Metadata for Digital Collections:

A How-To-Do-It Manual®

Steven J. Miller

HOW-TO-DO-IT MANUALS®

NUMBER 179

Neal-Schuman Publishers, Inc.

New York **London**

Published by Neal-Schuman Publishers, Inc.
100 William St., Suite 2004
New York, NY 10038

Copyright © 2011 Neal-Schuman Publishers, Inc.

"A How-To-Do-It Manual®" and "A How-To-Do-It Manual for Librarians®" are registered trademarks of Neal-Schuman Publishers, Inc.

All rights reserved. Reproduction of this book, in whole or in part, without written permission of the publisher, is prohibited.

Printed and bound in the United States of America.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1992.

Library of Congress Cataloging-in-Publication Data

ISBN: 978-1-55570-746-0

CIP Pending

Contents

List of Illustrations ix
Foreword xv
Preface xvii
Acknowledgments xxiii

1. Introduction to Metadata for Digital Collections 1

1.1. What Is Metadata? 1
1.2. What Is a Digital Collection? 6
1.3. What Does Metadata Do? 9
1.4. Types of Metadata 10
1.5. Metadata Standards 12
1.6. Creating a Digital Collection 14
1.7. Metadata for Digital Collections 17
1.7.1. Designing and Documenting a Metadata Scheme 18
1.7.2. Creating Metadata for Digital Objects 19
1.7.3. Metadata Sharing, Harvesting, and Aggregating 22
1.8. Summary 22
References 23

2. Introduction to Resource Description and Dublin Core 25

2.1. Resource Description Fundamentals 25
2.1.1. Resource Description 26
2.1.1.1. Resources 26
2.1.1.2. Metadata Descriptions and Records 28
2.1.1.3. Granularity of Description 31
2.1.1.4. Element Repeatability 33
2.1.1.5. Element Functionality 34
2.1.2. Local versus Standard, Shareable Element Sets 37
2.1.3. Describing Digital versus Original Resources 40
2.1.3.1. The One-to-One Principle 41
2.1.3.2. Content versus Carrier 42
2.1.3.3. Problems with the One-to-One Principle in Practice 43
2.1.3.4. Practical Options for Maintaining One-to-One 44

2.1.4. Descriptive versus Administrative Metadata 46
2.1.5. The Need for Research 46
2.2. Introduction to the Dublin Core Metadata Element Set 49
2.2.1. Simple (Unqualified) Dublin Core 50
2.2.2. Qualified Dublin Core 52
2.2.3. Creation and Use of Dublin Core Metadata 54
2.2.4. The Dublin Core Elements in Practice 56
2.3. Summary 57
References 58

3. Resource Identification and Responsibility Elements 59

3.1. Basic Resource Identification Elements 60
3.1.1. Titles 60
3.1.2. Dublin Core *Title* 63
3.1.3. Identifiers 66
3.1.4. Dublin Core *Identifier* 66
3.1.5. Dates 68
3.1.6. Dublin Core *Date* 70
3.1.7. Languages 74
3.1.8. Dublin Core *Language* 75
3.1.9. Resource Attributes Not Readily Accommodated in Dublin Core 76
3.2. Name, Responsibility, and Intellectual Property Elements 78
3.2.1. Names and Roles of Agents Responsible for Resources 78
3.2.2. Dublin Core *Creator* and *Contributor* 80
3.2.3. Publishers and Publication 83
3.2.4. Dublin Core *Publisher* 84
3.2.5. Rights, Ownership, and Restrictions on Access 85
3.2.6. Dublin Core *Rights* 86
3.3. Summary 86
References 87

4. Resource Content and Relationship Elements 89

4.1. Resource Content and Carrier Elements 89
4.1.1. Content Types and Genres 89
4.1.2. Dublin Core *Type* 91
4.1.3. Formats and Physical Description 94
4.1.4. Dublin Core *Format* 94
4.2. Subject Content Elements 98
4.2.1. Subjects 99
4.2.1.1. Subject Analysis, Representation, and Retrieval 99
4.2.1.2. Analyzing and Identifying Subject Content 100
4.2.1.3. Aboutness, Ofness, Isness, and Facets 101
4.2.1.4. Exhaustivity: Number of Subject Terms 103
4.2.1.5. Specificity: Specific versus General Subject Terms 104

