used to represent a concept. Terms are selected from natural language for inclusion in a controlled vocabulary.
Content Object	5.2.2	A content object is any item that is to be described for inclusion in an information retrieval system, website, or other source of information.
Indexing	5.2.3	Indexing is the process of assigning preferred terms or headings to describe the concepts and other metadata associated with a content object.
Principles	**5.3**	
Ambiguity	5.3.1	A controlled vocabulary ***must*** compensate for the problems caused by ambiguity by ensuring that each term has one and only one meaning.
Synonymy	5.3.2	A controlled vocabulary ***must*** compensate for the problems caused by synonymy by ensuring that each concept is represented by a single preferred term. *NOTE: A synonym ring is an exception to the above rule. See section 5.4.2 for more information on this type of vocabulary.* The vocabulary ***should*** list the other synonyms and variants as non-preferred terms with USE references to the preferred term.
Semantic Relationships	5.3.3	Various types of semantic relationships ***may*** be identified among the terms in a controlled vocabulary. These include equality relationships, hierarchical relationships, and associative relationships, which ***may*** be defined as required for a particular application.

Topic		Section	Description
	Facet Analysis	5.3.4	Facets are a kind of structural metadata. They ***may*** be applied either to content objects themselves or to the metadata that represent them. Whether to define facets for content objects is the responsibility of the database designer or information architect. This Standard does not address guidelines for determining which aspects or attributes to create as facets for content objects.
	Using Warrant to Select Terms	5.3.5	The process of selecting terms for inclusion in controlled vocabularies involves consulting various sources of words and phrases as well as criteria based on: • the natural language used to describe content objects (literary warrant), • the language of users (user warrant), and • the needs and priorities of the organization (organizational warrant).
	Impact on Information Retrieval	5.3.6	Information retrieval effectiveness is traditionally measured by two parameters: recall and precision. Controlled vocabulary design can have a positive impact on both of these measures.
Structure		**5.4**	
	List	5.4.1	A list (also sometimes called a pick list) is a limited set of terms arranged as a simple alphabetical list or in some other logically evident way.
	Synonym Ring	5.4.2	A synonym ring is a set of terms that are considered equivalent for the purposes of retrieval.
	Taxonomy	5.4.3	A taxonomy is a controlled vocabulary consisting of preferred terms, all of which are connected in a hierarchy or polyhierarchy.
	Thesaurus	5.4.4	A thesaurus is a controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators. Relationship indicators ***should*** be employed reciprocally.
Metadata and Metadata Schemes		**5.5**	There are a number of initiatives underway to address using metadata and a metadata schema to represent the entire contents of the controlled vocabulary. Controlled vocabulary developers ***should*** evaluate these metadata initiatives for relevance to their community and information domain prior to developing any new metadata element sets or schema.

Topic	Section	Description
Term Choice, Scope, and Form	**6**	
Choice of Terms	**6.1**	Many issues need to be considered in selecting terms for a controlled vocabulary including: • The information space or domain to which the vocabulary will be applied • Literary, user, and organizational warrant • Specificity or granularity of the terms • Relationship with other, related controlled vocabularies
Scope of Terms	**6.2**	The scope of terms is restricted to selected meanings within the domain of the controlled vocabulary.
Homographs	6.2.1	The use of homographs as terms in a controlled vocabulary sometimes requires clarification of their meaning through a qualifier. A qualifier is a word or words used with a term to make the specific meaning unambiguous. The qualifier, which is enclosed in parentheses, is part of the term. Use of qualifiers ***should*** be avoided whenever possible because of the problems that parentheticals can cause in filing and in retrieval. Rules for using qualifiers: • A compound term ***should*** be used instead of a single-word term with a parenthetical qualifier. • A qualifier ***should*** be added to each homograph. • A homograph's parenthetical qualifier ***may*** be left out when a term is used in only one of its meanings within a given controlled vocabulary domain and the meaning is obvious to users within that domain. • Parenthetical qualifiers ***should not*** be used to represent compound concepts. • The qualifier ***should*** be as brief as possible, ideally consisting of one word, but ***should not*** be a homograph. • Qualifiers ***should*** be standardized within a given controlled vocabulary. • A qualified term ***may*** have a scope note appended to it. • Qualifiers ***should*** also be added to entry terms when their meaning is ambiguous.

Topic			Section	Description
	Scope Note		6.2.2	A scope note ***should*** state the chosen meaning of a term. It ***may*** also indicate other meanings that are recognized in natural language, but which have been deliberately excluded from the controlled vocabulary. Changes in the scope of terms ***should*** be recorded in History Notes.
		Reciprocal Scope Notes	6.2.2.1	When reference is made to other terms in a scope note, a reciprocal scope note ***should*** generally be provided for each term mentioned.
	History Note		6.2.3	A History Note is used to track the development of terms over time. It is especially important to indicate when and how a term has changed over time. If appropriate, the history note ***may*** also include the date discontinued, the term that succeeded the term, and/or the term that preceded it.
Term Form			**6.3**	
	Single Word vs. Multiword Term		6.3.1	Each term included in a controlled vocabulary ***should*** represent a single concept (or unit of thought). A single concept is frequently expressed by a single-word term but in many cases a multiword term is required to represent the concept.
	Types of Concepts		6.3.2	The concepts represented by terms ***may*** be grouped into general types, such as: • Terms for things and their physical parts • Terms for materials • Terms for activities of processes • Terms for events or occurrences • Terms for properties or states • Terms for disciplines or subject fields • Terms for units of measurement
	Unique Entities		6.3.3	Unique entities, or “classes-of-one,” are usually expressed as proper nouns.
Grammatical Form of Terms			**6.4**	
	Nouns and Noun Phrases		6.4.1	The grammatical form of a term ***should*** be a noun or noun phrase.
		Verbal Nouns	6.4.1.1	Verbs expressed as infinitives (without "to") or participles ***should not*** be used alone as terms. Activities ***should*** be represented by nouns or gerunds.

Topic			Section	Description
		Noun Phrases	6.4.1.2	Noun phrases are compound terms that ***may*** be established as terms if they represent a single concept. Adjectival noun phrases (premodified) are the preferred form. Prepositional noun phrases (postmodified) are also allowed. When possible, noun phrases ***should*** exclude prepositions.
	Adjectives		6.4.2	Adjectives and adjectival phrases used alone ***may*** be established as terms in controlled vocabularies under certain special circumstances.
		Single Adjectives	6.4.2.1	Single adjectives ***may*** be used in a “nominal” way; that is, the noun is obvious from the context or the adjective is used to describe an attribute of the content object other than topic.
		Adjectives Used to Limit the Number of Compound Terms	6.4.2.2	As an alternative to the creation of multiple compound terms, adjectives ***may*** appear as separate terms when designed to be precoordinated in indexing or postcoordinated in searching.
		Adjectives Used as Cross-references	6.4.2.3	Adjectives ***may*** be used alone in general cross references to direct the user to or from a group of terms beginning with a corresponding noun.
	Adverbs		6.4.3	Single adverbs are rarely used. Adverbs such as "very" or "highly" ***should not*** be used alone as terms.
	Initial Articles		6.4.4	Initial articles in terms ***should*** be avoided except when required to convey essential information.
		When to Remove Initial Articles	6.4.4.1	Remove the initial article when the term is clear without it. Otherwise, use a parenthetical qualifier.
		When to Keep Initial Articles	6.4.4.2	If the initial article is an integral part of a proper name, and should be searchable, it ***should*** be included in the term in direct order. Otherwise, omit the article. In cases where the article is included in direct order, a cross-reference from the element following the article ***should*** be provided.
Nouns			**6.5**	
	Count Nouns (plural)		6.5.1	Count nouns are names of objects or concepts that are subject to the question "How many?" but not "How much?". These ***should*** normally be expressed as plurals.
	Mass Nouns (singular)		6.5.2	Mass (noncount) nouns are names of materials or substances that are subject to the question "How much?" but not "How many?". These ***should*** be expressed in the singular.

Topic			Section	Description
	Other Types of Singular Nouns		6.5.3	
		Abstract Concepts (singular)	6.5.3.1	The names of abstract concepts ***should*** be expressed in the singular.
		Unique Entities (singular)	6.5.3.2	The names of unique entities, whether concrete or abstract, ***should*** be expressed in the singular.
	Coexistence of Singular and Plural Forms		6.5.4	Where the singular and plural forms of a term represent different concepts, separate terms for each ***should*** be entered in the controlled vocabulary as appropriate. The distinction ***should*** be indicated by a qualifier.
Selecting the Preferred Form			**6.6**	The authority for the form selected ***should*** be recorded in the term record.
	Usage		6.6.1 6.6.1.1 6.6.1.2 6.6.1.3	Terms ***should*** reflect the usage of people familiar with the domain of the controlled vocabulary. Consider: • Literary warrant • Organizational warrant • User warrant
	Spelling		6.6.2	The most widely accepted spelling of words, based on warrant, ***should*** be adopted. If variant spellings exist and are commonly recognized, each ***should*** be entered in the controlled vocabulary, and a cross-reference ***should*** be made from the non-preferred to the preferred form. Spelling ***should*** follow the practice of well established dictionaries or glossaries. If a choice between spellings is made for dialectal reasons, the choice ***should*** be adhered to consistently throughout the controlled vocabulary.
	Abbreviations, Initialisms, and Acronyms		6.6.3	
		Preference for Abbreviations	6.6.3.1	Abbreviations and acronyms ***should*** be selected as terms only when they have become so well established that the full form of the term or proper name is rarely used. Cross-references ***should*** be made from the full forms.
		Preference for Full Form	6.6.3.2	The full form of terms ***should*** be selected when the abbreviated form is not widely used and generally understood. Cross references ***should*** be made from the abbreviated forms. When an acronym or abbreviation can stand for more than one word or phrase; the full form of the term ***should*** be selected in preference to the abbreviated form.

Topic			Section	Description
	Neologisms, Slang, and Jargon		6.6.4	When no widely accepted alternative exists, the neologism, slang, or jargon term ***should*** be accepted as a term.
	Popular and Scientific Terms		6.6.5	If a popular and a scientific name refer to the same concept, the form most likely to be sought by the users of the controlled vocabulary ***should*** be chosen as the term, and a cross-reference provided from the non-preferred term.
	Loanwords (words borrowed from another language)		6.6.6	If loanwords are well established, they ***should*** be admitted into the controlled vocabulary. Diacritics ***should*** be included if required by the orthographic authorities for the borrowing language.
	Proper Names		6.6.7	Proper names of persons, institutions, organizations, places, and titles ***may*** be: a) controlled by inclusion in a controlled vocabulary of subject (topical) terms; b) controlled through a separate name authority file; or c) left uncontrolled. In options b) or c), proper names ***may*** be assigned to content objects as identifiers, differentiated from topical terms. Different methods ***may*** be adopted for controlling various types of names. When proper names are included in a controlled vocabulary, the form of the name ***should*** be selected in accordance with a recognized code of cataloging practice.
		Place Names	6.6.7.1	The form most familiar to the users of the controlled vocabulary ***should*** be designated as the entry term, and cross-references ***should*** be provided from the variants. Other things being equal, preference ***should*** be given to the official rather than the popular name. The short form of the official name ***should*** be preferred. Standard authorities ***should*** be consulted for the official forms.
Capitalization and Non-alphabetic Characters			**6.7**	
	Capitalization		6.7.1	It is recommended that predominantly lowercase characters be used for terms in controlled vocabularies. Capitals ***should*** be used only for the initial letter(s) of proper names, trade names, and for those components of taxonomic names, such as genus, which are conventionally capitalized. Capitals ***should*** be used for all the letters of initialisms or where featured in unusual positions in product or corporate names.

Topic			Section	Description
	Non-alphabetic Characters		6.7.2	To simplify filing and searching, the use of symbols and punctuation marks in terms and entry terms ***should*** be minimized. Established orthographic authorities ***should*** be used to determine when such characters are essential.
		Parentheses	6.7.2.1	To eliminate problems in filing and searching, parentheses ***should*** be avoided in controlled vocabularies whenever possible. Parentheses ***should*** be used only to enclose qualifiers and trademark indicators, or when they constitute part of a term.
		Hyphens	6.7.2.2	To eliminate problems in filing and searching, hyphens generally ***should not*** be used in controlled vocabularies. Hyphens ***should*** be retained in topical terms when dropping them would lead to ambiguity. Hyphens ***should*** be retained where they occur as part of abbreviations, trademarks, chemical names, or proper nouns.
		Apostrophes	6.7.2.3	To eliminate problems in filing and searching, apostrophes generally ***should not*** be used in controlled vocabularies. Apostrophes indicating the possessive case, whether in the singular or plural, ***should*** be retained. Apostrophes that are part of proper names ***must*** be retained.
		Diacritical Marks	6.7.2.4	Diacritical marks ***may*** be used if they are required for proper names or by the accepted standards of a discipline.
		Other Symbols and Punctuation Marks	6.7.2.5	Symbols, such as the ampersand, and punctuation marks ***should not*** be used except in trademarks and proper names because they create filing and searching problems.
	Romanization		6.7.3	Commonly accepted spellings for terms or proper names from languages written in non-roman scripts, as found in authoritative reference sources, ***should*** be used in preference to systematic romanization

Topic		Section	Description
Compound Terms		**7**	
Purpose of Guidelines on Compound Terms.		**7.2**	As a general principle of vocabulary control, terms ***should*** represent single concepts, expressed by a single word or by a multiword term unless it is a precoordinated term
	Precoordinated Terms	7.2.1	A precoordinated term combines two or more words to name a specific concept more precisely.
	Postcoordinate and Precoordinate Retrieval	7.2.2	Terms can be combined at the time of search and retrieval using Boolean or other operators to express complex concepts. This is called postcoordinate retrieval. Terms ***may*** also be combined to represent more complex concepts in the controlled vocabulary itself. In this instance, terms ***may*** be constructed that are as specific as needed to describe the concepts in a particular domain. Individual terms then ***may*** be precoordinated into semantically linked, heading subheading combinations. Rules ***may*** be developed for combining terms to form precoordinate indexing terms.
Factors to be Considered When Establishing Compound Terms		**7.3**	The factors listed below ***should*** be used to decide which multiword terms ***should*** be split into separate terms and which ***should*** be retained in compound form. • Literary warrant. • Regulating the number of terms in the controlled vocabulary. • Print versus information retrieval systems. • Avoiding false hits in retrieval. • The nature of terminology in a given field.
Elements of Compound Terms		**7.4**	Elements of a compound term have one of two possible components: • The focus, the noun component that identifies the broader class of things or events to which the term as a whole refers. • The modifier, the part of a compound term that refers to a characteristic, or logical difference, which narrows the denotation of the focus by specifying a subclass of the broader concept represented by the focus.

