

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This standard may be revised or withdrawn at any time.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

Copyright © 2005 by the National Information Standards Organization

All rights reserved under International and Pan-American Copyright Conventions. For noncommercial purposes only, this publication may be reproduced or transmitted in any form or by any means without prior permission in writing from the publisher, provided it is reproduced accurately, the source of the material is identified, and the NISO copyright status is acknowledged. All inquiries regarding translations into other languages or commercial reproduction or distribution should be addressed to: NISO Press, 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814.

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

Contents

Foreword	viii
1 Introduction	1
1.1 Need for Vocabulary Control	1
1.2 How Vocabulary Control is Achieved	1
2 Scope	1
2.1 Applying the Standard	2
2.2 Monolingual Controlled Vocabularies	2
2.3 Types of Terms Covered by the Standard	2
2.4 Interoperability	3
2.5 Maintenance	3
2.6 Testing and Evaluation	3
3 Referenced Standards	3
4 Definitions, Abbreviations, and Acronyms	3
4.1 Definitions	3
4.2 Abbreviations and Acronyms	10
5 Controlled Vocabularies – Purpose, Concepts, Principles, and Structure	10
5.1 Purpose	10
5.2 Concepts	11
5.2.1 Terms	11
5.2.2 Content Objects	11
5.2.3 Indexing	12
5.3 Principles	12
5.3.1 Ambiguity	13
5.3.2 Synonymy	13
5.3.3 Semantic Relationships	14
5.3.4 Facet Analysis	14
5.3.5 Using Warrant to Select Terms	16
5.3.6 Controlled Vocabulary Impact on Information Retrieval	16
5.4 Structure	16
5.4.1 List	17
5.4.2 Synonym Ring	18
5.4.3 Taxonomy	18
5.4.4 Thesaurus	18
5.5 Controlled Vocabulary Metadata and Metadata Schemas	19
6 Term Choice, Scope, and Form	20
6.1 Choice of Terms	20

6.2	Scope of Terms	20
6.2.1	Homographs	20
6.2.2	Scope Notes	22
6.2.3	History Notes	22
6.3	Term Form	23
6.3.1	Single-Word vs. Multiword Terms	23
6.3.2	Types of Concepts	23
6.3.3	Unique Entities	24
6.4	Grammatical Forms of Terms	25
6.4.1	Nouns and Noun Phrases	25
6.4.2	Adjectives	26
6.4.3	Adverbs	27
6.4.4	Initial Articles	27
6.5	Nouns	28
6.5.1	Count Nouns	28
6.5.2	Mass Nouns	29
6.5.3	Other Types of Singular Nouns	29
6.5.4	Coexistence of Singular and Plural Forms	29
6.6	Selecting the Preferred Form	30
6.6.1	Usage	30
6.6.2	Spelling	30
6.6.3	Abbreviations, Initialisms, and Acronyms	31
6.6.4	Neologisms, Slang, and Jargon	31
6.6.5	Popular and Scientific Names	32
6.6.6	Loanwords, Translations of Loanwords, and Foreign-Language Equivalents	32
6.6.7	Proper Names	33
6.7	Capitalization and Non-alphabetic Characters	34
6.7.1	Capitalization	34
6.7.2	Non-alphabetic Characters	34
6.7.3	Romanization	36
7	Compound Terms	36
7.1	General	36
7.2	Purpose of Guidelines on Compound Terms	36
7.2.1	Precoordinated Terms	37
7.2.2	Retrieval Considerations	37
7.3	Factors to be Considered When Establishing Compound Terms	37
7.4	Elements of Compound Terms	38
7.5	Criteria for Establishing Compound Terms	39
7.6	Criteria for Determining When Compound Terms Should be Split	40
7.6.1	Factors to be Considered	40
7.6.2	Hierarchical Structure	40
7.7	Node Labels	41
7.8	Order of Words in Compound Terms	41
7.8.1	Cross-references from Inversions	41
8	Relationships	42
8.1	Semantic Linking	42