4.2.1.6. Subject Analysis and Indexing of Images 105
4.2.2. Dublin Core *Subject* 109
4.2.3. Dublin Core *Coverage* 113
4.2.4. Descriptions, Abstracts, and Tables of Contents 115
4.2.5. Dublin Core *Description* 117
4.3. Resource Relationship Elements 118
4.3.1. Relationships among Different Resources 118
4.3.2. Dublin Core *Relation* and *Source* 119
4.4. Dublin Core Full Record Examples 124
4.5. Mapping Local Elements to Dublin Core 126
4.6. Summary 126
References 128

5. Controlled Vocabularies for Improved Resource Discovery 129

5.1. Improving Resource Discovery 129
5.2. Types of Controlled Vocabularies 132
5.2.1. Lists 134
5.2.2. Synonym Rings 136
5.2.3. Authority Files 136
5.2.4. Taxonomies and Classification Schemes 138
5.2.5. Thesauri 138
5.2.6. Subject Heading Lists 140
5.3. Using Established Vocabularies 140
5.4. Creating Your Own Vocabularies 144
5.5. Summary 146
References 147

6. XML-Encoded Metadata 149

6.1. XML Metadata Basics 149
6.1.1. Introduction to Metadata Encoding and XML 149
6.1.2. XML Syntax: Elements and Attributes 150
6.1.3. Well-Formed versus Valid XML 152
6.1.4. XML Namespaces and Metadata Modularity 153
6.1.5. Creating Metadata in XML 154
6.2. XML Metadata Record Examples 155
6.2.1. Dublin Core in XML 155
6.2.2. MODS XML 157
6.3. Anatomy of an XML Metadata Record 159
6.4. Summary 161
References 161

7. MODS: The Metadata Object Description Schema 163

7.1. Introduction and Overview 164
7.1.1. MODS Implementation Projects 164
7.1.2. MODS Documentation 165
7.1.3. MODS XML Structure 166

7.1.3.1. Container Elements and Subelements 166
7.1.3.2. Element Attributes 167
7.1.4. Flexibility in MODS Level of Detail and Granularity 168
7.2. MODS Elements: An Overview with Examples 168
7.2.1. *TitleInfo* 170
7.2.2. *Name* 172
7.2.3. *TypeOfResource* 175
7.2.4. *Genre* 176
7.2.5. *OriginInfo* 177
7.2.6. *Language* 181
7.2.7. *PhysicalDescription* 182
7.2.8. *Abstract* 183
7.2.9. *TableOfContents* 184
7.2.10. *TargetAudience* 184
7.2.11. *Note* 185
7.2.12. *Subject* 186
7.2.13. *Classification* 186
7.2.14. *RelatedItem* 189
7.2.15. *Identifier* 191
7.2.16. *Location* 192
7.2.17. *AccessCondition* 192
7.2.18. *Part* 193
7.2.19. *Extension* 194
7.2.20. *RecordInfo* 194
7.3. MODS Records 195
7.3.1. Complete MODS Record Example 195
7.3.2. Creating MODS XML Records 195
7.3.3. Displaying and Transforming MODS XML Records 201
7.3.4. Qualified Dublin Core and MODS Record Comparison 204
7.4. Mapping from Dublin Core to MODS 208
7.4.1. Automated Mapping from Simple Dublin Core to Simple MODS 208
7.4.2. Human Mapping from Qualified Dublin Core to Richer MODS 209
7.5. Summary 211
References 212

8. VRA Core: The Visual Resources Association Core Categories 213

8.1. Introduction and Overview 213
8.1.1. Metadata for Museum Objects 213
8.1.2. Metadata Standards for Museum Objects 214
8.2. VRA 3.0 Overview 216
8.3. VRA 3.0 Record Examples 218
8.4. VRA 4.0 Overview 219
8.5. VRA 4.0 Record Examples 221