Topic	Section	Description
Criteria for Establishing Compound Terms	**7.5**	Compound or multiword terms ***should*** be established as terms in the following circumstances: • Splitting the parts would lead to ambiguity or a loss of meaning. • One component of the term is not relevant to the scope of the controlled vocabulary or is too vague to exist as an independent term. • The modifier in the term has lost its original meaning, so the meaning of the compound term as a whole is not the sum of the meaning of its parts. • The modifier suggests a resemblance, as in a metaphor, to an unrelated thing or event. • The term contains an adjective that does not define a subclass of the focus, and is not actually a member of that class at all. • The term is a proper name, or includes proper nouns or proper adjectives. • The compound term has become so familiar in common use, or in the field covered by the controlled vocabulary, that it is considered for practical purposes to represent a single concept.
Criteria for splitting compound terms	**7.6**	A compound term ***should*** be split when: • its focus refers to a property or part, and its modifier represents the whole or possessor of that property or part. • it consists of a term representing a transitive action modified by a term for the object on which the action is performed. • it consists of a term for an intransitive action modified by a term for the performer (agent) of the action.
Node Labels	**7.7**	A compound term ***should not*** be created solely for the reason that it forms a logical level in a hierarchy and would serve to group a set of narrower terms. For this purpose node labels ***should*** be used instead. Phrases ***may*** be used as node labels if the term forms a logical level in the hierarchy.
Order of words in Compound Terms	**7.8**	Noun phrases ***should*** be entered in a controlled vocabulary in natural language order, i.e., direct order.
Cross references from inverted form	7.8.1	The inverted form of a noun phrase ***may*** be included as a cross-reference to the preferred term in natural language order.

Topic	Section	Description
Relationships	**8**	
Semantic linking	**8.1**	The relationships among terms in a controlled vocabulary are indicated by semantic linking. Semantic linking encompasses various techniques and conventions for indicating the relationships among terms.
Indicating Relationships Among Terms	8.1.1	A basic property of relationships in controlled vocabularies is that they are reciprocal; that is, each relationship indicated between **Term A** and **Term B *must*** have a corresponding relationship from **Term B** to **Term A**. This rule ***must be*** observed for all types of relationships.
Equivalence Relationships	**8.2**	When the same concept can be expressed by two or more terms, one of these is selected as the preferred term. A cross-reference to the preferred term ***should*** be made from any "equivalent" entry term. The equivalence relationship is expressed by the following conventions: • U or USE, which leads from a non-preferred (entry) term to the preferred term, and • UF or USED FOR, the reciprocal relationship, which leads from the entry terms to the preferred term.
Types of Equivalence Relationships	8.2.1 8.2.2 8.2.3 8.2.4 8.2.5	• Synonyms • Lexical variants • Near-synonyms • Generic posting • Cross reference to elements of compound terms
Hierarchical Relationships	**8.3**	Hierarchical relationships are based on degrees or levels of superordination and subordination, where the superordinate term represents a class or a whole, and subordinate terms refer to its members or parts. Reciprocity ***may*** be expressed by the following relationship indicators: • BT (Broader Term), a label for the superordinate (parent) term • NT (Narrower Term), a label for the subordinate (child) term
Generic Relationships (also called IsA)	8.3.1	The generic nature of a relationship ***may*** be identified by the BT/NT coding or, if more refined coding is desired for the various types of hierarchical relationship, by the following abbreviations: • BTG = Broader term (generic) • NTG = Narrower term (generic)

Topic		Section	Description
	Instance Relationships (also called IsA)	8.3.2	The hierarchical instance relationship ***may*** be indicated specifically by the following abbreviations: • BTI = Broader term (instance) • NTI = Narrower term (instance)
	Whole-part Relationships	8.3.3	The hierarchical whole-part relationship ***may*** be indicated specifically by the following abbreviations: • BTP = Broader term (partitive) • NTP = Narrower term (partitive)
	Polyhierarchical Relationships	8.3.4	Some concepts belong, on logical grounds, to more than one category. They are then said to possess polyhierarchical relationships.
	Node labels	8.3.5	Node labels ***may*** be inserted into the display to show the principles of division among a set of sibling terms (terms that share a broader term). Node labels are not terms, and ***must not*** be used as indexing terms.
Associative Relationships		**8.4**	This relationship covers associations between terms that are neither equivalent nor hierarchical; yet the terms are semantically or conceptually associated to such an extent that the link between them ***should*** be made explicit in the controlled vocabulary, on the grounds that it ***may*** suggest additional terms for use in indexing or retrieval. The most common associative relationship used in thesauri is symmetrical and is generally indicated by the abbreviation RT (related term), but more semantically refined associations may also be developed to capture both symmetric and asymmetric associations.
	Relationships Between Terms Belonging to the Same Hierarchy	8.4.1	Related Term (RT) references are required for sibling terms with overlapping meanings, such as **ships** and **boats**, where each of the terms can be precisely defined (so they do not form an equivalence set), yet they are sometimes used loosely and almost interchangeably. The user interested in one ***should*** be reminded of the other. Uses include: • Overlapping sibling terms • Mutually exclusive sibling terms • Derivational relationships • Terms in different hierarchies

Topic		Section	Description
	Relationships Between Terms Belonging to Different Hierarchies	8.4.2	It is possible to establish many grounds for associating terms belonging to different hierarchies. Related term references are often made between etymologically related terms, i.e., that contain the same root, but which do not represent the same kind of thing. Uses include: • Process / Agent • Process / Counteragent • Action / Property • Action / Product • Action / Target • Cause / Effect • Concept or Object / Property • Concept or Object / Origins • Concept or Object / Units or Mechanisms of Measurement • Raw Material / Product • Discipline or Field / Object or Practitioner • Phrases where the noun is not a true BT • Antonyms
	Node Labels	8.4.3	In order to bring closely related concepts together in the alphabetical array under a given term, related terms ***may*** be divided into categories that do not form part of a logical hierarchy. These related terms ***should*** then be identified by a node label.
	Specifying Types of Related Term Relationships	8.4.4	In certain controlled vocabularies, it ***may*** be desirable to refine related term references in order to make the nature of the relationships explicit. Codes for such relationship indicators and their reciprocals ***may*** be developed locally. These local codes ***should*** be clearly explained.
Displays		**9**	
General considerations		**9.1**	
	Elements to Address	9.1.1	The vocabulary developer ***should*** address the following elements of the vocabulary display: • Presentation • Type of display • Format • Documentation

Topic		Section	Description
	User Categories	9.1.2	The design of displays for controlled vocabularies ***should*** take into account the needs of each anticipated class of user: • Controlled vocabulary maintenance personnel • Indexers and expert searchers • End users Controlled vocabulary developers ***may*** want to produce different versions of the vocabulary: • a basic list of terms, references, and relationships designed for the end user or occasional searcher, and • a more complete version designed for the indexer and the expert searcher, which may include detailed scope notes, indexing instructions, information on term history, and postings data.
Presentation		**9.2**	
	Equivalence Relationships	9.2.1	USE references from non-preferred terms ***should*** be incorporated into the main listing of a controlled vocabulary rather than being relegated to an auxiliary "access vocabulary" or separate list of entry terms. The permuted display ***should not*** be used as a substitute for the inclusion of useful inversions as USE references. USE references ***should*** be created from the inverted form of a term to the preferred term in direct order. The type of hierarchical relationship display chosen affects decisions on the necessity of USE references.
	Hierarchical and Associative Relationships	9.2.2	When hierarchical and associative relationships have been established for a vocabulary, each term display ***should*** include some or all of the following information about the term, preferably in this order: • Scope notes • USE references • USED FOR (UF) references • Broader Term (BT) references to terms that are one level broader • Narrower Term (NT) references to terms that are one level narrower • Related Term (RT) references to terms that have an associative relationship with the term

Topic	Section	Description
Indentation	9.2.3	Indentation ***should*** be used as an additional visual cue for hierarchical and associative relationship layouts.
Typography	9.2.4	Terms, non-preferred terms, relationship indicators, and textual notes ***should*** be typographically distinguished. Suggested typographic specifications are: • lightface or italics for non-preferred terms and notes, • all capitals for relationship indicators such as USE, and • boldface for terms.
Capitals And Lower Case Letters	9.2.5	• The controlled vocabulary ***should*** serve as an orthographic authority in addition to noting preferred terminology. A combination of capitals and lowercase letters ***should*** therefore be used in controlled vocabulary terms as dictated by usage. See section 6.7.1 for specific guidelines.
Filing and sorting	9.2.6	In electronic systems, sorting rules and handling of non-alphabetic characters ***may*** be pre-established by the software. Some systems provide options for developers to select the sorting rules. Where such options exist, the recommendations in this Standard ***should*** be selected.
Alphabetic Characters	9.2.6.1	Terms consisting of letters ***should*** be filed word-by-word rather than letter-by-letter.
Numerals	9.2.6.2	Computer systems ***may*** differ in the way numbers are filed. Some systems file numerals at the beginning of the alphabet and some at the end. Either position is acceptable, provided that the filing position of numerals is explained in the introduction to the controlled vocabulary. Special sorting instructions are usually required for Roman numerals or if numbers are to be ignored in filing.
Non-Alphanumeric Characters (Punctuation Marks)	9.2.6.3	Punctuation marks ***should*** generally be ignored in filing entry terms and terms. Commas in inverted USE references ***should*** be ignored when filing or sorting. Parentheses within terms ***should*** be ignored in filing. Parentheses around qualifiers ***should*** be treated as a special character in filing.
Relation of Character Set to Search Commands	9.2.6.4	Non-alphabetic characters used in a machine-readable controlled vocabularies ***should not*** conflict with special characters used in search commands.

Topic	Section	Description
Types of Displays	**9.3**	
Alphabetical Displays	9.3.1	An alphabetical listing is the most basic type of vocabulary display. It ***should*** contain both preferred terms with USED FOR references and entry terms with USE references. The flat format is the most commonly used controlled vocabulary display format. It consists of all the terms arranged in alphabetical order, including their term details, and one level of BT/NT hierarchy.
Permuted Displays	9.3.2	A permuted or rotated display of the terms lists each term multiple times in the alphabetic sequence of the controlled vocabulary for each of the words in the term. The two types of permuted displays are Key Word In Context (KWIC) and Key Word Out of Context (KWOC)
Term Detail Displays	9.3.3	The complete details for each term ***should*** be presented in at least one of the displays offered.
Hierarchical Displays	9.3.4	Taxonomies, thesauri, and any vocabularies with established relationships between terms ***should*** include a hierarchical display that illustrates the relationships.
Multilevel hierarchical display	9.3.4.1	In a multilevel hierarchical display, all levels of the broader and narrower terms related to a given term are immediately visible. The sibling terms (the related terms at the same level) of the given term are only displayed by viewing the broader term.
Multilevel Broader and Narrower Terms	9.3.4.1.1	This type of multilevel display employs special notation, such as BT1, BT2 (one level broader, two levels broader) and NT1, NT2 (one level narrower, two levels narrower), to show the full hierarchy for each term.
Generic Structure	9.3.4.1.2	Multiple levels of hierarchy ***may*** be indicated without broader term / narrower term (BT/NT) notation by using the abbreviation GS (generic structure) with indentation and punctuation marks, such as periods, as cues to the levels of hierarchy
Tree Structure	9.3.4.2	In a tree structure, each term is assigned a classification notation or line number. This leads the user from the alphabetic display to the full hierarchical display, which is placed in a separate sequence.
Top Term Structure	9.3.4.3	In a top term structure, the alphabetic display includes all the relationships found in the flat format, with the addition of a relationship indicator for the top term (TT) of the hierarchy.

Topic			Section	Description
		Two-Way Hierarchical Structure	9.3.4.4	A two-way hierarchical display option is generally appended to a flat format controlled vocabulary. Each term is an access point, and all levels of broader and narrower terms are displayed, generally without notation and with indentation as a cue to hierarchy.
		Broad Categories	9.3.4.5	The alphabetic arrays of some controlled vocabularies include numbers that identify the broad category to which each term belongs. Such controlled vocabularies generally feature a separate section that displays the terms for each numbered category or subcategory in an alphabetic sequence, undifferentiated in terms of hierarchical levels.
		Faceted Display	9.3.4.6	Some controlled vocabularies provide a display of the terms organized according to the broad categories or facets to which the term belongs. Facets ***may*** have a hierarchical arrangement as well so that narrower facets are arranged within broader categories.
	Graphic Displays		9.3.5	Graphic displays can communicate relationships among concepts more effectively than linear displays to some users. Graphic displays are logically equivalent to a tree structure, but usually do not have a notation.
Display Formats			**9.4**	Controlled vocabularies ***may*** be produced in multiple formats including: • A traditional print format • An electronic file, usually PDF, of the print document. • An online, interactive version that is searchable and browsable. This ***may*** be part of an associated information retrieval system or a Web browser application. • A stand-alone software application that can be installed on the end user's personal computer.
	Print Format		9.4.1	
		Minimizing Double Lookups	9.4.1.1	A major principle in the design of a printed controlled vocabulary is the minimization of double lookups, i.e., the need to consult more than one sequence of terms to find the desired entries.

Topic			Section	Description
		Juxtaposition of Terms	9.4.1.2	Juxtaposition of terms plays a role in the creation of USE references for a printed controlled vocabulary An entry term that would immediately precede or follow the term to which it leads ***may*** be suppressed. Some singular and plural forms of terms can be widely separated in a printed alphabetic list, however, and a USE reference from one to the other would then be warranted.
		Running Heads	9.4.1.3	Whenever a printed controlled vocabulary includes multiple sequences, e.g., an alphabetic list, a rotated list, and a hierarchical display, each page ***should*** feature a running head to identify the sequence.
	Screen Format		9.4.2	
		User Interface Design	9.4.2.1	Usability and accessibility standards ***should*** be applied rigorously to all controlled vocabulary display designs.
		Keyword Searching	9.4.2.2	Electronic controlled vocabularies ***should*** provide keyword searching of all of the terms in the vocabulary. Keyword searches ***should*** retrieve all occurrences of the term, especially in compound terms.
		Term Detail Display	9.4.2.3	Users ***should*** have the option of viewing a term's history, scope note, or definition, as well as all term relationships— equivalence, hierarchical, and associative—plus any specialized relationships created for the controlled vocabulary.
		Hierarchy Level Display	9.4.2.4	It ***should*** be possible to display the hierarchies for controlled vocabularies at various levels. For example, if the hierarchy of a full controlled vocabulary has five levels, it ***should*** be possible to display an outline of the first three levels
		Pick Lists	9.4.2.5	Pick lists are frequently used to display vocabulary that has been sub-divided into facets, simple lists of terms that are appropriate for various metadata elements such as: • Formats (e.g. full text, PDF, CD-ROM, print) • Languages • States • Countries • Intellectual level (e.g. K-12, news, scholarly journal) Each of these lists ***may*** comprise an entire controlled vocabulary in itself or ***may*** be part of a larger vocabulary.