8.1.1	Indicating Relationships Among Terms	43
8.2	Equivalence Relationships	43
8.2.1	Synonyms	44
8.2.2	Lexical Variants	45
8.2.3	Near-Synonyms	45
8.2.4	Generic Posting	45
8.2.5	Cross-references to Elements of Compound Terms	46
8.3	Hierarchical Relationships	46
8.3.1	Generic Relationships	47
8.3.2	Instance Relationships	48
8.3.3	Whole-Part Relationships	49
8.3.4	Polyhierarchical Relationships	49
8.3.5	Node Labels in Hierarchies	51
8.4	Associative Relationships	51
8.4.1	Relationships Between Terms Belonging to the Same Hierarchy	51
8.4.2	Relationships Between Terms Belonging to Different Hierarchies	53
8.4.3	Node Labels for Related Terms	56
8.4.4	Specifying Types of Related Term References	57
9	Displaying Controlled Vocabularies	57
9.1	General Considerations	57
9.1.1	Elements to Address	57
9.1.2	User Categories	57
9.2	Presentation	58
9.2.1	Displaying the Equivalence Relationship	58
9.2.2	Displaying Hierarchical and Associative Relationships	60
9.2.3	Indentation	61
9.2.4	Typography	62
9.2.5	Capitals and Lowercase Letters	63
9.2.6	Filing and Sorting	63
9.3	Types of Displays	64
9.3.1	Alphabetical Displays	64
9.3.2	Permuted Displays	65
9.3.3	Term Detail Displays	66
9.3.4	Hierarchical Displays	68
9.3.5	Graphic Displays	73
9.4	Display Formats – Physical Form	74
9.4.1	Print Format – Special Considerations	74
9.4.2	Screen Format – Special Considerations	75
9.4.3	Web Format – Special Considerations	79
9.5	Documentation	81
10	Interoperability	82
10.1	The Need for Interoperability	82
10.2	Factors Affecting Interoperability	83
10.3	Multilingual Controlled Vocabularies	84
10.4	Searching	85
10.5	Indexing	85

10.6 Merging Databases	85
10.7 Merging Controlled Vocabularies	86
10.8 Achieving Interoperability	86
10.9 Storage and Maintenance of Relationships among Terms in Multiple Controlled Vocabularies	86
10.9.1 Authority Records	86
10.9.2 Vocabulary Mapping	86
10.9.3 Semantic Network	87
10.9.4 Lexical Database	89
11 Construction, Testing, Maintenance, and Management Systems	90
11.1 Construction	90
11.1.1 Avoid Duplicating Existing Vocabularies	91
11.1.2 Determine the Structure and Display Formats	91
11.1.3 Construction Methods	91
11.1.4 Term Records	93
11.1.5 Term Verification	93
11.1.6 Candidate Terms	93
11.1.7 Levels of Specificity	94
11.1.8 Unassigned Terms	94
11.2 Testing and Evaluation of Controlled Vocabularies	94
11.2.1 Testing Methods	95
11.2.2 Evaluation Criteria	95
11.3 Maintenance	96
11.3.1 Updating the Vocabulary	96
11.3.2 Vocabulary Updates and Database Records	97
11.3.3 Hyperlink Maintenance	99
11.4 Management Systems	99
11.4.1 General System Considerations	99
11.4.2 Term Records and Displays	100
11.4.3 Cross-References	100
11.4.4 Term Deletion	101
11.4.5 Candidate Terms	101
11.4.6 Error Checking	101
11.4.7 Searching and Browsing	102
11.4.8 Reports	102
11.4.9 Testing and Evaluating Systems for Use with Controlled Vocabularies	102
Appendix A Summary of Standard Requirements / Recommendations	104
Appendix B Comparison of Vocabulary Types	135
Appendix C Characteristics and Uses of Controlled Vocabulary Display Options	137
Appendix D Methods for Achieving Interoperability	142
Appendix E Sample Candidate Term Forms	147
Appendix F References	149
Bibliography	152
Glossary	157
Index	168