8.6. Summary 225
References 225

9. Metadata Interoperability, Shareability, and Quality 227

9.1. Interoperability 227
9.2. Short- and Long-Term Metadata Viability 227
9.3. Metadata Sharing, Harvesting, and Aggregating 228
9.4. OAI Metadata Harvesting 229
9.5. Metadata Mapping and Crosswalks 231
9.6. Metadata Conversion and Processing 234
9.7. Example of Metadata Harvesting, Processing, and Aggregating 237
9.8. Good Quality and Shareable Metadata 241
9.9. Assessing Metadata Quality 242
9.10. Five Ways to Improve Your Metadata Quality and Interoperability 244
9.11. Summary 247
References 248

10. Designing and Documenting a Metadata Scheme 251

10.1. Metadata Scheme Design and Documentation 251
10.1.1. Introduction 251
10.1.2. Analyze Context, Content, and Users, and Determine Functional Requirements 252
10.1.3. Select and Develop an Element Set 255
10.1.3.1. General/Cross-Collection Metadata Scheme Design 255
10.1.3.2. Collection-Specific Metadata Scheme Design 256
10.1.3.3. Factors in Choice of Metadata Element Set 257
10.1.4. Establish Element and Database Specifications 258
10.1.5. Establish Controlled Vocabularies and Encoding Schemes 259
10.1.6. Develop Content Guidelines 260
10.1.7. Document the Scheme 261
10.2. Metadata Design Examples 262
10.2.1. General Application Profile Examples 263
10.2.1.1. Collaborative Digitization Program Dublin Core Metadata Documentation 263
10.2.1.2. OhioLINK Dublin Core Metadata Documentation 272
10.2.1.3. Indiana Memory Dublin Core Metadata Documentation 272
10.2.1.4. DLF/Aquifer MODS Metadata Documentation 272
10.2.2. Collection-Specific Application Profile Examples 284

10.2.2.1. University of Washington's Architecture Collection Metadata Documentation 288
10.2.2.1. University of Washington's Musical Instruments Collection Metadata Documentation 288
10.2.2.1. University of Wisconsin–Milwaukee's Transportation Collection Metadata Documentation 288
10.2.3. CONTENTdm Examples 292
10.3. Summary 300
References 301

11. Metadata, Linked Data, and the Semantic Web

11.1. What Are Linked Data and the Semantic Web and Why Care about Them? 303
11.2. Linked Open Data and the Resource Description Framework 305
11.2.1. Statements, Properties, Values, and RDF Triples 306
11.2.2. URIs: Uniform Resource Identifiers 307
11.2.3. Literals, Strings, and Things 310
11.2.4. The Power of Linking and Querying in the Linked Data Cloud 312
11.2.5. RDF/XML 314
11.3. Linked Data and Digital Collections 316
11.4. Dublin Core: From a Core Metadata Element Set for the Web to a Core Vocabulary for Linked Data 317
11.4.1. The DCMI Abstract Model (DCAM) 317
11.4.2. Dublin Core Application Profiles 319
11.5. Metadata Registries 319
11.6. What Does All of This Have to Do with Me? 321
11.7. Summary 322
References 323

Bibliography 325
Index 333
About the Author

List of Illustrations

Figures

Figure 1.1	Library Catalog Record: Metadata about a Book	4
Figure 1.2	Digital Image in an Online Digital Collection	5
Figure 1.3	American Variety Stage: Vaudeville and Popular Entertainment, 1870–1920	7
Figure 1.4	Southern Oral History Program Interview Database	8
Figure 1.5	Smithsonian American Art Museum George Catlin Indian Paintings	8
Figure 1.6	CONTENTdm "Collection Field Properties" Screen Example	20
Figure 1.7	CONTENTdm Metadata Record Creation/ Editing Screen Example	21
Figure 2.1	Digital Image	27
Figure 2.2	Digital Text: Book Digitized as Separate Image Files (Complex Digital Object)	28
Figure 2.3	Metadata Record for Digital Image	29
Figure 2.4	Metadata Record for Complex Digital Object (Book Digitized as Set of Image Files)	30
Figure 2.5	Element Repeatability: Multiple Values in a Single Field versus Separate Fields	33
Figure 2.6	Search Options Based on Metadata Fields	34
Figure 2.7	Drop-Down Browse Menus Based on Metadata Fields and Values	34
Figure 2.8	Metadata Record for a Digital Image	35
Figure 2.9	Search Options Based on Metadata Fields	36
Figure 2.10	Browse Options Based on Metadata Fields and Values	37
Figure 2.11	Metadata Fields and Values That Underlie Functionality	38
Figure 2.12	Functionality: Description versus Indexing	39
Figure 2.13	Functionality: Results of Indexed/Hyperlinked Subject Terms	40
Figure 2.14	The Need for Research	47

