

Introduction to

Controlled Vocabularies

Terminology for Art, Architecture,
and Other Cultural Works

First edition

Patricia Harpring

Murtha Baca, Series Editor

Published by the Getty Research Institute

The Getty Research Institute Publications Program

Thomas W. Gaechtgens, *Director, Getty Research Institute*
Gail Feigenbaum, *Associate Director*

Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works

Lauren Edson, *Manuscript Editor*

Elizabeth Zozom, *Production Coordinator*

Designed by Hesperheide Design, Newbury Park, California

Printed and bound by Odyssey Press, Inc., Gonic, New Hampshire

© 2010 J. Paul Getty Trust

Published by the Getty Research Institute, Los Angeles

Getty Publications

Gregory M. Britton, *Publisher*

1200 Getty Center Drive, Suite 500

Los Angeles, California 90049-1682

www.getty.edu

14 13 12 11 10 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Harpring, Patricia.

Introduction to controlled vocabularies : terminology for art,
architecture, and other cultural works / Patricia Harpring.

p. cm.

Includes bibliographical references.

ISBN 978-1-60606-018-6 (pbk.)

1. Subject headings—Cultural property. 2. Subject headings—Art.

3. Subject headings—Architecture. 4. Information retrieval. I. Title.

Z695.1.C85H37 2010

025.4'7—dc22

Cover: The story of the Tower of Babel (Genesis 11) was an allegory to explain why different societies spoke different languages (in addition to the obvious warnings against pride toward the deity and urban evils). Babel was a city in Babylon, where after the great flood, humanity was united in one large urban center, speaking a single language. In their pride, the inhabitants began construction of the Tower of Babel, with the intention of reaching the clouds of heaven. Their arrogant plan was foiled by God, who scattered them across the earth and confused their language so they could no longer understand each other.

Draftsman: Lieven Cruyl (Flemish, ca. 1640–ca. 1720); etcher: Coenraet Decker (Dutch, 1651–1685); *Tower of Babel*; etching; folio height 39 cm (15 $\frac{3}{8}$ inches); in Athanasius Kircher (German, 1601/1602–1680); *Athanasii Kircheri e Soc. Jesu Turris Babel*; published: Amsterdam: Ex officina Janssonio-Waesbergiana, 1679; Research Library; The Getty Research Institute (Los Angeles, California); 85-B16716-pl.[2].

Back cover: Antoine Babuty Desgodets, after George Marshall, *The Temple of Vesta at Tivoli* (detail). See fig. 24.

Contents

x	Foreword
xii	Acknowledgments
1	1. Controlled Vocabularies in Context
2	1.1. What Are Cultural Works?
2	1.1.1. Fine Arts
3	1.1.2. Architecture
3	1.1.3. Other Visual Arts
3	1.2. Creators of Art Information
3	1.2.1. Museums
4	1.2.2. Visual Resources Collections
5	1.2.3. Libraries
6	1.2.4. Special Collections
6	1.2.5. Archival Collections
7	1.2.6. Private Collections
7	1.2.7. Scholars
7	1.3. Standards for Art Information
8	1.3.1. Standards for the Creation of Vocabularies
9	1.3.2. Issues in Sharing Data
12	2. What Are Controlled Vocabularies?
12	2.1. Purpose of Controlled Vocabularies
12	2.2. Display Information and Controlled Information
13	2.2.1. Display Information with Controlled Vocabularies
14	2.2.2. Controlled Vocabularies vs. Controlled Format
16	2.3. Types of Controlled Vocabularies
16	2.3.1. Relationships in General
18	2.3.2. Subject Heading Lists
19	2.3.2.1. Other Headings
19	2.3.3. Controlled Lists
20	2.3.4. Synonym Ring Lists
21	2.3.5. Authority Files
22	2.3.6. Taxonomies
23	2.3.7. Alphanumeric Classification Schemes
24	2.3.8. Thesauri
24	2.3.9. Ontologies
26	2.3.10. Folksonomies
27	3. Relationships in Controlled Vocabularies
27	3.1. Equivalence Relationships
27	3.1.1. Synonyms