Topic			Section	Description
	Web Format		9.4.3	
		Path Hierarchy Display	9.4.3.1	Web displays often include a simplified hierarchy "path" on each display screen. This enables the user to place the current display in the hierarchy context and, when hyperlinked, to navigate back up the hierarchy, and even jump multiple levels.
		Web Navigation Techniques	9.4.3.2	Use of hyperlinks and other Web navigation tools can make it much easier to move around a controlled vocabulary enabling users to: • move easily around the hierarchies, navigating from level to level to find the desired level of specificity; • link to and view definitions, scope notes, and history notes only when needed; • move easily from one type of display to another; and • link to the full record for each term in the controlled vocabulary from within any display.
		Browsing	9.4.3.3	Controlled vocabularies that have been implemented for Web access ***should*** allow the user to browse through an alphabetical listing of the terms. The user ***may*** move through the listing by: • using the scroll slider bar on the right-hand edge of the window; • using the up and down navigation arrows (↑↓); • using the page up (PgUp) and page down (PgDn) keys; or • clicking on a plus sign (+) or equivalent symbol next to a term to expand or collapse hierarchy levels.
		Hyperlink Navigation	9.4.3.4	A typical Web-enabled controlled vocabulary might include links from a term to the following portions of a controlled vocabulary: • Narrower and broader terms • Related terms • Individual term records • Scope notes • History notes • Facets • Tree structures • Classification code or structure

Topic	Section	Description
Documentation	**9.5**	All controlled vocabularies ***should*** include comprehensive supporting documentation that clearly states: • the purpose of the controlled vocabulary; • the scope, i.e., the subject field(s) covered, with core and fringe areas separately identified; • the meaning of all conventions, abbreviations, and any punctuation marks used in nonstandard ways; • the rules and authorities adopted in selecting the preferred forms of terms and in establishing their relationships; • whether the controlled vocabulary complies with a national or international standard for controlled vocabulary construction; • the filing rules employed, citing an appropriate standard or guideline when used; • the total number of terms, with separate totals for terms and entry terms; • the date on which the controlled vocabulary was last updated; • a statement on the updating policy; • the contact information (e.g., name and address) of the responsible organization to which comments and suggestions can be sent; and • any special online navigation conventions or special options.
Interoperability	**10**	
Need For Interoperability	**10.1**	• Metasearching of multiple content resources using the searcher's preferred query vocabulary • Indexing of content in a domain using the controlled vocabulary from another domain • Merging of two or more databases that have been indexed using different controlled vocabularies • Merging of two or more controlled vocabularies to form a new controlled vocabulary that will encompass all the concepts and terms contained in the originals • Multiple language searching, indexing, and retrieval

Topic	Section	Description
Factors Affecting Interoperability	**10.2**	• The similarity of the content subject matter in the different domains. • Different controlled vocabularies used to index content from similar domains • The degree of specificity or granularity of the controlled vocabularies used to index different content domains or databases. • Differences in how synonyms and near-synonyms are handled. • The search methodologies expected by the databases being searched • The literary, organizational, and user warrants used in developing the vocabulary • The intended purpose of the databases or systems
Multilingual Controlled Vocabularies	**10.3**	For the purposes of this Standard, multilingual controlled vocabularies are included as a special case of interoperability. For all other issues related to multilingual vocabularies consult ISO 5964.
Searching	**10.4**	Three approaches are used to facilitate searching: • Combined controlled vocabularies – Individual controlled vocabularies are integrated to form a larger controlled vocabulary that is applicable to all of the specific domains. • Micro vocabularies – A generic controlled vocabulary structure is used as the guideline and template for creating multiple, highly specific controlled vocabularies. • Switching languages – Various mappings are employed to translate the terms contained in one controlled vocabulary into equivalent terms in another.
Indexing	**10.5**	A single controlled vocabulary ***may*** be used to index content from another domain if all the conditions described in section 10.2 are satisfied.
Merging Databases	**10.6**	Merging databases that have been indexed using different controlled vocabularies can present nearly insurmountable problems.

Topic	Section	Description
Merging Controlled Vocabularies	**10.7**	If the controlled vocabularies to be merged are very different in their specificity, a lot of work could be required to create new hierarchies and other relationships among terms to make the resulting vocabulary structurally consistent. Lexicographers are frequently needed to address issues of semantics and form, specificity, and usability, especially when the same or similar terms are used with different meanings in the vocabularies being merged or when different terms are used to represent identical content.
Achieving Interoperability	**10.8**	Considerable work, both practical and scholarly, has been done to develop methods that will enable controlled vocabularies to be used across multiple databases and systems and to enable them to be shared by indexers and searchers. All are still evolving. Further discussion of these methods can be found in Appendix D.
Storage And Maintenance Of Relationships	**10.9**	
Authority Files	10.9.1	Special fields in authority formats ***may*** be used to store information about the relationships among terms in different controlled vocabularies.
Vocabulary Mapping	10.9.2	One controlled vocabulary is selected as the master with others as subsidiaries. Two features are needed: • The vocabularies overlap with each other in terms of coverage. • The terms are ordered by means of hierarchical structures and other types of relationships that are shared by all. This enables them to be mapped against a shared classification scheme.
Semantic Network	10.9.3	A semantic network ***may*** be used to cluster terms by grouping the terms according to their relationship to some conceptual scheme that is common to all of the candidate controlled vocabularies. In establishing a semantic network, it is critical that the **Types** of concepts and **Types** of relationships be defined.
Lexical Database	10.9.4	A lexical database ***may*** be used to associate terms from multiple controlled vocabularies into clusters of related concepts. Various types of relationships ***may*** be accommodated including synonyms, antonyms, hierarchical relationships of various sorts, IsA, and HasA

Topic	Section	Description
Construction, testing, maintenance, and management systems	**11**	
Construction	**11.1**	
Avoid Duplicating Existing Vocabularies	11.1.1	Before embarking on the creation of a new controlled vocabulary, the sponsoring organization ***should*** ascertain whether an existing controlled vocabulary covers the same or an overlapping domain of knowledge .
Determine the Structure and Display Formats	11.1.2	The structure of the controlled vocabulary (e.g., flat, generic structure, hierarchical display, or graphic display) and display format (print, online, web-enabled) ***should*** be decided before terms are collected and considered. and before relationships among terms are constructed.
Construction Methods	11.1.3	
Committee Approach	11.1.3.1	Two main methods of creating controlled vocabularies by committee have been used. 1. **Top Down** – The broadest terms are identified first and then narrower terms are selected to reach the desired level of specificity. The necessary hierarchical structures and relationships are created as the work proceeds. 2. **Bottom Up** – This case frequently occurs when lists of terms have been derived from a corpus of content objects and are then to be incorporated in a controlled vocabulary. As in the case above, the necessary hierarchical structures and relationships are created as the work proceeds, but starting from the terms having the narrowest scope and moving to the more generic ones. If a new controlled vocabulary is being created, the "top down" approach is preferred. Once a controlled vocabulary is in place, the "bottom up" approach is most often used to add new terms to cover new concepts.

Topic			Section	Description
		Empirical Approach	11.1.3.2	The empirical approach has two basic methods:: 1. **The deductive method –** Terms are extracted from content objects but no attempt is made to control the vocabulary, or to determine relationships between terms, until a sufficient number of terms have been collected. All terms are then reviewed by a group of experts, preferably consisting of both information and subject specialists. They ***should*** first identify terms that represent the broadest classes, and then allocate remaining terms to these classes on the basis of their logical relationships, so that the hierarchies tend to be established on a broader-to-narrower basis. Vocabulary control ***should*** be developed at the stage where hierarchies and other relationships are established, following the principles described in sections 7 and 8. 2. **The inductive method –** New terms are selected for potential inclusion in the controlled vocabulary as they are encountered in content objects. Vocabulary control is applied from the outset. If the vocabulary being constructed is to have some sort of hierarchical arrangements, each term, as it is admitted, is designated as a member of one or more broader classes that are constructed on an ad hoc basis at an early stage. The controlled vocabulary is therefore established on a narrower-to-broader term basis. Controlled vocabulary construction is regarded from the outset as a continuous operation. Assistance from subject experts is strongly recommended; these experts ***may*** serve as members of a formal editorial board or committee.
		Combination of Methods	11.1.3.3	In practice, more than one of these approaches is likely to be employed at one stage or another during the construction of a controlled vocabulary.
		Machine Assistance	11.1.3.4	It is assumed throughout this Standard that controlled vocabulary construction calls for intellectual decisions. Machine assistance ***may*** be used for term identification tasks such as the following: • Identification of candidate terms • Registering frequency of term assignment • Recording terms from user queries

Topic		Section	Description
	Term Records	11.1.4	An individual record ***should*** be created for every term, and optionally for every entry term. Records for entry terms ***may*** include source notes as well as the date of admission into the controlled vocabulary. For terms, the record ***may*** contain any or all of the following elements: • term • source(s) consulted for terms and entry terms. • scope note • USED FOR references • nondisplayable variations • broader terms • narrower terms • related terms • locally established relationships • category or classification number • history note
	Term Verification	11.1.5	Before a term is admitted into a controlled vocabulary, it ***should*** be validated according to the rules in sections 6 and 7 concerning scope, form, and choice of terms. The compiler ***should*** also review relationships between each new term and other terms in the hierarchy to which it is assigned. Authorities within the subject domain ***should*** be checked before candidate terms are accepted for inclusion.
	Candidate Terms	11.1.6	Candidate terms (also called provisional terms) are proposed terms that have not gone through all acceptance procedures. These terms ***should*** be marked by a special symbol or phrase in the term record. As soon as a candidate term is approved as a term, the symbol or phrase ***should*** be deleted.
	Levels of Specificity	11.1.7	The addition of highly specific terms is usually restricted to the core area of the subject field covered by a controlled vocabulary because the proliferation of such terms in fringe areas is likely to lead to a controlled vocabulary that is difficult to manage.
	Unassigned Terms	11.1.8	When hierarchies are being established in a controlled vocabulary, terms that have not yet been used in indexing are frequently admitted into the controlled vocabulary on the grounds that they are necessary to complete a hierarchy (for example, as broader terms), and that they have potential value as indexing terms in their own right.

Topic	Section	Description
Testing And Evaluation	**11.2**	The two major reasons for testing and evaluating controlled vocabularies are: 1. To determine if the controlled vocabulary being used to describe content objects provides adequate search results (i.e. high relevance and recall) for most users, and 2. To determine if the controlled vocabulary matches users' expectations of the terms contained therein. For example, if the entry vocabulary is not sufficiently rich, users ***may*** not discover their desired terms.
Testing Methods	11.2.1	The following methods ***should*** be considered when measuring the quality or effectiveness of a controlled vocabulary: • Heuristic Evaluation • Affinity Modeling • Usability Testing
Evaluation Criteria	11.2.2	In evaluating controlled vocabularies, begin by conducting a summary evaluation of the vocabulary's adequacy for the stated purpose.
Maintenance	**11.3**	
Updating the Vocabulary	11.3.1	Policies and procedures ***should*** be established for periodic review of terminology, establishment of new terms, and replacement of obsolete terms. Controlled vocabulary editors ***should*** update the controlled vocabulary at intervals that will be determined by the frequency and volume of changes made, and by the method of distribution.
Addition Of Terms	11.3.1.1	Whenever an appropriate term or combination of terms cannot be found in a controlled vocabulary, the indexer or searcher ***should*** nominate a new term as a candidate term. The editor ***may*** want to use a print or electronic form for term nominations to ensure that all needed review information is submitted.
Modification Of Existing Terms	11.3.1.2	Indexers and searchers ***should*** be able to propose modifications to existing terms or their relationships, explaining the rationale and supplying supporting documentation for the proposed changes. Such proposals ***may*** be communicated electronically or via printed forms.
Deletion Of Terms	11.3.1.3	Overused terms and terms that have been assigned very infrequently in indexing ***should*** be considered candidates for modification or deletion, as both kinds of term are generally ineffective in retrieval.
Vocabulary Updates and Database Records	11.3.2	

Topic			Section	Description
		Modified Or Deleted Terms	11.3.2.1	Whenever a term is modified or deleted from a controlled vocabulary, the impact of the change on the ability to search previously indexed database records ***must*** be considered, unless the modified or deleted term has never been assigned. • A deleted or modified term ***may*** be retained in the controlled vocabulary for retrieval purposes only. If it is retained, it ***must*** be marked, e.g., "for retrieval purposes only," and the date of its change in status ***must*** be recorded in the history note and displayed to users. • One-for-one term changes, where one term simply replaces another, can be implemented in the database automatically if resources permit. A history note ***should*** be made in the term record to indicate that the term replacements were made. • More complicated term changes ***may*** require human intervention, e.g., where two or more specific terms are added to replace a more general term or become narrower terms to it.
		History of Changes	11.3.2.2	Each term ***should*** have a history note recording its date of entry. The note ***should*** also track the history of modifications, recording previous forms with reasons for the change. Where obsolete terms are retained in the controlled vocabulary but are not assigned in indexing, the date of replacement and reasons for it ***should*** be given.
	Hyperlink Maintenance		11.3.3	When a controlled vocabulary utilizes hyperlinks in screen displays, e.g. in a Web-based system, routine checking and maintenance of the hyperlinks ***should*** be done.

Topic	Section	Description
Management Systems	**11.4**	Because controlled vocabularies are generally used in the context of information storage and retrieval systems, the design or choice of a controlled vocabulary management system is necessarily dependent on and ***should*** be an integral part of (although not necessarily integrated into) the overall design of the information system of which it is a component. Developers and maintainers of controlled vocabularies ***should*** have access to systems that can deal with the special needs of a controlled vocabulary. It ***should not*** be necessary to adapt the controlled vocabulary to an inadequate system. The system ***should*** be flexible enough to allow controlled vocabulary managers to take advantage of emerging technologies.
General System Considerations	11.4.1	Systems for managing controlled vocabulary ***must*** have several basic capabilities: • Typography -- The system ***should*** allow printing or screen display in both capitals and lowercase letters, and it ***should*** be able to handle any special characters required by the subject field or language of the controlled vocabulary. • Sorting -- The sorting functionality ***should*** comply with the filing and sorting guidelines presented in section 9.2.5 • Display types -- The system ***should*** be capable of handling alphabetical and hierarchical controlled vocabulary displays, where applicable. Additional display types ***may*** be warranted depending on user needs. • Field definitions -- The system ***should*** provide for the definition of fields other than those enumerated for term records for purposes such as assigning codes from other systems to terms or providing term definitions as well as scope notes.