Figure 3.1	Example of a Typical Title Display	61
Figure 3.2	Digital Image: Many Possible Supplied Titles	62
Figure 4.1	Political Cartoon Example of *Ofness* versus *Aboutness*	107
Figure 4.2	Digital Image Example of *Ofness* versus *Aboutness*	108
Figure 4.3	Description Element Example	116
Figure 5.1	Controlled Vocabulary Term Browse	130
Figure 5.2	Personal Name Browse	131
Figure 5.3	Simple Hierarchical Browse	133
Figure 5.4	Typology of Controlled Vocabularies from ANSI/NISO Z39.19	134
Figure 5.5	Controlled Subject Terms Used as Flat List in CONTENTdm	135
Figure 5.6	LC/NACO Name Authority Example	137
Figure 5.7	National Agricultural Thesaurus Record Example	139
Figure 5.8	TGM Record for the Term *Drawbridges*	141
Figure 5.9	AAT Record for the Term *drawbridges*	142
Figure 5.10	Subject Browse with Equivalence Relationships Implemented	145
Figure 7.1	Public Display of MODS Record	198
Figure 7.2	oXygen XML Editor Example	199
Figure 7.3	oXygen XML Editor <name> Subelement Example	200
Figure 7.4	oXygen XML Editor <roleTerm> Attribute Example	201
Figure 7.5	oXygen XML Editor Grid View Example	202
Figure 7.6	University of Alberta MODS Editor <titleInfo> Example	202
Figure 7.7	University of Alberta MODS Editor <originInfo> Example	203
Figure 7.8	University of Alberta MODS Editor <subject> Example	204
Figure 9.1	OAI Metadata Harvesting	230
Figure 9.2	OAIster Advanced Search: Fielded Search Options	231
Figure 9.3	OAIster Advanced Search: Resource Type Limit Options	232
Figure 9.4	Digital Image	237
Figure 10.1	CDP MAP Table of Contents Page	264
Figure 10.2	CDP MAP List of Mandatory and Optional Elements	265
Figure 10.3	CDP MAP General Input Guidelines	266
Figure 10.4	CDP MAP *Title* Element Specifications, First Page	267
Figure 10.5	CDP MAP *Title* Element Specifications, Second Page	268
Figure 10.6	CDP MAP *Coverage* Element Specifications, First Page	269
Figure 10.7	CDP MAP *Coverage* Element Specifications, Second Page	270

Figure 10.8	CDP MAP *Coverage* Element Specifications, Third Page	271
Figure 10.9	OhioLINK MAP Table of Contents	273
Figure 10.10	OhioLINK MAP Core Element Set and General Input Guidelines	274
Figure 10.11	OhioLINK MAP General Input Guidelines, Continued	275
Figure 10.12	OhioLINK MAP *Title* Element Specifications	276
Figure 10.13	OhioLINK MAP *Creator* Element Specifications	277
Figure 10.14	OhioLINK MAP *Subject* Element Specifications	278
Figure 10.15	Indiana Memory MAP Introductory and *Title* Element Documentation	279
Figure 10.16	Indiana Memory MAP *Subject* Element Documentation, Beginning	280
Figure 10.17	Indiana Memory MAP *Subject*, *Date.Original*, and *Date.Digital* Elements	281
Figure 10.18	Indiana Memory MAP Recommended, *Creator*, and *Publisher* Elements	282
Figure 10.19	DLF/Aquifer MAP Summary, First Page	283
Figure 10.20	DLF/Aquifer MAP Summary, Second Page	284
Figure 10.21	DLF/Aquifer MAP <typeOfResource> Element, First Page	285
Figure 10.22	DLF/Aquifer MAP <typeOfResource> Element, Second Page	286
Figure 10.23	DLF/Aquifer MAP <typeOfResource> Element, Third Page	287
Figure 10.24	Transportation Around the World: Graphical Map Browse Screen	292
Figure 10.25	Transportation Around the World: Browse Method Screen	293
Figure 10.26	CONTENTdm Add Collection Screen	294
Figure 10.27	CONTENTdm Metadata Fields Screen Before Customizing	295
Figure 10.28	CONTENTdm Edit Field Screen	296
Figure 10.29	CONTENTdm Select Controlled Vocabulary Screen	297
Figure 10.30	CONTENTdm Set Default Values Screen	297
Figure 10.31	CONTENTdm Metadata Fields Screen After Customizing	298
Figure 10.32	CONTENTdm Metadata Entry/Record Creation Screen	299
Figure 10.33	CONTENTdm Export Metadata Screen	300
Figure 11.1	The Linking Open Data Cloud Diagram: Selection	304
Figure 11.2	The Structure of an RDF Statement or Triple	306
Figure 11.3	Subject Statement Shown as Graph	308