29	3.1.1.1. Lexical Variants
30	3.1.1.2. Historical Name Changes
30	3.1.1.3. Differences in Language
32	3.1.2. Near Synonyms
33	3.1.3. Preferred Terms
33	3.1.4. Homographs
35	3.1.4.1. Qualifiers
36	3.1.4.1.1. How to Choose a Qualifier for a Term
36	3.1.4.2. Other Ways to Disambiguate Names
37	3.2. Hierarchical Relationships
37	3.2.1. Whole/Part Relationships
38	3.2.2. Genus/Species Relationships
39	3.2.3. Instance Relationships
40	3.2.4. Facets and Guide Terms
41	3.2.5. Polyhierarchies
42	3.3. Associative Relationships
43	3.3.1. Types of Associative Relationships
45	3.3.2. When to Make Associative Relationships
49	4. Vocabularies for Cultural Objects
49	4.1. Types of Vocabulary Terms
51	4.2. The Getty Vocabularies
52	4.2.1. <i>Art & Architecture Thesaurus (AAT)</i>
53	4.2.1.1. Scope
54	4.2.1.1.1. Facets and Hierarchies in the <i>AAT</i>
56	4.2.1.2. What Constitutes a Term in the <i>AAT</i> ?
56	4.2.1.2.1. Warrant for a Term
57	4.2.1.2.2. Discrete Concepts
57	4.2.1.3. What Is Excluded from the <i>AAT</i> ?
57	4.2.1.4. Fields in the <i>AAT</i>
59	4.2.2. <i>Getty Thesaurus of Geographic Names (TGN)</i>
59	4.2.2.1. Scope
60	4.2.2.1.1. Nations, Cities, Archaeological Sites
60	4.2.2.1.2. Physical Features
60	4.2.2.1.3. Places That No Longer Exist
60	4.2.2.2. What Is Excluded from the <i>TGN</i> ?
60	4.2.2.2.1. Built Works
60	4.2.2.2.2. Cultural and Political Groups
61	4.2.2.3. Fields in the <i>TGN</i>
62	4.2.3. <i>Union List of Artist Names (ULAN)</i>
62	4.2.3.1. Scope

63	4.2.3.1.1. Artists
63	4.2.3.1.2. Architects
63	4.2.3.1.3. Non-Artists
63	4.2.3.1.4. Workshops and Families
64	4.2.3.1.5. Anonymous and Unknown Artists
64	4.2.3.1.6. Amateur Artists
64	4.2.3.2. What Is Excluded from the <i>ULAN</i> ?
65	4.2.3.3. Fields in the <i>ULAN</i>
65	4.2.4. <i>Cultural Objects Name Authority (CONA)</i>
65	4.2.4.1. Scope
67	4.2.4.1.1. Built Works
67	4.2.4.1.2. Movable Works
68	4.2.4.2. What Is Excluded from <i>CONA</i> ?
68	4.2.4.3. Fields in <i>CONA</i>
68	4.2.5. <i>Conservation Thesaurus (CT)</i>
71	4.3. <i>Chenhall's Nomenclature for Museum Cataloging</i>
71	4.3.1. Organization and Scope of <i>Nomenclature for Museum Cataloging</i>
71	4.3.2. Terms in <i>Nomenclature for Museum Cataloging</i>
72	4.3.3. <i>Nomenclature for Museum Cataloging</i> vs. the <i>AAT</i>
73	4.4. <i>Library of Congress Authorities</i>
74	4.4.1. <i>Library of Congress/NACO Authority File (LCNAF)</i>
75	4.4.2. <i>Library of Congress Subject Headings (LCSH)</i>
76	4.5. <i>Thesaurus for Graphic Materials (TGM)</i>
77	4.5.1. Scope of the <i>TGM</i>
77	4.5.2. The <i>TGM</i> vs. the <i>AAT</i>
80	4.6. <i>Iconclass</i>
80	4.6.1. Structure and Scope of <i>Iconclass</i>
83	5. Using Multiple Vocabularies
83	5.1. Interoperability of Vocabularies
84	5.2. Maintenance of Mappings
84	5.3. Methods of Achieving Interoperability
85	5.3.1. Direct Mapping
86	5.3.2. Switching Vocabulary
86	5.3.3. Factors for Successful Interoperability of Vocabularies
89	5.3.4. Semantic Mapping
90	5.4. Interoperability across Languages
90	5.4.1. Issues of Multilingual Terminology
92	5.4.2. Dominant Languages
92	5.5. Satellite and Extension Vocabularies