Topic	Section	Description
Term Records and Displays	11.4.2	The system ***should*** support the following for term records: • Term detail – At a minimum, these records ***should*** contain the data and specify the relationships recommended in this Standard. • Term length – There ***should*** be no limits on the number of characters in a term. • Preferred and entry terms – The system ***should*** be able to distinguish between the non-preferred terms that ***should*** be displayed to controlled vocabulary users and those that ***should not*** be displayed. • Term relationships – There ***should*** be no limits on the number or types of relationships that can be established for a given term. • Hierarchies -- It ***should*** be possible to edit a term's position in a hierarchy, and to simultaneously display multiple hierarchies for a term where these exist.
Cross References	11.4.3	When cross-reference links are created between terms, the system ***should***: • Check for the existence of a term chosen as a broader, narrower, related, or USE term, and check for the validity of references, so that: – no cross-references are made to nonexistent terms; – the only type of reference that leads from a non-preferred to a preferred term is a USE reference. – a term is not related to itself; – no conflicting references are made for the same term pairs – duplicate cross-references are eliminated. • Create reciprocal references for BT/NT (or BTG/NTG and BTP/NTP), RT/RT, and USE/ UF pairs. • Maintain these references reciprocally, so that when a term is modified or deleted, the references to it are changed accordingly in the records for all the terms related to it.
Term Deletion	11.4.4	When a term is deleted, the system ***should*** prompt for verification of the deletion. The system ***must*** make all relevant reciprocal deletions so that all links to and from deleted terms are also deleted . The system ***should*** inform the editor when the deletion creates an orphan term (a term without associative or hierarchical relationships).

Topic	Section	Description
Candidate Terms	11.4.5	The system ***should*** include a mechanism for the entry and designation of candidate terms, and candidate relationships, which are not admitted to the controlled vocabulary until reviewed and approved by controlled vocabulary editors.
Error Checking	11.4.6	A good controlled vocabulary maintenance system ***should*** enable error checking by identifying: • Duplicate terms – The system ***should*** be able to detect duplicate terms, whether entry terms or preferred terms. Duplicates detected by the system ***should*** be displayed to the controlled vocabulary editor, who will judge whether they are true duplicates or homographs requiring qualification. • Potential duplicates – The system ***should*** be able to identify potential duplicates with typographic variations (capitals and lowercase letters) and qualifiers, e.g., mercury (metal) and Mercury (planet). • Structural inconsistencies – The system ***should*** be able to check for discrepancies in term relationships, missing reciprocal relationships and inconsistent modification or deletion of terms. •
Searching And Browsing	11.4.7	The system ***should*** provide the following search and browse capabilities: • Search by the complete term or a truncation of the term • Search for non-preferred terms and retrieve references to preferred terms • Search without case sensitivity, i.e. a search term in capitals ***should*** match a lowercase term or entry term and vice versa • Browsing within hierarchical and alphabetical displays • View a term in the context of its relationships and its complete term record from any display

Topic	Section	Description
Reports	11.4.8	Controlled vocabulary reports ***may*** include, but are not limited to, the following: a) Term listings, including: • terms added or modified since a certain date • terms that do or do not have a particular type of note or relationship (SN, UF, BT, NT, or RT) • orphan terms, with no hierarchical or associative relationships • top terms b) Displays of term relationships, including: • all the relationships for an individual term • hierarchical displays • listings at various levels of a hierarchy • reports of special links created for a given controlled vocabulary • the entire hierarchical context for an individual term c) Statistical reports on terms and their characteristics, including: • the number of preferred terms and of all terms in the controlled vocabulary • the number of terms featuring a certain characteristic, for example, those that consist of three words or those that end in a certain string of characters, e.g., "itis" • the number of terms added, modified, deleted, or merged since a certain date • the number of BTs, NTs, UFs, RTs, SNs, or other value in the controlled vocabulary • the average number of characters per term and the size of the longest term • postings reports for controlled vocabularies linked to databases
Testing and Evaluating Systems	11.4.9	Some studies have been done which compare the features and strengths and weaknesses of controlled vocabulary software packages. Because they become out-of-date quickly, interested readers should check the Internet for the latest available information.

Appendix B
Comparison of Vocabulary Types

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

B.1. Properties of Lists, Synonym rings, Taxonomies, and Thesauri

Table 3: Controlled vocabulary properties

Property	List	Synonym Ring	Taxonomy	Thesaurus
Types of Terms				
Preferred terms	Yes	No	Yes	Yes
Entry terms	No	Yes	No	Yes
Candidate terms	No	No	No	Optional
Provisional terms	No	No	No	Optional
Deleted terms	No	No	No	Optional
Relationships	No	Yes	Yes	Yes
Equivalence		Yes	No	Yes
Hierarchy		No	Yes	Yes
Part/Whole		No	Yes	Yes
IsA		No	Yes	Yes
HasA		No	Yes	Yes
Classification		No	Optional	Optional
Related terms		No	No	Yes
Facet		No	No	Optional
Notes	No	No	Optional	Optional
Scope note			No	Optional
History note			No	Optional
Other notes			No	Optional

B.2. Advantages and Disadvantages of Lists, Synonym rings, Taxonomies, and Thesauri

B.2.1. Lists:

- Simple to implement, use, and maintain
- Provide little or no guidance for the user

B.2.2. Synonym Rings:

- Are constructed manually and are not used in indexing
- Can be useful in retrieval as they allow synonyms and near-synonyms to be treated equally in searching.

B.2.3. Taxonomies

- Good information about hierarchical relationships among terms
- Useful for both indexers and searchers who need to discover the most appropriate, specific terms for their purposes
- There is no entry vocabulary, (i.e. USE/USED FOR terms)
- Taxonomies do not indicate other types of relationships among terms

B.2.4. Thesauri

- Good information about hierarchical relationships among terms
- Good information about relationships among terms
- Entry vocabulary to help users locate the correct terms
- Thesauri are useful for both indexers and searchers who need to discover the most appropriate, specific terms for their purposes
- Thesauri are time-consuming and labor-intensive to develop and maintain

B.3. Typical applications of Lists, Synonym Rings, Taxonomies, and Thesauri

B.3.1. Lists

Lists are frequently used to display small sets of terms that are to be used for quite narrowly defined purposes such as a Web pull-down list or list of menu choices.

B.3.2. Synonym Rings

Synonym rings are frequently used behind-the-scenes to enhance retrieval, especially in an environment in which the indexing uses an uncontrolled vocabulary and/or there is no indexing as when searching full text.

B.3.3. Taxonomies

Taxonomies are often created and used in indexing applications and for Web navigation. Because of their (usually simple) hierarchical structure, they are effective at leading users to the most specific terms available in a particular domain.

B.3.4. Thesauri

Thesauri are the most typical form of controlled vocabulary developed for use in indexing and searching applications because they provide the richest structure and cross-reference environment. Thesauri can be narrow in scope and cover a limited domain or they can be broad in scope and widely applicable to many different types of content.

Appendix C
Characteristics and Uses of Controlled Vocabulary Display Options

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

C.1. Alphabetical Listings

Characteristics:

- Format is immediately obvious to the user
- Easy to locate desired terms in the listing
- A rich entry term vocabulary helps users to find the terms they need
- Closely related terms can be scattered widely throughout the listing because of differences in term format

  NOTE: This is a major reason for using the inverted form for terms, e.g., ***pharmacology, clinical.***
- Can be difficult for users to envision the complete hierarchy for a term if a flat format is used
- A separate listing and a double lookup are required to view hierarchical relationships

Typical uses:

- Pick lists
- Small sets of simple terms
- Synonym rings

C.2. Permuted Displays

Characteristics:

- Enables users to locate desired terms by looking for any one of the words in the desired term
- Very useful as an aid in finding the necessary terms in a very large controlled vocabulary

  NOTE: The keyword search capabilities available for electronic controlled vocabularies make it easy to locate needed terms without this display option.
- Requires a lot of space

  NOTE: This is less of a problem in electronic versions of controlled vocabularies than it is for printed products where the sheer size of the permuted display may require producing a second printed volume.

Typical uses:

- Large controlled vocabularies in printed format where it can be difficult for the user to locate desired terms by perusing the alphabetical listing

  NOTE: This type of display has been largely replaced by keyword searching in online versions of controlled vocabularies.

C.3. Term Detail Displays

Characteristics:

- Provides all of the information about a given term

  NOTE: This is especially useful for indexers and for controlled vocabulary maintenance personnel

- Provides exact information on the context of the term and how it should be used to describe content (useful to searchers)
- Provides other useful retrieval information such as showing broader/narrower terms, related terms, scope notes, and history notes if they have been used for the term
- Takes up a lot of space

  NOTE: Electronic controlled vocabularies can generally mask this problem by allowing users to view the term detail information only when requested.

Typical uses:

- Large controlled vocabularies that contain terms with extensive definitions, scope notes, history notes, etc., to show all of the information that is available about a term in a single place

C.4. Hierarchical Displays

Characteristics:

- Shows how a term fits into a given context
- Allows users to move up or down through a hierarchy to locate the desired level of specificity
- Allows users to explore other areas of the vocabulary if a term belongs to more than one hierarchy
- Facilitates search expansion by enabling the straightforward inclusion of narrower terms
- Can make it more difficult to find terms that are similar in format but belong to different hierarchies
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.5. Flat Format Displays

Characteristics:

- Economical of space
- Simple and straightforward display; easy to understand
- Requires multiple lookups to trace entire tree structure or hierarchy
- Entire hierarchical structure not visible at any one term

Typical uses:

- Smaller controlled vocabularies where displaying the entire hierarchy may not be needed

C.6. Multilevel Hierarchical Displays

Characteristics:

- Entire hierarchical structure visible at each term display
- Requires additional space
- Can be more difficult to understand, depending on the notation used
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.7. Multilevel Broader and Narrower Term displays

Characteristics:

- Entire hierarchical structure visible at each term display
- Requires additional space
- Can be more difficult to understand, depending on the notation used
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.8. Generic Structure Displays

Characteristics:

- Entire hierarchical structure visible at each term display
- Requires additional space
- Can be more difficult to understand, depending on the notation used
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.9. Tree Structures

Characteristics:

- No redundancy in display – economical of space
- Provides the ability to view the complete hierarchy in one place
- The complete array of classes and subclasses serves as a check on the logical consistency of the hierarchies

- Notation coding can be confusing to novice users
- Requires a second lookup
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.10. Top Term Structure Displays

Characteristics:

- Economical in space
- Does not provide guidance to the location of a term within the array under its top term
- Requires second lookup
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.11. Two-Way Hierarchical Structure Displays

Characteristics:

- Can view all of the broader / narrower terms in one display
- Used in combination with flat format, requires more space
- Can be confusing to novice users
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.12. Broad Categories Display

Characteristics:

- Groups terms into broad classes, which can make it easier to locate needed terms
- Can be easier to miss terms unless all single categories of terms are consulted
- Provides a convenient way for users to view the entire hierarchy to which a term belongs

Typical uses:

- Controlled vocabularies that have a deep structural relationship among terms

C.13. Faceted Displays

Characteristics:

- Groups terms into broad classes according to topical areas or other criteria (e.g. genre, format)
- Provides a view of the vocabulary that is complementary to any strict hierarchical arrangement
- Can be difficult for users to locate a specific term

Typical uses:

- Large controlled vocabularies covering a broad domain or discipline with complex relationships among terms

C.14. Graphic Displays

Characteristics:

- Provides information on how a given term fits into the broader context of a particular domain
- Introduces end users to the idea that terms can be used in multiple contexts (e.g., stock as in livestock, financial portfolios, cooking, and inventory control)
- Shows closely related concepts that can be described by entirely different terms (e.g., cardiac arrest and heart attack)
- Enhances the visibility of a controlled vocabulary and the domain it represents
- Especially effective for browsing by pulling together on one screen many terms / concepts that could be widely separated in a flat form display or require multiple hyperlink navigations in a hierarchical display
- Can take more full advantage of Web / online capabilities by getting away from the "page" orientation
- Relationships will reflect the latest information added to the vocabulary or information base since the displays are often built dynamically (in real-time)

  Note: See section 11.3 for a discussion of maintenance issues.

- Cannot always show the full hierarchy for a term or the term details

  Note: Consequently, graphic displays of controlled vocabularies are almost always provided as an additional view to the more standard alphabetical and hierarchical displays.

- Display of complex relationships with many levels may be too large to fit on a page or screen and still be legible
- Users familiar with flat hierarchical display can find the display initially confusing

Typical uses:

- Large vocabularies covering extensive domains
- Where there is a desire to show relationships that go beyond the traditional hierarchical relationships

Appendix D
Methods for Achieving Interoperability

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

NOTE: The information in this appendix and the figures are taken from the article by Marcia Lei Zeng and Lois Chan in the Journal of the American Society for Information Science and Technology [42]. Figures reprinted with permission of John Wiley & Sons, Inc. and the authors.

D.1. Derivation/Modeling

In this approach, a comprehensive controlled vocabulary is selected as a model for developing or deriving other controlled vocabularies that will be more or less interoperable. This approach encourages consistency in term selection, hierarchical structure, and form across the various derivative controlled vocabularies. For example, many users of the Art & Architecture Thesaurus® (AAT) use only the portions of AAT that apply to their own collections. To these core AAT terms, they often add their own local terminology.

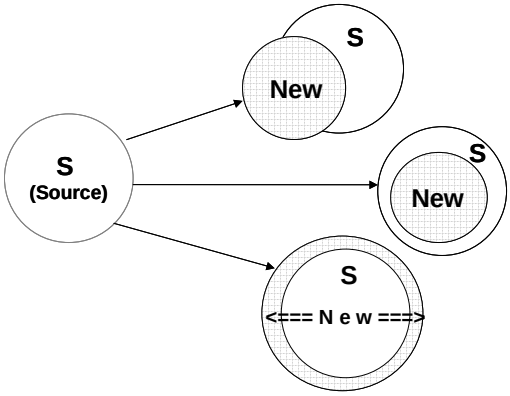


Figure 8: Derivation modeling

D.2. Translation/Adaptation

Some new controlled vocabularies have been created by translating terms from an existing controlled vocabulary or adapting terms in an existing controlled vocabulary for use in a translated version. In this case, the newly constructed vocabulary shares the general structure and guidelines of the original controlled vocabulary, thus promoting interoperability. This approach does, however, run the risk of encountering the many issues involved in developing multilingual controlled vocabularies. (See section 10.3.)

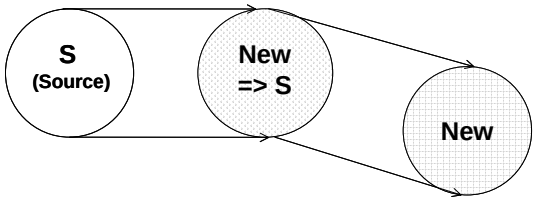


Figure 9: Translation / Adaptation

D.3. Satellite Controlled Vocabularies

The satellite approach to constructing new controlled vocabularies that will be interoperable with a source work is similar to the derivation/modeling approach covered in 11.3.2. In this case, the source controlled vocabulary functions as a superstructure that includes controlled vocabularies that are appropriate to narrower specialties. This approach enables the newly constructed controlled vocabularies to include highly specific terms (if necessary) that might overwhelm the more generic controlled vocabulary if they were included there.