Figure 11.4 *Thesaurus of Graphic Materials* Term in RDF 309
Figure 11.5 Graph of Three Triples Expressed as Literals/Strings 311
Figure 11.6 Graph with Two Triples Expressed as URIs/Things 311
Figure 11.7 Graph of Triples for a Related Resource 312
Figure 11.8 Two Sets of Triples Linked Based on Matching URI 313
Figure 11.9 The DCMI Resource Model 319
Figure 11.10 The Singapore Framework for DCAPs 320

Tables

Table 1.1 Metadata in a Microsoft Word Document 3
Table 1.2 Metadata for an Album of Songs in iTunes 3
Table 1.3 Technical Metadata about a Digital Photograph 4
Table 1.4 Metadata about the Digital Image 5
Table 1.5 Functionality: Faceted Navigation 11
Table 2.1 Collection-Level Record (Selected Elements) 31
Table 2.2 Item-Level Record (Selected Elements) 32
Table 2.3 Customized Local versus Standard Dublin Core Elements 41
Table 2.4 Content versus Carrier 42
Table 2.5 One-to-One Principle: Separate Linked Records for Original and Digital Resources 43
Table 2.6 One-to-One Principle: Single Record Using Free-Text Source Element for Information about the Original Resource 44
Table 2.7 Local Metadata after Harvesting as Simple Dublin Core: Value of *Source* Element for Clarifying Various Dates, Creator Roles, and Identifier Numbers in Body of Metadata 45
Table 2.8 Descriptive versus Administrative Metadata 46
Table 2.9 The Results of Research (Selected Elements) 48
Table 2.10 The Dublin Core Metadata Element Set 50
Table 2.11 DCMES Grouping Example 1 51
Table 2.12 DCMES Grouping Example 2 51
Table 2.13 Simple Dublin Core Record Example 52
Table 2.14 Dublin Core Qualifiers 54
Table 2.15 Qualified Dublin Core Representation Example 55
Table 3.1 Title Elements in DC, MODS, and VRA 64
Table 3.2 Identifier Elements in DC, MODS, and VRA 67
Table 3.3 Date Range Search Limit Example 69
Table 3.4 Individual Date Search Limit Example 69
Table 3.5 Example of Potential Problems with Indexing Uncertain Dates and Ranges 69