94	6.	Local Authorities
96	6.1.	Which Fields Should Be Controlled?
97	6.2.	Structure of the Authority
97	6.3.	Unique IDs in the Authority
99	6.4.	Person/Corporate Body Authority
101	6.4.1.	Sources for Terminology
102	6.4.2.	Suggested Fields
106	6.5.	Place/Location Authority
107	6.5.1.	Sources for Terminology
108	6.5.2.	Suggested Fields
113	6.6.	Generic Concept Authority
114	6.6.1.	Sources for Terminology
115	6.6.2.	Suggested Fields
119	6.7.	Subject Authority
121	6.7.1.	Sources for Terminology
122	6.7.2.	Suggested Fields
130	6.8.	Source Authority
130	6.8.1.	Sources for Terminology
130	6.8.2.	Suggested Fields
133	7.	Constructing a Vocabulary or Authority
133	7.1.	General Criteria for the Vocabulary
133	7.1.1.	Local or Broader Use
134	7.1.2.	Purpose of the Vocabulary
134	7.1.3.	Scope of the Vocabulary
135	7.1.4.	Maintaining the Vocabulary
135	7.2.	Data Model and Rules
135	7.2.1.	Established Standards
136	7.2.2.	Logical Focus of the Record
136	7.2.3.	Data Structure
137	7.2.4.	Controlled Fields vs. Free-Text Fields
137	7.2.5.	Minimum Information
138	7.2.6.	Editorial Rules
138	7.3.	Imprecise Information
140	7.4.	Rules for Constructing a Vocabulary
140	7.4.1.	Establishing Terms
141	7.4.1.1.	Capitalization
141	7.4.2.	Regulating Hierarchical Relationships
142	7.4.2.1.	Mixing Relationships
142	7.4.2.2.	Incorporating Facets and Guide Terms

144	7.5. Displaying a Controlled Vocabulary
144	7.5.1. Display for Various Types of Users
145	7.5.2. Technical Considerations
145	7.5.2.1. Display Independent of Database Design
145	7.5.3. Characteristics of Displays
146	7.5.3.1. Format of Display
146	7.5.3.2. Documentation
147	7.5.3.3. Displaying Hierarchies
147	7.5.3.3.1. Indentation vs. Notations
149	7.5.3.3.2. Alternative Hierarchical Displays
149	7.5.3.3.3. Display of Polyhierarchy
152	7.5.3.3.4. Sorting of Siblings
153	7.5.3.3.5. Faceted Displays and Guide Terms
154	7.5.3.3.6. Classification Notation or Line Number
155	7.5.3.4. Full Record Display
155	7.5.3.5. Displaying Equivalence and Associative Relationships
157	7.5.3.5.1. Permuted Lists and Inverted Forms
157	7.5.3.5.2. Displaying Homographs
158	7.5.3.5.3. Sorting and Alphabetizing Terms
159	7.5.3.5.4. Diacritics in Sorting
160	7.5.3.5.5. Display of Diacritics
160	7.5.3.6. Search Results Displays
160	7.5.3.6.1. Headings or Labels
162	7.5.3.6.2. Ascending or Descending Order of Parents
162	7.5.3.6.3. Displaying the User's Search Term
164	7.5.3.7. Pick Lists
165	8. Indexing with Controlled Vocabularies
165	8.1. Technical Issues of Indexing
165	8.1.1. Availability of Indexing Terms to the Cataloger
167	8.2. Methodologies for Indexing
167	8.2.1. Indexing Display Information
167	8.2.2. When Fields Do Not Display to End Users
168	8.2.3. Specificity and Exhaustivity
168	8.2.3.1. Specificity Related to the Authority Records
169	8.2.3.2. General and Specific Terms
170	8.2.3.3. Preferred or Variant Terms
170	8.2.3.4. How Many Terms
170	8.2.3.4.1. How to Establish Core Elements