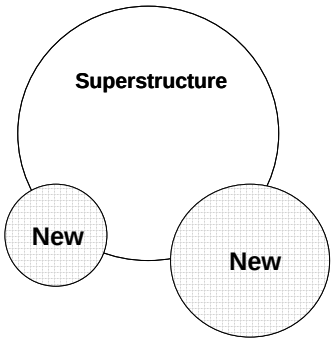


Figure 10: Satellite vocabularies

D.4. Node or Leaf Linking

Another closely related approach uses various nodes in a hierarchical structure in a source controlled vocabulary to link to more detailed controlled vocabularies that are applicable to a single node of the parent hierarchy. Once again, this approach ensures that the family of controlled vocabularies is consistent in terms of structure, term format, scope, and guidelines and allows the desired levels of specificity in the new controlled vocabularies without swamping the original controlled vocabulary with detail that most users do not need.

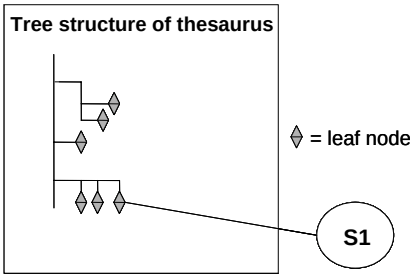


Figure 11: Node or leaf linking

D.5. Direct Mapping

Mapping involves the process of establishing equivalence and other relationships among terms in different controlled vocabularies. This approach generally requires extensive human intellectual analysis of the terms in both the source and target controlled vocabularies. Numerous problems can arise: for example, terms that are treated as synonyms in one controlled vocabulary could have a hierarchical relationship in another because that controlled vocabulary had greater specificity.

Similarly, mapping various types of related term relationships from one controlled vocabulary to another can lead to massive intellectual effort to resolve subtle differences between them.

Some work has also been done with computer-assisted mapping. This work generally depends on the use of dictionaries, semantic maps, and other devices to recognize related concepts in heterogeneous controlled vocabularies.

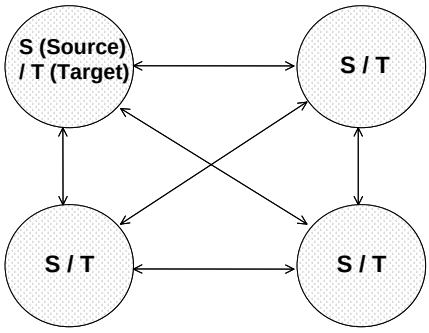


Figure 12: Direct mapping

D.6. Co-occurrence Mapping

This approach uses the co-occurrence of terms in the underlying databases to find families of related terms. The result is a controlled vocabulary consisting of sets of loosely mapped terms. The terms used for the co-occurrence mapping can be selected from metadata fields in the underlying databases, from uncontrolled keywords or other metadata that have been assigned to the content, or from words selected from the full text of the content object itself.

The loosely mapped term clusters discovered using this approach can be used in mapping between controlled vocabularies or directly for retrieval.

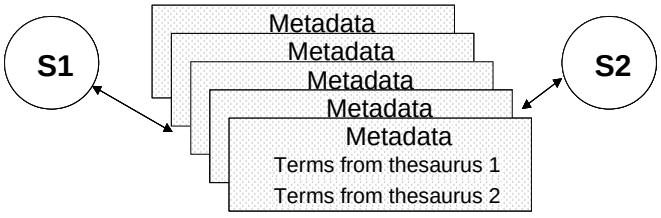


Figure 13: Co-occurrence mapping

D.7. Switching

The process of translating equivalent terms from one controlled vocabulary to another may not take place directly, but may instead use a switching language to facilitate the process. The switching system may take advantage of an existing controlled vocabulary or may be a new one. The terms contained in the switching system vocabulary serve as focal points or notation (such as a classification scheme) around which candidate terms from various controlled vocabularies are grouped. This approach enables a single, unifying hierarchical display to be used to group and display terms selected from multiple sources.

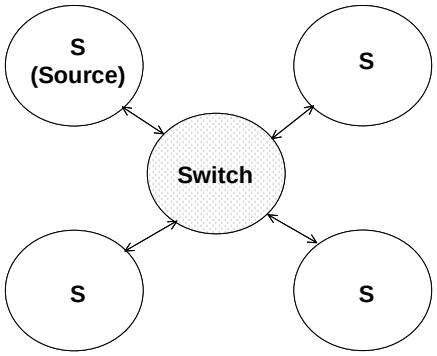


Figure 14: Switching

D.8. Linking Through a Temporary Union List

In this approach, different controlled vocabularies are linked by software mapping of terms that are not conceptual equivalents but are closely related in some way. Capturing these clusters of closely related terms is intended to enhance retrieval. The mapping is done dynamically in response to a specific user query and the results are displayed in a temporary union list. No mapping table is produced and no new controlled vocabulary is generated. Some systems also use an embedded semantic network generated from one or more underlying databases to map the user's query terminology into the words that occur in the various content objects.

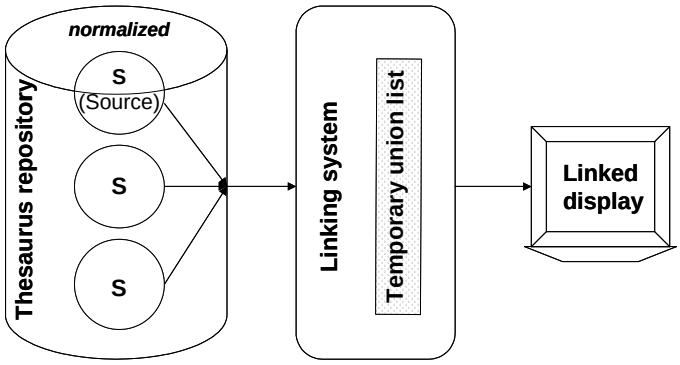


Figure 15: Temporary union list linking

D.9. Linking Through Controlled Vocabulary Servers

One additional approach to interoperability involves transmitting a query to one or more controlled vocabulary servers. Individual controlled vocabularies register with a vocabulary server service and accept queries against their vocabularies in the protocol specified by the service. User queries are submitted to the central server, which passes the queries to the registered vocabularies, collects the various results, and presents them in some consolidated form to the user.

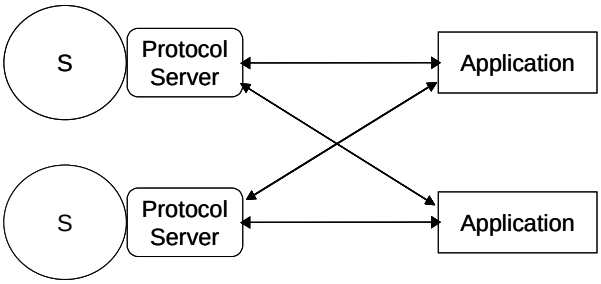


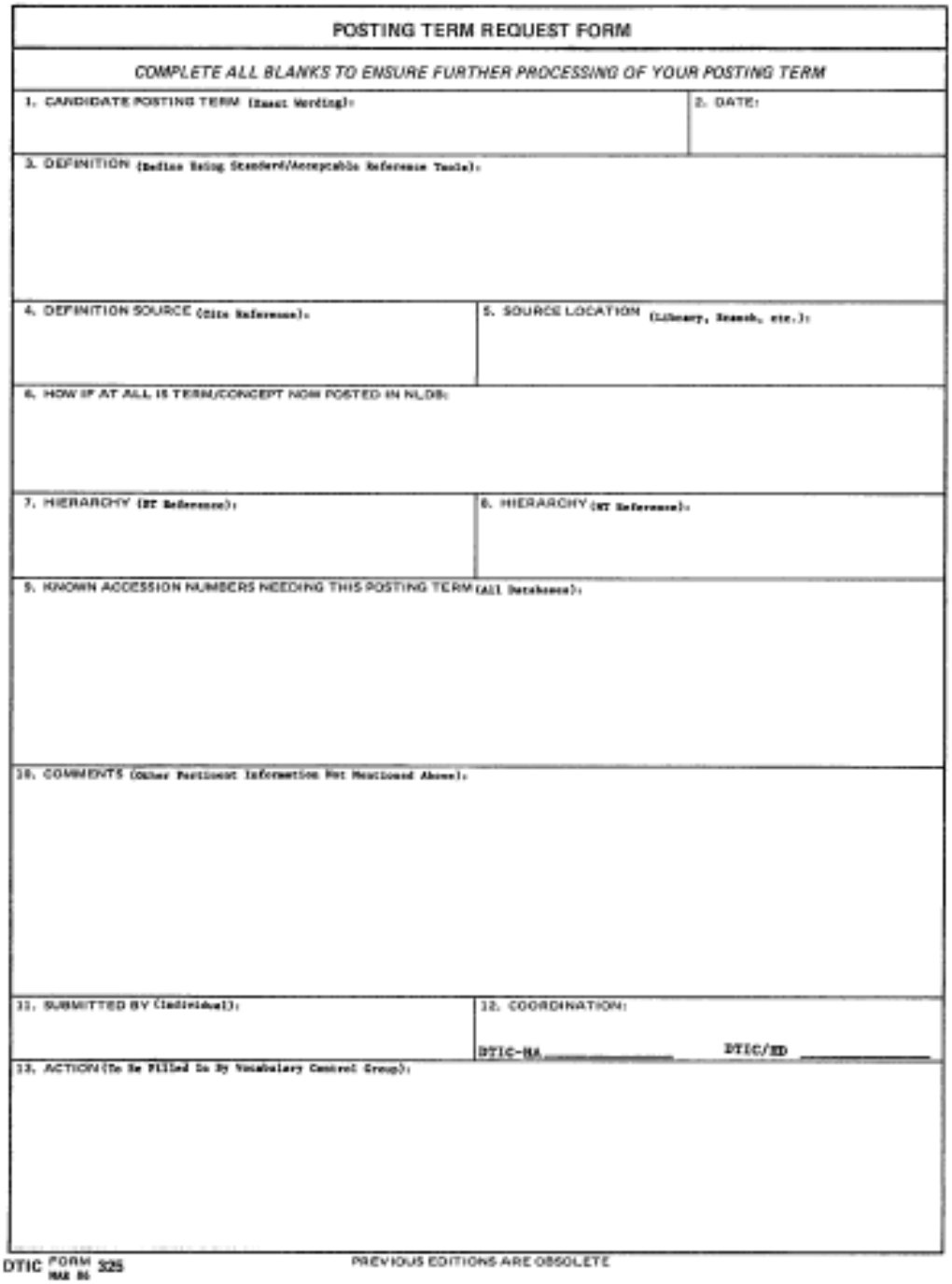
Figure 16: Server linking

A related approach is employed by some of the metasearch initiatives where the query terminology employed in a search is reformatted to conform to the protocols of a target system or database.

Appendix E
Sample Candidate Term Forms

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

Figure 17: DTIC Posting Term Request Form



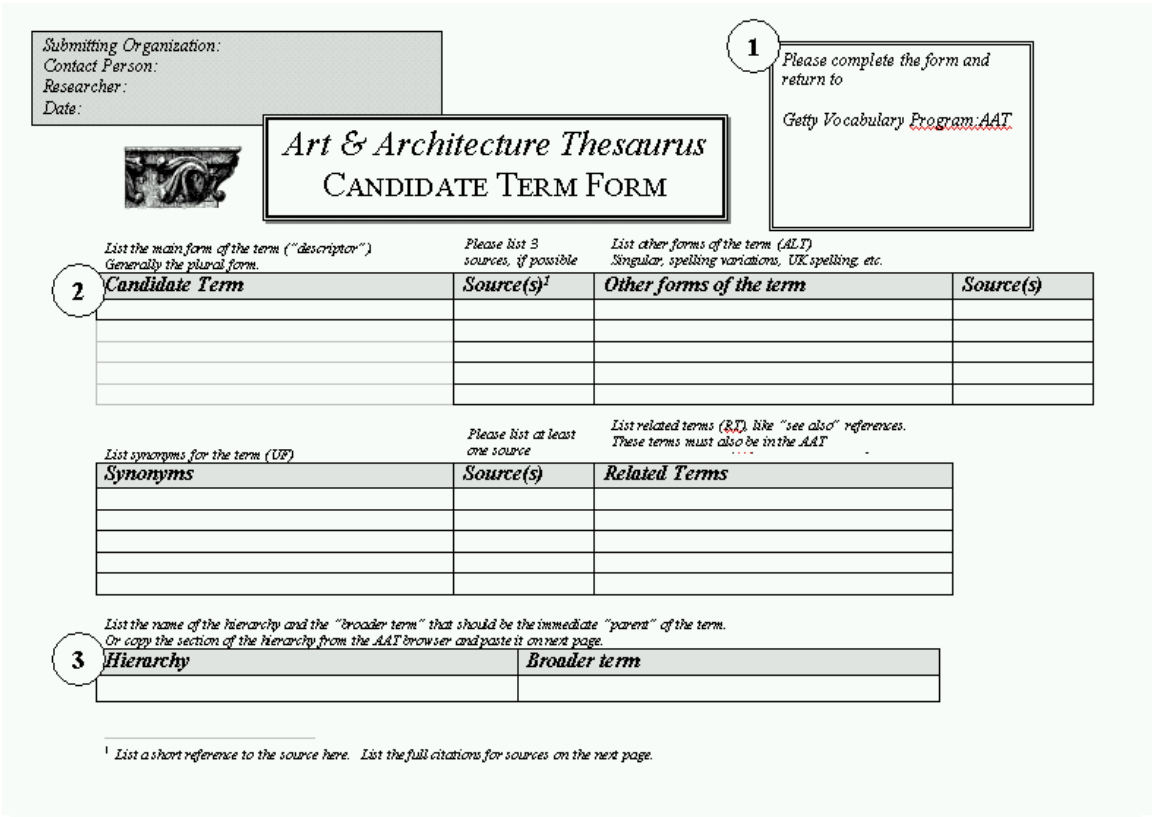
POSTING TERM REQUEST FORM

COMPLETE ALL BLANKS TO ENSURE FURTHER PROCESSING OF YOUR POSTING TERM

1. CANDIDATE POSTING TERM (Exact Wording):
2. DATE:
3. DEFINITION (Indicate Using Standard/Acceptable Reference Tools):
4. DEFINITION SOURCE (Cite Reference):
5. SOURCE LOCATION (Library, Branch, etc.):
6. HOW IF AT ALL IS TERM/CONCEPT NOW POSTED IN NLDB:
7. HIERARCHY (BT Reference):
8. HIERARCHY (NT Reference):
9. KNOWN ACCESSION NUMBERS NEEDING THIS POSTING TERM (All Databases):
10. COMMENTS (Other Pertinent Information Not Mentioned Above):
11. SUBMITTED BY (Individual):
12. COORDINATION:
DTIC-HA _____ DTIC/ED _____
13. ACTION (To Be Filled In By Vocabulary Control Group):

DTIC FORM 325 MAR 86 PREVIOUS EDITIONS ARE OBSOLETE

Figure 18: Art & Architecture Thesaurus® Candidate Term Form



Submitting Organization:
Contact Person:
Researcher:
Date:

1 Please complete the form and return to

Getty Vocabulary Program:AAT

Art & Architecture Thesaurus
CANDIDATE TERM FORM

2

List the main form of the term ("descriptor"). Generally the plural form. Candidate Term	Please list 3 sources, if possible Source(s)[1]	List other forms of the term (ALT) Singular, spelling variations, UK spelling, etc. Other forms of the term	Source(s)

List synonyms for the term (UF) Synonyms	Please list at least one source Source(s)	List related terms (RT), like "see also" references. These terms must also be in the AAT Related Terms

3 List the name of the hierarchy and the "broader term" that should be the immediate "parent" of the term. Or copy the section of the hierarchy from the AAT browser and paste it on next page.