Figures

Figure 1: Content objects	12
Figure 2: Ambiguity caused by homographs and polysemes.....	13
Figure 3: Information scatter caused by synonyms.....	13
Figure 4: Facet analysis applied to content objects and controlled vocabularies	14
Figure 5: Increasing structural complexity among controlled vocabularies.....	17
Figure 6: All and some relationship – passes test.....	47
Figure 7: All and some relationship – does not pass test.....	48
Figure 8: Derivation modeling.....	142
Figure 9: Translation / Adaptation	143
Figure 10: Satellite vocabularies	143
Figure 11: Node or leaf linking.....	143
Figure 12: Direct mapping	144
Figure 13: Co-occurrence mapping	144
Figure 14: Switching	145
Figure 15: Temporary union list linking	145
Figure 16: Server linking.....	146
Figure 17: DTIC Posting Term Request Form.....	147
Figure 18: Art & Architecture Thesaurus Candidate Term Form.....	148

Tables

Table 1: Selected semantic relationships among terms.....	42
Table 2: Interoperability needs	83
Table 3: Controlled vocabulary properties.....	135

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 Haas 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

Wherever used in this standard, all terms that are trademarks or service marks are and remain the property of their respective owners.

Permission to reprint copyrighted material was granted from: Lois Chan; Energy Technology Data Exchange and International Atomic Energy Agency; Food and Agriculture Organization of the United Nations; Gallery Systems; Getty Research Institute; John Wiley & Sons, Inc.; MACS project of the Conference of European National Librarians; Multisystems; OCLC, Inc.; Synaptica; Thomson Gale; WordNet; Xrefer; and Marcia Lei Zeng.

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. Coutts

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.

<u>Term</u>	<u>Definition</u>
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 <i>symmetric</i> and <i>reciprocity</i> .

<u>Term</u>	<u>Definition</u>
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 <i>provisional term</i> .
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 <i>tree structure</i> .
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 <i>tree structure</i> .
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 <i>bound term</i> (in the glossary) and <i>precoordination</i> .
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.

<u>Term</u>	<u>Definition</u>
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. <i>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.</i> 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. <i>NOTE: This rule does not apply to synonym rings.</i> 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 <i>associative relationship</i>; <i>equivalence relationship</i>; <i>hierarchical relationship</i>.
descriptor	See <i>preferred term</i> .
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 <i>document</i> is a <i>content object</i> .
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 <i>preferred term</i> .
entry vocabulary	The set of non-preferred terms (USE references) which lead to the preferred terms in a controlled vocabulary. <i>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.</i>
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 <i>head noun</i> . See also <i>modifier</i> .

<u>Term</u>	<u>Definition</u>
generic posting	In controlled vocabularies, the treatment of narrower terms as equivalents, e.g., furniture UF beds; UF sofas. See also <i>up-posting</i> , 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 <i>qualifier</i> .
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 <i>broader term</i> ; <i>narrower term</i> .
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 <i>qualifiers</i> .
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 <i>organizational warrant</i> and <i>user warrant</i> .

<u>Term</u>	<u>Definition</u>
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 <i>difference</i> .
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 <i>entry term</i> . See also <i>preferred term</i> .
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 <i>literary warrant</i> and <i>user warrant</i> .
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 <i>precoordination</i> .

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.

<u>Term</u>	<u>Definition</u>
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 <i>precoordination</i> .
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 <i>precoordinated term</i> and <i>precoordination</i> .
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 <i>asymmetric</i> and <i>reciprocity</i> .
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 <i>compound term</i> , <i>entry term</i> , and <i>precoordinated term</i> .
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 <i>systematic display</i> .

<u>Term</u>	<u>Definition</u>
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 <i>literary warrant</i> and <i>organizational warrant</i> .
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 <i>controlled vocabulary</i> .

4.2 Abbreviations and Acronyms

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

<u>Abbreviation / Acronym</u>	<u>Definition</u>
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).

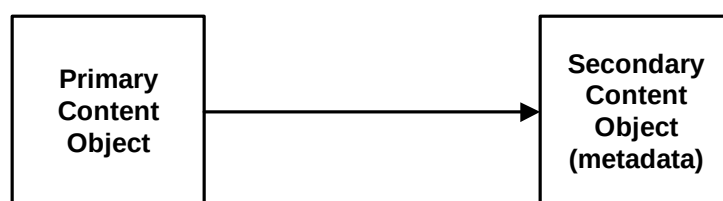
5.2.2 Content Objects

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.