171	8.2.3.4.2. Minimal Records
171	8.2.3.4.3. Missing Information
172	8.2.3.5. Size and Focus of the Collection
172	8.2.3.5.1. Different Works Require Different Indexing
172	8.2.3.5.2. Cataloging in Phases
172	8.2.3.5.3. Indexing Groups vs. Items
173	8.2.3.5.4. Expertise of End Users
174	8.2.3.5.5. Expertise of Catalogers and Indexers
174	8.2.4. Indexing Uncertain Information
176	8.2.4.1. Knowable vs. Unknowable Information
176	8.2.4.1.1. Knowable Information
176	8.2.4.1.2. Debated Information
177	9. Retrieval Using Controlled Vocabularies
177	9.1. Identifying the Focus of Retrieval
178	9.2. User Intervention or Behind the Scenes
178	9.2.1. Retrieval by Browsing
181	9.2.2. Retrieval via Search Box
182	9.2.3. Retrieval by Querying in a Database
185	9.2.3.1. Reports and Ad Hoc Queries of the Database
186	9.2.4. Querying across Multiple Databases
186	9.2.5. Seeding Tags with Vocabulary Terms
187	9.3. Processing Vocabulary Data for Retrieval
188	9.3.1. Know Your Audience
188	9.3.2. Using Names for Retrieval
189	9.3.3. Truncating Names
190	9.3.4. Keyword Searching
191	9.3.5. Normalizing Terms
192	9.3.5.1. Case Insensitivity in Retrieval
193	9.3.5.2. Compound Terms and Names in Retrieval
193	9.3.5.3. Diacritics and Punctuation in Retrieval
194	9.3.5.4. Phonetic Matching
195	9.3.5.5. Singulars and Plurals in Retrieval
196	9.3.5.6. Abbreviations
196	9.3.5.7. Trunk Names
197	9.3.5.8. Form and Syntax of the Name
197	9.3.5.8.1. First and Last Names
197	9.3.5.8.2. Pivoting on the Comma
198	9.3.5.8.3. Multiple Commas

199	9.3.5.9. Articles and Prepositions
200	9.3.6. Reserved Character Sets
201	9.3.7. Stop Lists
201	9.3.8. Boolean Operators
201	9.3.9. Context of Terms in Retrieval
202	9.3.9.1. Qualifiers in Retrieval
202	9.3.9.2. Hierarchical Relationships in Retrieval
204	9.3.9.3. Associative Relationships in Retrieval
204	9.4. Other Data Used in Retrieval
204	9.4.1. Unique Identifiers as Search Criteria
205	9.4.2. Other Vocabulary Data Used in Retrieval
205	9.5. Results Lists
207	Appendix: Selected Vocabularies and Other Sources for Terminology
210	Glossary
239	Selected Bibliography

Foreword

The Getty Vocabulary Program has devoted almost three decades to building thesauri that can be used as knowledge bases, cataloging and documentation tools, and online search assistants. In addition to building tools for use by art and cultural heritage professionals and the general public, we also provide training opportunities and educational materials on how to build and implement controlled vocabularies. Part of our mission as an institution devoted to research and education is to share our knowledge and expertise with the international art and cultural heritage communities in their broadest sense.

Elisa Lanzi's *Introduction to Vocabularies*, which appeared in print in 1998 and was updated in an online version in 2000, offers a general overview of vocabularies for art and material culture. *Introduction to Controlled Vocabularies* is a much more detailed "how-to" guide to building controlled vocabulary tools, cataloging and indexing cultural materials with terms and names from controlled vocabularies, and using vocabularies in search engines and databases to enhance discovery and retrieval in the online environment.

"How forceful are right words!" is written in Job 6:25. The King James Version of the Bible uses the word *forcible*, meaning "forceful" or "powerful," instead. In the online environment, words have the power to lead users to the information resources that they seek. But we should not force users to know what we consider to be the "right" word or name in order for them to be able to obtain the best search results.

We recognize that a single concept can be expressed by more than one word, and that a single word can express more than one concept. Words can change over time and take a variety of forms, and they can be translated into many languages. A carefully constructed controlled vocabulary provides catalogers and others who create descriptive metadata with the "right" or "preferred" name or term to use in describing collections and other resources, but it also clusters together all of the synonyms, orthographic and grammatical variations, historical forms, and even in some cases "wrong" names or terms in order to enhance access for a broad range of users without constraining them to the use of the "right" term. With millions of searches being

conducted by millions of users each day via Web search engines and in proprietary databases, the power of words is a crucial factor in providing access to the wealth of information resources now available in electronic form. We hope that this book will provide organizations and individuals who wish to enhance access to their collections and other online resources with a practical tool for creating and implementing vocabularies as reference tools, sources of documentation, and powerful enhancements for online searching.

Murtha Baca, Getty Research Institute

Acknowledgments

I wish to thank Murtha Baca for her lasting support, thoughtful guidance, and expert editing. I am also grateful to Joan Cobb, Gregg Garcia, Marcia Zeng, and Karim Boughida, who provided invaluable advice on the technical aspects of this book. I extend sincere thanks to the indefatigable Getty Vocabulary Program editors Antonio Beecroft, Ming Chen, Robin Johnson, and Jonathan Ward, who proofed the manuscript and provided important feedback.

Finally, thanks go to the scores of earlier vocabulary editors and users of the Getty vocabularies, who have provided countless insights and advice over the last three decades.

Patricia Harpring