Hierarchy	Broader term

[1] List a short reference to the source here. List the full citations for sources on the next page.

You may paste section of the hierarchy here:

4 List sources for this record: author, title, place of publication, publisher, date of publication, page numbers.

Scope Note:	Source(s):

5 List sources for this record: author, title, place of publication, publisher, date of publication, page numbers.

Sources for the Record:

Appendix F
References

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

This appendix lists full citations for the controlled vocabularies that were used as examples and to other sources that were referenced in the text of the Standard.

F.1. Controlled Vocabularies Used as Examples

[1] *AGROVOC, Multilingual agricultural thesaurus*. Rome (IT): Food and Agriculture Organization of the United Nations. Print version: Fourth edition, 1999. English version: ISBN 92-5-104378-7. Also available in French and Spanish. Available from: http://www.fao.org/agrovoc/

[2] Dickert, John, compiler and editor. *DTIC Thesaurus*. Alexandria (VA): U.S. Defense Technical Information Center; October 1996. Report No.: AD-A321 038. Available from: http://www.osti.gov/bridge/servlets/purl/486136-SyMBRY/webviewable/486136.pdf

[3] Getty Vocabulary Program, *The Art & Architecture Thesaurus (AAT)*®, version 3.0-Web. Los Angeles (CA): The J. Paul Getty Trust; c 2000. Data updated monthly. Available from: http://www.getty.edu/research/conducting_research/vocabularies/aat/

[4] *Joint Thesaurus: ETDE/INIS Joint Reference Series* Report No.: IAEA-ETDE/INIS-1 (Rev.1). Vienna (AT): International Atomic Energy Agency (IAEA); April 2004. ISBN 92-0-105604-4. Distributed by: Energy Technology Data Exchange, Oak Ridge (TN). Available from: http://www.etde.org/edb/JRS1r1_web.pdf

[5] *Legislative Indexing Vocabulary.* Washington (DC): Congressional Research Service; October 14, 1997. Available from: http://thomas.loc.gov/liv/livtoc.html

[6] *Multilingual Access to Subjects (macs)* [prototype database]. Le Tilburg, The Netherlands: InfoLab, c2003. Data supplied by Index Data (Denmark) and Tilburg University Library (Netherlands). Available from: http://macs.cenl.org

[7] *MeSH (Medical Subject Headings)*. Bethesda (MD): National Library of Medicine; 2004. Available from: http://www.nlm.nih.gov/mesh

[8] *MultiTes: Evaluate MultiTes demo information*. Miami (FL): Mulitsystems; 2004. [Note: Product information available from: http://www.multites.com/products.htm]

[9] *NASA Thesaurus: Volume 1, Hierarchical Listing with Definitions*. Hanover (MD): NASA Center for Aerospace Information; January 2004. Report No: NASA/SP-2004-7501/Vol. 1. Available from: http://www.sti.nasa.gov/98Thesaurus/vol1.pdf

[10] *NASA Thesaurus: Volume 2, Rotated Term Display.* Hanover (MD): NASA Center for Aerospace Information; January 2004. Report No: NASA/SP-2004-7501/Vol. 2. Available from: http://www.sti.nasa.gov/98Thesaurus/vol2.pdf

[11] *Photosynthesis Concept Map.* Boston (MA): Xrefer Ltd.; c2004. Available from: http://www.xrefer.com/research/

[12] *The Museum System* [software]. New York (NY): Gallery Systems. [Note: For further information see http://gallerysystems.com/]

[13] *Thesaurus for Graphic Materials I: Subject Terms (TGM I)*. Washington (DC): Library of Congress, Prints and Photographs Division; 1995. Available from: http://www.loc.gov/rr/print/tgm1/

[14] *Thesaurus of ERIC Descriptors*. Landham (MD): Educational Resource Information Center (ERIC) Processing & Reference Facility. Available from: http://eric.ed.gov/ERICWebPortal/Home.portal [Select *Thesaurus* tab.]

[15] *Thomson Gale Master Thesaurus*. Farmington Hills (MI): Thomson Gale. [Examples generated using *Synaptica* [software]. Denver (CO): Synapse, the Knowledge Link Corporation.]

[16] *UNESCO Thesaurus: A Structured List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social and Human Science, Culture, Communication and Information*. Paris: UNESCO Publishing, 1995. ISBN 92-3-003100-3. Available from: http://www.ulcc.ac.uk/unesco/

[17] *Unified Medical Language System (UMLS) Documentation, Section 3, UMLS Semantic Network. 16th Release, 2004AB.* Bethesda (MD): U.S. National Library of Medicine; updated July 20, 2004. Available from: http://www.nlm.nih.gov/research/umls/meta3.html

[18] *Unified Medical Language System (UMLS) Metathesaurus.* Version 2004AB. Bethesda (MD): U.S. National Library of Medicine. Available from: http://www.nlm.nih.gov/research/umls/

[19] *WordNet 2.0 Search*. Princeton (NJ): Princeton University, Cognitive Sciences Laboratory; August 2003. Available from: http://wordnet.princeton.edu/cgi-bin/webwn

F.2. Documents Referenced in the Standard

[20] *AACR2: Anglo-American Cataloguing Rules*. Second edition. Chicago (IL): American Library Association; 2002; revision 2004. ISBN: 0-8389-3546-X.

[21] *ALA Filing Rules*. Chicago (IL): American Library Association; January 1, 1980. ISBN 0-838932-55-X.

[22] Anderson, James D. *Guidelines for Indexes and Related Information Retrieval Devices*. Bethesda (MD): National Information Standards Organization; 1997. Report No: NISO TR02-1997. ISBN: 1-880124-36-X. Available from: http://www.niso.org/standards/resources/tr02.pdf

[23] *ANSI/NISO Z39.50-2003, Information Retrieval: Application Service Definition & Protocol Specification*. Bethesda (MD): National Information Standards Organization; 2003. ISBN: 1-880124-55-6. Available from: http://www.niso.org/standards/resources/Z39-50-2003.pdf [Note: ISO 23950:1998 is the equivalent international standard.]

[24] *ANSI/NISO Z39.85-2001, Dublin Core Metadata Element Set*. Bethesda (MD): National Information Standards Organization, 2001. ISBN: 1-880124-53-X. Available from: http://www.niso.org/standards/resources/Z39-85.pdf [Note: ISO 15836: 2003 is the identical international standard.]

[25] *Barry, Randall K., compiler and editor. ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts*. *Washington (DC): Library of Congress Network Development and MARC Standards Office;* 1997. ISBN 0-8444-0940-5. Available from: http://www.loc.gov/catdir/cpso/roman.html

[26] *BS 1749:1985, British Standard Alphabetical Arrangement Filing Order of Numerals and Symbols*. London (UK): BSI Group.

[27] *Education Index*. New York: H.W. Wilson Co., 1932- . ISSN: 0013-1385. Annual, updated monthly (except July and Aug.); updated daily on WilsonWeb and monthly on WilsonDisc.

[28] *Geographic Names Information System (GNIS)* [homepage]. Reston (VA): U.S. Geologic Survey and U.S. Board on Geographic Names. Available from: http://geonames.usgs.gov/

[29] Hodge, Gail. *Systems of knowledge organization for digital libraries: Beyond traditional authority files*. Washington, D.C.: Council on Library and Information Resources, pub91, April 2000. ISBN 1-887334-76-9. Available from: http://www.clir.org/pubs/abstract/pub91abst.html

[30] *ISO 5964:1985, Documentation—Guidelines for the establishment and development of multilingual thesauri*. Geneva (CH): International Organization for Standardization; 1985.

[31] *Library of Congress Filing Rules*. Washington (DC): Library of Congress; 1980. ISBN 0-8444-0347-4.

[32] *Library of Congress Subject Cataloging Manual: Subject Headings*. Washington (DC): Library of Congress; 2002. Cumulation: fifth edition (1996) with all updates through 2002 interfiled. Updated semiannually. ISBN 0-8444-0906-5.

[33] *Library of Congress Subject Headings (LCSH)*. 27th edition. Washington (DC): Library of Congress; 2004. Five volumes. Hardbound. Published annually. ISSN 1048-9711.

[34] *Macrothesaurus for Information Processing in the Field of Economic and Social Development*. 5th ed. Paris (FR): Organization for Economic Cooperation for Development (OECD); June 1998. ISBN: 92-64-16025-6. Also available in French.

[35] Olson, Tony; Strawn, Gary. *Mapping the MeSH and LCSH Systems*. Information Technology and Libraries. March 1997; 16(1): 5-19. [Note: Authority files for this project are available from: http://www.library.northwestern.edu/public/lcshmesh/]

[36] *PubMed* [database on the Internet]. Bethesda (MD): National Library of Medicine, National Center for Biotechnology Information; 1960-[cited October 2004]. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi

[37] *Renardus* [homepage]. Renardus Consortium. http://www.renardus.org/

[38] *Thesaurus for the Global Legal Information Network*. Washington (DC): Library of Congress. Available from: http://www.loc.gov/lexico/servlet/lexico?usr=pub-311:0&op=frames&db=GLIN

[39] Soergel, Dagobert. *Thesauri and Ontologies in Digital Libraries: Tutorial*. In: Evaluation of thesauri. Joint Conference on Digital Libraries, Portland, Oregon, July 14, 2002 (JCDL 2002), pp 107-ff. Available from: http://www.dsoergel.com/cv/B63.pdf

[40] The Unicode Consortium. *The Unicode Standard*, Version 4.0.0. Boston (MA): Addison-Wesley, 2003. ISBN 0-321-18578-1. Available online from: http://www.unicode.org/versions/Unicode4.0.0/

[41] Vizine-Goetz, Diane; Hickey, Carol; Houghton, Andrew; Thomson, Roger. *Vocabulary Mapping for Terminology Services*. Journal of Digital Information. March 2004; 4(4): article no. 272. Copyright © 2004, OCLC Online Computer Library Center, Inc. Used with permission. Available from: http://jodi.ecs.soton.ac.uk/Articles/v04/i04/Vizine-Goetz/

[42] Zeng, Marcia Lei; Chan, Lois Mai. *Trends and Issues in Establishing Interoperability Among Knowledge Organization Systems*. Journal of the American Society for Information Science and Technology. 2004; 55(5): 377-395.

Bibliography

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

Other Controlled Vocabulary Standards

ISO 2788: 1986, Documentation—Guidelines for the establishment and development of monolingual thesauri. Geneva (CH): International Organization for Standardization; 1986. [Note: UK equivalent: BS 5723: 1987]

ISO 5964: 1985, Documentation—Guidelines for the establishment and development of multilingual thesauri. Geneva (CH): International Organization for Standardization; 1985. [Note: UK equivalent: BS 6723: 1985]

I*SO 12200: 1999, Computer applications in terminology—Machine-readable terminology interchange format (MARTIF)—Negotiated interchange*. Geneva (CH): International Organization for Standardization; 1999.

ISO 12620: 1999, Computer applications in terminology—Data categories. Geneva (CH): International Organization for Standardization; 1999.

ISO 16642:2003, Computer applications in terminology—Terminological markup framework. Geneva (CH): International Organization for Standardization; 2003.

ISO/IEC 13250: 2003, Information technology—SGML applications—Topic maps. Geneva (CH): International Organization for Standardization; 2003.

BS 8723, Structured vocabularies for information retrieval—Guide. London (UK): BSI Group. [Note: In progress.]

Controlled Vocabulary Fundamentals

Buckland, Michael. *Vocabulary as a central concept in library and information science*. In: Arpanac, T., ed. Digital libraries: Interdisciplinary concepts, challenges, and opportunities. Proceedings of the third international conference on conceptions of library and information science (CoLIS3), Dubrovnik, Croatia, 23-26 May 1999. Zagreb: Lokve, pp 3-12. ISBN 953-6003-37-6. Available from: http://www.sims.berkeley.edu/~buckland/colisvoc.htm

Hodge, Gail. *Systems of knowledge organization for digital libraries: Beyond traditional authority files*. Washington, D.C.: Council on Library and Information Resources, pub91, April 2000. ISBN 1-887334-76-9. Available from: http://www.clir.org/pubs/abstract/pub91abst.html

Iyer, Hemalata. *Classificatory structures: concepts, relations and representation*. Frankfurt/Main: Indeks Verlag, 1995. ISBN: 3886725014.

Kwasnik, Barbara H. *The role of classification in knowledge representation and discovery*. Library Trends. 1999; 48(1): 22-47.

Lancaster, F.W. *Vocabulary control for information retrieval*. 2nd ed. Arlington (VA): Information Resources Press, 1986. ISBN: 0878150536. [Note: First edition, 1972, contains useful material omitted from the 2nd ed.]

Lanzi, Elisa. *Introduction to vocabularies: A guide to enhance access to art and material culture information*. Los Angeles (CA): The Getty Research Institute; first published in 1998; updated by Patricia Harpring, March 31, 2000. Available from: http://www.getty.edu/research/conducting_research/vocabularies/introvocabs/index.html

Svenonius, Elaine. *The intellectual foundation of information organization*. Cambridge, MA: MIT Press, 2000. ISBN: 0262194333.

Controlled Vocabulary Construction and Maintenance

Aitchison, Jean; Gilchrist, Alan; Bawden, David. *Thesaurus construction and use: A practical manual*. 4th edition. New York (NY): Fitzroy Dearborn Publishers; 2002. ISBN: 1579582737.

Craven, Tim. *Thesaurus construction: An introductory tutorial*. London (Ontario, CA): University of Western Ontario, Faculty of Information and Media Studies; February 19, 2002. Available from: http://instruct.uwo.ca/gplis/677/thesaur/main00.htm

Miller, Uri. *Thesaurus construction: problems and their roots*. Information Processing & Management. July 1997; 33(4): 481-493.