Examples

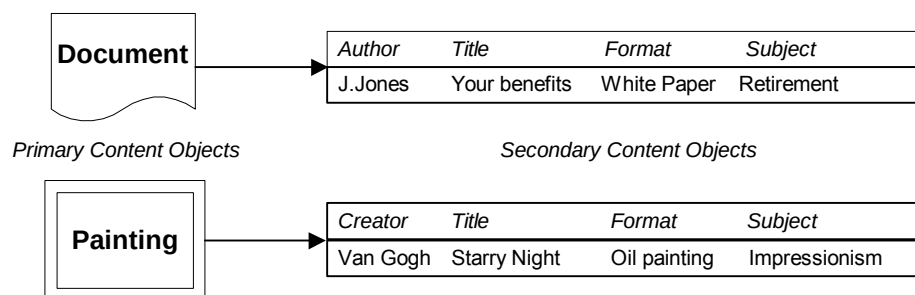


Figure 1: Content objects

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.

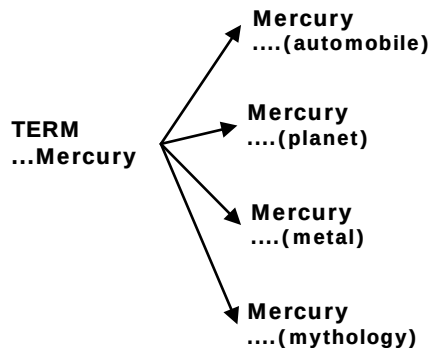


Figure 2: Ambiguity caused by homographs and polysemes

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

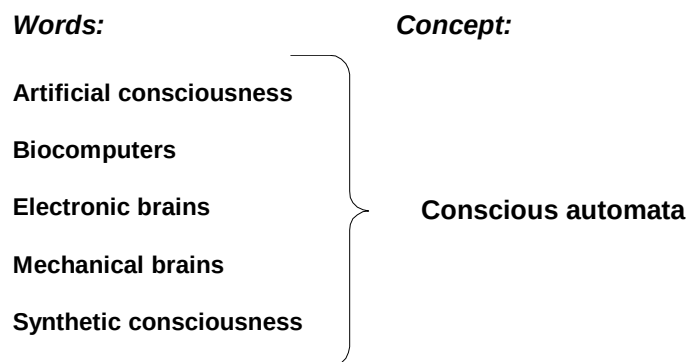


Figure 3: Information scatter caused by synonyms

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 than 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.