Morville, Peter. *How do you build a thesaurus?* Web Review. October 30, 1998. Available from: http://www.ddj.com/documents/s=2879/nam1012433845/index.html

RBMS Bibliographic Standards Committee. *Thesaurus construction and maintenance guidelines*. RBMS manual 13. Chicago (IL): Association of College and Research Libraries, Rare Books and Manuscripts Section; 1998. Available from: http://www.rbms.nd.edu/rbms_manual/thesaurus_construction.shtml

Soergel, Dagobert. *Indexing languages and thesauri: Construction and maintenance*. Los Angeles (CA): Melville, 1974. ISBN: 0471810479.

Townley, Helen M.; Gee, Ralph D. *Thesaurus-making: Grow your own word-stock*. London (UK): Andre Deutsch, 1980. Distributed by Westview Press, Boulder (CO). ISBN: 0233972250.

Warner, Amy J. *A taxonomy primer*. Ann Arbor (MI): Lexonomy, 2002. Available from: http://www.lexonomy.com/publications/aTaxonomyPrimer.html

Faceted Analysis

Canow, Joanne; Kerr, David; Whittaker, Patricia. *Faceted classification, a group perspective: History, current and future applications*. Vancouver (BC, CA): The University of British Columbia, School of Library, Archival, and Information Studies; March 27, 2002. Available from: http://www.slais.ubc.ca/courses/libr517/01-02-wt2/www/faceted/Home.html

Fast, Karl; Leise, Fred; Steckel, Mike. *All about facets and controlled vocabularies*. Boxes and arrows; December 9, 2002. Available from: http://www.boxesandarrows.com/archives/all_about_facets_controlled_vocabularies.php

Grunenberg, Louise. *Facet analysis: using faceted classification techniques to organize site content and structure*. In: ASIS&T 2002 Information Architecture Summit, Baltimore, Maryland, March 15-17, 2002. PowerPoint presentation: 32 slides. Available from: http://www.asis.org/Conferences/Summit2002/Gruenberg.ppt

Pollitt, A. Steven. *The key role of classification and indexing in view-based searching*. 63rd IFLA General Conference. Copenhagen, Denmark, August 31 - September 5, 1997. Available from: http://www.ifla.org/IV/ifla63/63polst.pdf

Spiteri, Louise. *A simplified model for facet analysis*. Canadian Journal of Information and Library Science. April-July 1998; 23: 1-30. Available from: http://aifia.org/pg/a_simplified_model_for_facet_analysis.php

Spiteri, Louise F. *The classification research group and the theory of integrative levels*. The Katharine Sharp Review (ISSN 1083-5261). Summer 1995; no. 1. Available from: http://alexia.lis.uiuc.edu/review/summer1995/spiteri.html

Spiteri, Louise F. *The use of facet analysis in information retrieval thesauri: an examination of selected guidelines for thesaurus construction*. Cataloging & Classification Quarterly. 1997; 25(1): 21-37.

Topic Maps

A practical introduction to topic maps. Oxford (UK): Techquila. Available from: http://www.techquila.com/practical_intro.html

Garshol, Lars Marius. *Metadata? Thesauri? Taxonomies? Topic maps!: Making sense of it all*. Journal of Information Science. 2004; 30(4): 378-391. Available from: http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html

TopicMaps.Org [homepage]. Available from: http://www.topicmaps.org/

Concept Mapping and Graphic Displays

Concept mapping. Fairfax (VA): George Mason University, Instructional Design & Development Immersion Program. Available from: http://chd.gse.gmu.edu/immersion/knowledgebase/strategies/cognitivism/conceptmap.htm

Kremer, Rob. *Concept mapping: Informal to formal*. In: ICCS '94 - Proceedings of the International Conference on Conceptual Structures. University of Maryland, 1994. Available from: http://pages.cpsc.ucalgary.ca/~kremer/papers/ICCS94.html

Van Schie, J.P. *Visualisation tools for knowledge management: A survey*. Observetory.com; May 2002. Available from: http://www.observetory.com/conceptmappingvs1.htm

Controlled Vocabulary-Related Metadata Element Sets and Schemas

eXchangeable Faceted Metadata Language (XFML) [homepage]. http://www.xfml.org/

Language Independent Metadata Browsing of European Resources (LIMBER) Project [homepage]. Available from: http://www.limber.rl.ac.uk/

Lee, Maria; Baillie, Stewart; Dell'Oro, Jon. *TML: a thesaural markup language*. Australia: CSIRO, 1999. Available from: http://www.ted.cmis.csiro.au/omt/tml.pdf

NKOS Registry, Reference document for data elements. Version 3. Networked Knowledge Organization Systems (NKOS). Draft, June 14, 2001. Available from: http://staff.oclc.org/~vizine/NKOS/Thesaurus_Registry_version3_rev.htm

W3C Recommendation: RDF Vocabulary Description Language 1.0: RDF Schema. WorldWideWeb Consortium, RDF Core Working Group, February 10, 2004. Available from: http://www.w3.org/TR/rdf-schema/

SKOS Core RDF Vocabulary. Semantic Web Advanced Development for Europe (SWAD-E) Project. Available from: http://www.w3.org/2004/02/skos/core/

Taylor, Mike. *Zthes: a Z39.50 Profile for thesaurus navigation*. Available from: http://www.lcweb.loc.gov/z3950/agency/profiles/zthes-04.html

Vocabulary Definition Exchange (VDEX). IMS Global Learning Consortium, Inc. Available from: http://www.imsglobal.org/vdex/

Vocabulary ML: Metacode strawman DTD. Networked Knowledge Organization Systems/Services. Draft, August 18, 2004. Available from: http://nkos.slis.kent.edu/VOCML-1.DOC

Bibliographies, Registries, and Lists of Controlled Vocabularies

Classification, indexing, metadata, and thesauri. Amherst (MA): UMass Amherst Libraries; updated: 07/08/03. Available from: http://www.library.umass.edu/catalog/class.html

Controlling your language - links to metadata vocabularies. Bristol (UK): TASI (Technical Advisory Service for Images); updated November 24, 2003. Available from: http://www.tasi.ac.uk/resources/vocabs.html

EUROBrokerS. *Thesaurus guide: analytical directory of selected vocabularies for information retrieval*. 2nd ed. Luxembourg (BE): European Communities; 1993. Report No.: EUR/92/14006. ISBN 9282649563.

Koch, Traugott. *Controlled vocabularies, thesauri and classification systems available in the WWW*. Created: 1996-12-10; last update: 2003-12-03. Available from: http://www.lub.lu.se/metadata/subject-help.html

Middleton, Michael. *Controlled vocabularies.* Available from: http://sky.fit.qut.edu.au/~middletm/cont_voc.html

Miles, A.J. *Thesaurus links*. Semantic Web Advanced Development for Europe (SWAD-E) Project; updated Oct. 14, 2003. Available from: http://www.w3c.rl.ac.uk/SWAD/thes_links.htm

Stephenson, Mary Sue. *Indexing resources on the WWW: Database indexing, controlled vocabularies and thesauri*. Vancouver (BC, CA): University of British Columbia, School of Library, Archival and Information Studies; 2002. Available from: http://www.slais.ubc.ca/resources/indexing/database1.htm

Taxonomy warehouse. Denver (CO): Synapse, the Knowledge Link Corporation. Available from: http://www.taxonomywarehouse.com/

Bibliographies, Lists, and Reviews of Controlled Vocabulary Software

Ganzmann, Jochen. *Criteria for the evaluation of thesaurus software*. International Classification. 1990; 17(3/4): 148-157. Available from: http://www.willpowerinfo.co.uk/ganzmann.htm

Maio, Paola. *Taxonomy software to the rescue*. Online Journalism Review. 2001-10-12. Available from: http://www.ojr.org/ojr/technology/1015016550.php

Milstead, Jessica. *Thesaurus management software*. Wheat Ridge (CO): American Society of Indexers; c2002. Available from: http://www.asindexing.org/site/thessoft.shtml

Milstead, Jessica L. *Specifications for thesaurus software*. Information Processing and Management. 1991; 27(2/3): 165-175.

Will, Leonard. *Software for building and editing thesauri*. Enfield (UK): Willpower Information; amended 2005-05-20. Available from: http://www.willpowerinfo.co.uk/thessoft.htm

User Interface, Web Design, and Accessibility

Johnson, Jeff. *Web bloopers: 60 common web design mistakes and how to avoid them*. San Francisco (CA): Morgan Kaufmann, 2003. ISBN 1-55860-840-0.

Krug, Steve. *Don't make me think: A common sense approach to Web usability*. Indianapolis (IN): Que, c2000. ISBN: 0789723107.

Mayhew, Deborah J. *The usability engineering lifecycle: A practitioner's handbook for user interface design*. 1st edition. San Francisco (CA): Morgan Kaufmann; 1999. ISBN: 1558605614.

Nielsen, Jakob. *Designing web usability: The practice of simplicity*. Berkeley (CA): New Riders, 1999. ISBN 1-56205-810-X.

Perlman, Gary. *HCI bibliography: Human-computer interaction resources*. Modified: 2004-10-04. Available from: http://www.hcibib.org/

Shneiderman, Ben. *Designing the user interface: Strategies for effective human-computer interaction*. 3rd ed. Reading (MA): Addison Wesley Longman, 1998. ISBN: 0201694972.

Usability Net [homepage]. Available from: http://www.usabilitynet.org/home.htm

Web Accessibility Initiative (WAI) [homepage]. World Wide Web Consortium. Available from: http://www.w3.org/WAI/

Glossary

(This appendix is not part of *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, ANSI/NISO Z39.19 – 2005. It is included for information only.)

The definitions included in this *Glossary* are derived from the previous version of this Standard, from related standards, specialized dictionaries, and industry experts.

Term	Definition
acronym	An abbreviation composed of the first letters of a compound term or phrase; e.g. Automatic Teller Machine = ATM, United Nations = UN.
associative relationship	A relationship between or among terms in a controlled vocabulary that leads from one term to other terms that are related to or associated with it; begins with the words SEE ALSO or related term (RT).
asymmetric	Lacking symmetry. In the context of controlled vocabularies, reciprocal relationships are asymmetric when the relationship indicator used between a pair of linked terms is different in one direction than it is in the reverse direction, e.g. BT / NT. See also *symmetric* and *reciprocity*.
authority file	A set of established headings and the cross-references to be made to and from each heading, often citing the authority for the preferred form or variants. Types of authority files include name authority files and subject authority files.
authorization / authorizing body	The process (authorization) or oversight group (authorizing body) responsible for selecting terms and establishing relationships for a controlled vocabulary.
blind reference	1. A term in a controlled vocabulary that has not been assigned to any content objects. These may be needed in some instances as place holders in taxonomies and other structured vocabularies. 2. A preferred term used in a SEE or USE reference where the term pointed to does not exist in the vocabulary.
bound term	A term consisting of a compound term or phrase that indicates a single concept. (The phrase was originated by Mortimer Taube in his *Studies in Coordinate Indexing,* vol. 1, 1953, p. 43.) See also *compound term.*
broader term	A term to which another term or multiple terms are subordinate in a hierarchy. In thesauri, the relationship indicator for this type of term is BT.
browsing	The process of visually scanning through organized collections of representations of content objects, controlled vocabulary terms, hierarchies, taxonomies, thesauri, etc.
candidate term	A term under consideration for admission into a controlled vocabulary because of its potential usefulness. Also known as *provisional term*.

Term	Definition
category	A grouping of terms that are semantically or statistically associated, but which do not constitute a strict hierarchy based on genus/species, parent/child, or part/whole relationships. See also *tree structure*.
classification scheme	A method of organization according to a set of pre-established principles, usually characterized by a notation system and a hierarchical structure of relationships among the entities.
compound term	A term consisting of more than one word or a phrase that represents a single concept. Compound terms ***must*** be constructed according to the guidelines of this Standard. See also *bound term* and *precoordination.*
concept	A unit of thought, formed by mentally combining some or all of the characteristics of a concrete or abstract, real or imaginary object. Concepts exist in the mind as abstract entities independent of terms used to express them.
concept map	A representation in two dimensions of the conceptual relationships among terms and the concepts they represent.
content object	An entity that contains data/information. A content object can itself be made up of content objects. For example, a journal is a content object made up of individual journal articles, which can each be a content object. The text, figures, and photographs included in a journal article can also be separate content objects. Paintings, sculpture, maps, photographs, and other non-textual objects are also content objects. The metadata for a content object can itself be a content object.
controlled vocabulary	A list of terms that have been enumerated explicitly. This list is controlled by and is available from a controlled vocabulary registration authority. All terms in a controlled vocabulary ***must*** have an unambiguous, non-redundant definition. *NOTE: This is a design goal that may not be true in practice; it depends on how strict the controlled vocabulary registration authority is regarding registration of terms into a controlled vocabulary.* At a minimum, the following two rules ***must*** be enforced: 1. If the same term is commonly used to mean different concepts, then its name is explicitly qualified to resolve this ambiguity. *NOTE: This rule does not apply to synonym rings.* 2. If multiple terms are used to mean the same thing, one of the terms is identified as the preferred term in the controlled vocabulary and the other terms are listed as synonyms or aliases.
cross-reference	1. A direction from one term to another. See *associative relationship*; *equivalence relationship*; *hierarchical relationship*.
descriptor	See *preferred term*.
difference	See *modifier*.

Term	Definition
document	Any item, printed or otherwise, that is amenable to cataloging and indexing. The term applies not only to written and printed materials in paper or microform versions (e.g., books, journals, maps, diagrams), but also to non-print media (e.g., machine-readable records, transparencies, audiotapes, videotapes) and, by extension, to three-dimensional objects or realia (e.g., museum objects and specimens). A *document* is a *content object*.
drop-down menu	See *pick list*.
entry term	The non-preferred term in a cross reference that leads to a term in a controlled vocabulary. Also known as "lead-in term." In thesauri, the relationship indicator for this type of term is U (USE); its reciprocal is UF (USED FOR). See also *preferred term.*
entry vocabulary	The set of non-preferred terms (USE references) that lead to terms in a controlled vocabulary. *NOTE: This term is used by some controlled vocabulary designers to represent the preferred as well as the non-preferred terms in a controlled vocabulary.*
eponym	A term incorporating the name of a real or mythical person, generally the discoverer of a phenomenon or inventor of an object, e.g., Herculean labor, Parkinson's disease, pasteurization.
equivalence relationship	A relationship between or among terms in a controlled vocabulary that leads to one or more terms that are to be used instead of the term from which the cross-reference is made; begins with the word SEE or USE.
facet	A grouping of concepts of the same inherent category. Examples of categories that may be used for grouping concepts into facets are: activities, disciplines, people, materials, places, etc.
facet indicator	See *node label.*
false hit	A content object retrieved whose content does not match the intent of the concepts represented by the search terms used. Previously called false drop.
federated searching	See *metasearching.*
filing rules	A set of guidelines that determine how letters and numbers, spaces, and special characters will be treated in assembling an alphabetical or other listing.
flat format	An alphabetical display format of controlled vocabularies in which only one level of broader terms and one level of narrower terms are shown for each term.
focus	In a compound term, the noun component that identifies the class of concepts to which the term as a whole refers. Also known as *head noun*. See also *modifier*.
free text	Antonym of controlled vocabulary. Natural language terms appearing in content objects, which can complement controlled vocabulary terms in an information storage and retrieval system. In free text searching, controlled vocabulary terms can also be retrieved. See also *keyword*.

Term	Definition
generic posting	1. In controlled vocabularies, the treatment of narrower terms as equivalents, e.g., **furniture** UF beds; UF sofas. See also *up-posting*. 2. In indexing and subject cataloging, the assignment of a broader term instead of a specific term, e.g., **furniture** to a content object on sofas.
generic structure	A controlled vocabulary format that indicates all hierarchical levels of terms within an alphabetic display by means of codes, indentation, and/or punctuation marks.
gloss	An explanation or definition of an obscure or ambiguous word in a text. See also *qualifier.*
graphics display	A method of representing information that uses space and distance in addition to words.
head noun	See *focus*
heading	A preferred name or term. Types of headings include proper name headings (which may be called identifiers), subject headings, and terms. A heading may include a *qualifier*.
hierarchical relationship	. A relationship between or among terms in a controlled vocabulary that depicts broader (generic) to narrower (specific) or whole-part relationships; begins with the words broader term (BT), or narrower term (NT).
hierarchy	Broader (generic) to narrower (specific) or whole-part relationships, which are generally indicated in a controlled vocabulary through codes or indentation. See also *broader term*; *narrower term.*
history note	A note in a term record in a controlled vocabulary that provides the date of entry of a term as well as the history of modifications to its scope, relationships, etc.
homograph	One of two or more words that have the same spelling, but different meanings and origins. In controlled vocabularies, homographs are generally distinguished by *qualifiers*.
HTML (Hyper Text Markup Language)	A markup language used to describe the layout and presentation of a document on the World Wide Web.
hyperlink	A method of using embedded links to connect different parts of a content object to one another.

Term	Definition
identifier	1. A proper name (or its abbreviation or acronym) of an institution, person, place, object, or process, optionally treated as a category of heading distinct from terms. Identifiers may be held in a separate file (compare authority file), and their form may be controlled (e.g., the name of an international organization having different names in various languages, only one of which is selected). 2. In some systems, a provisional term that may be upgraded to approved status, or a highly specific term that is not eligible for term status, but which is considered useful for retrieval and is assigned to one or more content objects without vocabulary control.
indexing	1. A method by which terms or subject headings from a controlled vocabulary are selected by a human or computer to represent the concepts in or attributes of a content object. The terms may or may not occur in the content object. 2. An operation intended to represent the results of the content analysis of a document by means of a controlled indexing language or by natural language. [ISO 5127/1]
indexing language	A controlled vocabulary or classification system and the rules for its application. An indexing language is used for the representation of concepts dealt with in documents [content objects] and for the retrieval of such documents [content objects] from an information storage and retrieval system. [ISO 5127/1]
indexing term	The representation of a concept in an indexing language, generally in the form of a noun or noun phrase. Terms, subject headings, and heading-subheading combinations are examples of indexing terms.
information storage and retrieval system	A set of operations and the associated equipment, software, and documentation by which content objects are indexed and the data are stored, so that selected content objects can be retrieved in response to requests employing commands that can be handled by the system.
initialism	A set of initials by which something is known in preference to the full form of its name. Example: IBM, ICBM. See also *acronym.*
interoperability	The ability of two or more systems or components to exchange information and use the exchanged information without special effort on the part of either system.
keyword	A word occurring in the natural language of a document that is considered significant for indexing and retrieval. See also *free text*.
KWIC (Key Word In Context) index.	A type of index, arranged alphabetically, in which each significant word in a string of text serves as an access point, by being graphically emphasized and surrounded by the rest of the string. The keyword is generally in a centered column and is followed on the right by the continuation of the string, which provides the context. The balance of the string, if any, is positioned to the left of the keyword.

Term	Definition
KWOC (Key Word Out Of Context) index	A type of index, arranged alphabetically, in which each significant word in a string of text serves as an access point, usually positioned in the left-hand column of a page, followed by the complete string. The keyword may therefore not be in the immediate context of the words that surround it.
lexeme	A fundamental unit of the vocabulary of a language.
lexical database	A database containing terms as well as information about the terms such as part of speech, type of term, etc.
lexicographer	A person who is knowledgeable about terms, their uses, parts of speech, etc. Lexicographers often construct controlled vocabularies.
literary warrant	Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of its frequent occurrence in the literature. See also *organizational warrant* and *user warrant*.
mapping	A set of correspondences between categories, schema element names, or controlled terms. Mappings are used for transforming data or queries from one vocabulary for use with another.
metasearching	The simultaneous searching across multiple databases, sources, platforms, and protocols. Also known as broadcast searching, cross-database searching, federated searching, or parallel searching.
microcontrolled vocabulary	A subset of a controlled vocabulary, covering a limited range of topics within the domain of the controlled vocabulary. A microcontrolled vocabulary may contain highly specialized terms that are not in the broad controlled vocabulary. Such terms should map to the hierarchical structure of the broad controlled vocabulary. A microcontrolled vocabulary is internally consistent with respect to relationships among terms.
modifier	In a compound term, one or more components that serve to narrow the extension of a focus and specify one of its subclasses. Also known as *difference*.
multilevel hierarchy	A set of hierarchical relationships among terms that has multiple levels of specificity extending from the most broadly defined terms to the most specific.
narrower term	A term that is subordinate to another term or to multiple terms in a hierarchy. In thesauri, the relationship indicator for this type of term is NT.
natural language	A language used by human beings for verbal communication. Words extracted from natural language texts for indexing purposes without vocabulary control are often called keywords.
navigation	The process of moving through a controlled vocabulary or an information space via some pre-established links or relationships. For example, navigation in a controlled vocabulary could mean moving from a broader term to one or more narrower terms using the predefined relationships.

Term	Definition
near-synonym	A term whose meaning is not exactly synonymous with that of another term, yet which may nevertheless be treated as its equivalent in a controlled vocabulary. Example: **salinity**, **saltiness**
node label	A "dummy" term, often a phrase, that is not assigned to documents when indexing, but which is inserted into the hierarchical section of some controlled vocabularies to indicate the logical basis on which a class has been divided. Node labels may also be used to group categories of related terms in the alphabetic section of a controlled vocabulary.
non-preferred term	See *entry term*. See also *preferred term*.
OPAC (Online Public Access Catalog)	A library or other catalog of content objects that is accessible online. The catalog may or may not be accessible to the public, but it is still called an OPAC.
organizational warrant	Justification for the representation of a concept in an indexing language or for the selection of a preferred term due to characteristics and context of the organization. See also *literary warrant* and *user warrant*.
orphan term	A term that has no associative or hierarchical relationship to any other term in a controlled vocabulary.
orthography	The art of writing words with the proper letters according to standard usage.
PDF (portable document format)	A file format developed by Adobe Systems that provides hardware- and software-independent viewing of a formatted document.
permuted display	A type of index where individual words of a term are rotated to bring each word of the term into alphabetical order in the term list. See also *KWIC* and *KWOC*.
pick list	A graphical user interface device that allows the user to select from a pre-set list of terms. Typically the list of terms is shown when the user clicks on a down arrow next to the entry box for the term.
polyhierarchy	A controlled vocabulary structure in which some terms belong to more than one hierarchy. For example, **rose** might be a narrower term under both **flowers** and **perennials** in a horticulture vocabulary.
polyseme	A word with multiple meanings. In spoken language, polysemes are called homonyms; in written language they are called homographs. Only the latter are relevant to controlled vocabularies designed for textual information.
postcoordination	The combining of terms at the searching stage rather than at the subject heading list construction stage or indexing stage. See also *precoordination*.
postings	The number of content objects to which a term is assigned.

Term	Definition
precision	A measure of a search system's ability to retrieve only relevant content objects. Usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the total number of content objects retrieved. A high-precision search ensures that, for the most part, the content objects retrieved will be relevant. However, a high-precision search may not retrieve all relevant content objects. See also *recall*. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down.
precoordination	The formulation of a multiword heading or the linking of a heading and subheadings to create a formally controlled, multi-element expression of a concept or object. Precoordination is often used to ensure logical sorting of related expressions. Examples of precoordinated headings: **New England—Genealogy—Handbooks, Manuals, etc.** **Searching, Bibliographic** **United States—History—Civil War, 1861-1865** See also *postcoordination*.
preferred term	One of two or more synonyms or lexical variants selected as a term for inclusion in a controlled vocabulary. See also *non-preferred term*. *NOTE: In the previous version of this Standard, a preferred term was known as a descriptor.*
provisional term	See *candidate term*.
qualifier	A defining term, used in a controlled vocabulary to distinguish homographs. A qualifier is considered part of a term, subject heading, or entry term, but is separated from it by punctuation. The qualifier is generally enclosed in parentheses. Example: **Mercury (metal)** See also *gloss*.
quasi-synonym	See *near synonym*.
recall	A measure of a search system's ability to retrieve all relevant content objects. Usually expressed as a percentage calculated by dividing the number of retrieved relevant content objects by the number of all relevant content objects in a collection. A high recall search retrieves a comprehensive set of relevant content objects from the collection. However, high recall increases the possibility that less relevant content objects will also be retrieved. See also *precision*. Recall and precision tend to be inverse ratios. When one goes up, the other usually goes down.
reciprocity	Semantic relationships in controlled vocabularies must be reciprocal, that is each relationship from one term to another must also be represented by a reciprocal relationship in the other direction. Reciprocal relationships may be symmetric, e.g. RT / RT, or asymmetric e.g. BT / NT. See also *asymmetric* and *symmetric*.
related term	A term that is associatively but not hierarchically linked to another term in a controlled vocabulary. In thesauri, the relationship indicator for this type of term is RT.

Term	Definition
relationship indicator	A word, phrase, abbreviation, or symbol used in thesauri to identify a semantic relationship between terms. Examples of relationship indicators are UF (USED FOR), and RT (related term).
romanization	The conversion of a non-roman script by means of transcription or transliteration or a combination of the two methods.
rotated listing	See *permuted display*.
running head	A page heading indicating the first and last entries that appear on that page. The heading changes on each page to reflect the changed content.
scope note	A note following a term explaining its coverage, specialized usage, or rules for assigning it.
semantic linking	A method of linking terms according to their meaning or meanings.
semantic web	A representation in two (or possibly three) dimensions of the semantic relationships between and among terms and the concepts they represent.
sibling	A term that shares the same broader term (one level higher) as other terms.
stop list	A list of words considered to be of no value for retrieval. It consists primarily of function words—articles, conjunctions, and prepositions— but may also include words that occur very frequently in the literature of a domain.
subheading	A term appended to a heading in order to modify or delimit the heading by indicating a particular aspect or relationship pertaining to it. A term with a subheading may be subject to further modification. See also *precoordination*.
subject heading	A word or phrase, or any combination of words, phrases, and modifiers used to describe the topic of a content object. Precoordination of terms for multiple and related concepts is a characteristic of subject headings that distinguishes them from controlled vocabulary terms. See also *precoordinated term* and *precoordination*.
subject heading list	An alphabetical list of subject headings with cross-references from non-preferred terms and links to related terms. These lists often include separate sequences of standardized subheadings that may be combined with all or only some subject headings. Rules for applying subheadings usually accompany such lists.
symmetric	Having symmetry. In the context of controlled vocabularies reciprocal relationships are symmetric when the relationship indicator used between a pair of linked terms is the same in one direction as it is in the reverse direction, e.g. RT / RT. See also *asymmetric* and *reciprocity*.
synonym	A word or term having exactly or very nearly the same meaning as another word or term.
synonym ring	A group of terms that are considered equivalent for the purposes of retrieval.
systematic display	See *tree structure*.

Term	Definition
taxonomy	A collection of controlled vocabulary terms organized into a hierarchical structure. Each term in a taxonomy is in one or more parent/child (broader/narrower) relationships to other terms in the taxonomy.
term	One or more words designating a concept. See also *compound term, entry term,* and *precoordinated term*.
term record	A collection of information associated with a term in a controlled vocabulary, including the history of the term, its relationships to other terms, and, optionally, authorities for the term.
thesaurus (plural: thesauruses, thesauri)	A controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators. Relationship indicators ***should*** be employed reciprocally. Its purpose is to promote consistency in the indexing of content objects, especially for postcoordinated information storage and retrieval systems, and to facilitate browsing and searching by linking entry terms with terms. Thesauri may also facilitate the retrieval of content objects in free text searching. *Note: The term "Thesaurus" is the Latin form of the Greek word thesauros, originally meaning "treasure store." In the 16th century, it began to be used as a synonym for "dictionary" (a treasure store of words), but later it fell into disuse. Peter Mark Roget resurrected the term in 1852 for the title of his dictionary of synonyms. The purpose of that work is to give the user a choice among similar terms when the one first thought of does not quite seem to fit. A hundred years later, in the early 1950s, the word "thesaurus" began to be employed again as the name for a word list, but one with the exactly opposite aim: to prescribe the use of only one term for a concept that may have synonyms. A similarity between Roget's Controlled Thesaurus and thesauri for indexing and information retrieval is that both list terms that are related hierarchically or associatively to terms, in addition to synonyms.*
top term	The broadest term in a controlled vocabulary hierarchy, sometimes indicated by the abbreviation TT.
transcription	The process of recording the phonological and/or morphological elements of a language in terms of a specific writing system.
transliteration	The process of recording the graphic symbols of one writing system in terms of the corresponding graphic symbols of another writing system.
tree structure	A controlled vocabulary display format in which the complete hierarchy of terms is shown. Each term is assigned a tree number or line number which leads from the alphabetical display to the hierarchical one. The hierarchical display is also known as a *systematic display*.
typography	The style, arrangement, appearance, or typeface used to represent information.
up-posting	The automatic assignment of broader terms in addition to the specific term by which a document is indexed. Also known as autoposting. See also *generic posting*.
user interface	The way in which a user interacts with a computer-based system.

Term	Definition
user warrant	Justification for the representation of a concept in an indexing language or for the selection of a preferred term because of frequent requests for information on the concept or free-text searches on the term by users of an information storage and retrieval system. See also *literary warrant* and *organizational warrant*.
vocabulary control	The process of organizing a list of terms (a) to indicate which of two or more synonymous terms is authorized for use; (b) to distinguish between homographs; and (c) to indicate hierarchical and associative relationships among terms in the context of a controlled vocabulary or subject heading list. See also *controlled vocabulary*.

Index

~Q~

~R~

~S~

~T~