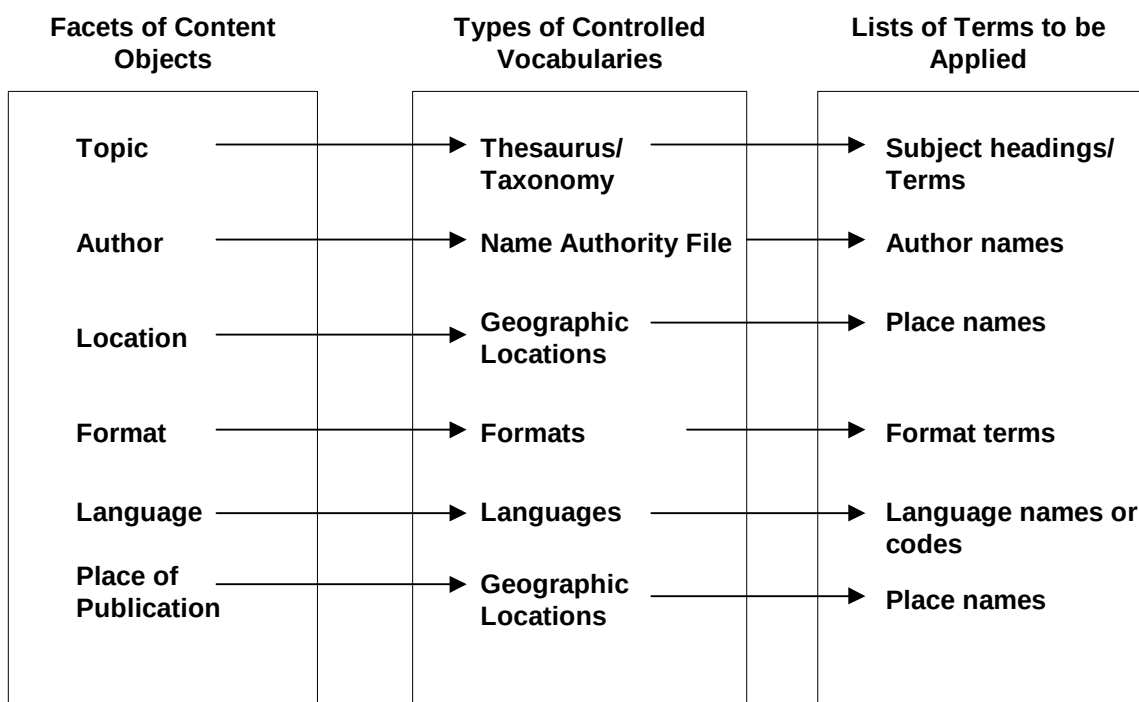


Figure 4: Facet analysis applied to content objects and controlled vocabularies

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

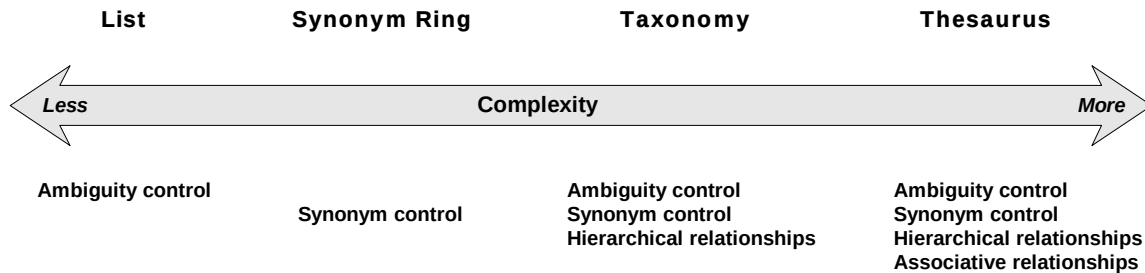


Figure 5: Increasing structural complexity among controlled vocabularies

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

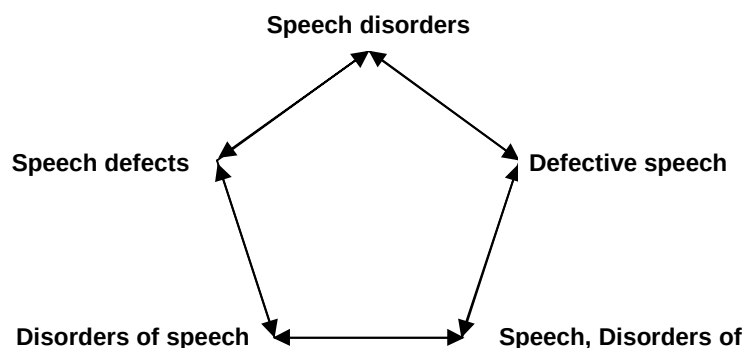
Mercury
Venus
Earth
Mars
Jupiter
Saturn
Uranus
Neptune
Pluto

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
      magnetohydrodynamics
  
```

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**

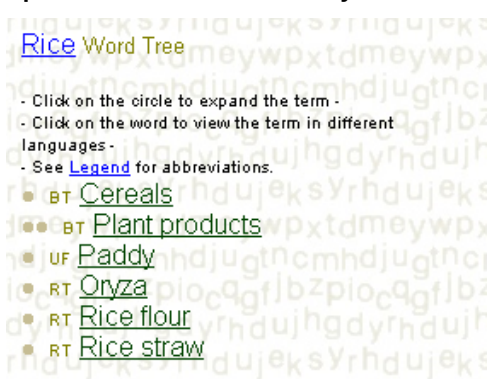
NT **BIOLOGICAL ABSORPTION**

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., **cooking** (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
cooking *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

<u>Adjectives used as terms</u>	<u>Precoordinated indexing terms</u>
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 dialectal 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

<u>Singular</u>	<u>Plural</u>
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.