Table 3.6	Date Elements in DC, MODS, and VRA	71
Table 3.7	Language Elements in DC, MODS, and VRA	75
Table 3.8	Selected Elements, Subelements, and Attributes in MODS and VRA Not Readily Accommodated in DC	79
Table 3.9	Name Elements in DC, MODS, and VRA	80
Table 3.10	Publication Elements in DC, MODS, and VRA	84
Table 3.11	Rights Elements in DC, MODS, and VRA	85
Table 4.1	Broad, Generic Type of Resource Search Limit Example	90
Table 4.2	Broad, Generic Type of Resource Browse Example	90
Table 4.3	More Specific Type of Resource or Genre Search Limit Example	90
Table 4.4	Type and Genre Elements in DC, MODS, and VRA	91
Table 4.5	Format and Physical Description Elements in DC, MODS, and VRA	95
Table 4.6	Format Elements for Both Original and Digital Manifestations in a Single Record	98
Table 4.7	Subject Elements in DC, MODS, and VRA	110
Table 4.8	Description Elements in DC, MODS, and VRA	117
Table 4.9	Relationship Elements in DC, MODS, and VRA	120
Table 4.10	Simple Dublin Core Record Example	125
Table 4.11	Qualified Dublin Core Record Example	125
Table 4.12	Simple Dublin Core Record Example Using Detailed Source Element	126
Table 4.13	Local Collection-Specific Elements Mapped to Dublin Core	127
Table 5.1	Uncontrolled and Inconsistent Values for Resource Type	130
Table 5.2	Variant Forms of a Playwright's Name	131
Table 5.3	Inclusion of Equivalence Relationships in Flat Lists	136
Table 5.4	Subject Metadata Example Using Term from TGM	143
Table 5.5	Subject Metadata Example Using Term from AAT	143
Table 5.6	More Detailed Subject Metadata Example Using Terms from AAT	144
Table 6.1	Simple *Place* Element Example in XML Format	151
Table 6.2	Simple *Place* Element Example in Tabular Database Format	151
Table 6.3	Simple XML *Place* Element with an XML Attribute	151
Table 6.4	Simple XML *Place* Element with Two XML Attributes	151
Table 6.5	Table Format Example of Element with Two Qualifiers	151
Table 6.6	Anatomy of an XML Metadata Record	160
Table 7.1	MODS Top-Level Elements and Subelements	169
Table 7.2	Side-by-Side Comparison of Qualified Dublin Core and MODS Records	205

Table 7.3	Simple Dublin Core Mapping to MODS Version 3	210
Table 7.4	Conversion of DC Resource Type Vocabulary to MODS TypeOfResource Values	211
Table 7.5	Three Simple Dublin Core Elements Mapped to MODS Lite	211
Table 7.6	Three Qualified Dublin Core Elements Mapped to Full MODS	211
Table 8.1	VRA 3.0 Categories, Qualifiers, Data Values, and DC Mappings	217
Table 8.2	VRA 3.0 Data Set for an Etching (Work Record)	218
Table 8.3	VRA 3.0 Data Set for an Image of the Etching (Image Record)	218
Table 8.4	VRA 4.0 Elements, Subelements, and Attributes	219
Table 9.1	Local Metadata with Mapping to Dublin Core	233
Table 9.2	Local Metadata after Mapping to Simple Dublin Core	234
Table 9.3	MODS to Dublin Core Metadata Element Set Mapping	235
Table 9.4	Conversion of MODS TypeOfResource Values to DC Resource Type Vocabulary	236
Table 9.5	Original Local Metadata	238
Table 9.6	Harvested Metadata in Statewide Consortial Repository	239
Table 9.7	Harvested Metadata in International OAIster Repository	240
Table 9.8	Metadata Records Viewed in Spreadsheet Table	243
Table 9.9	Spreadsheet Table Sorting to Reveal Missing and Disparate Values	244
Table 10.1	Architecture of the Pacific Northwest Data Dictionary: Selections	289
Table 10.2	Musical Instruments in the University of Washington Ethnomusicology Division Data Dictionary: Selections	290
Table 10.3	Transportation Around the World Database Fields	291
Table 11.1	Dublin Core Metadata Record in Table Format	307
Table 11.2	Dublin Core Metadata Record in XML Format	307
Table 11.3	Statements about a Digital Image	308
Table 11.4	URIs Used as Subjects in RDF Triples	308
Table 11.5	URIs Used as Properties in RDF Triples	309
Table 11.6	URIs Used as Values in RDF Triples	310
Table 11.7	Three Triples Expressed as Literals/Strings	310
Table 11.8	Two Triples Expressed as URIs/Things	311

Foreword

Back in the dark ages of the last decade, when first talking to publishers about putting together my own contribution to the metadata canon, it was suggested to me that I write more of a how-to book, but I demurred, largely because of the enormous effort required. As it happens, I broke my right arm early in the process of creating *Metadata in Practice*, so it's just as well that my actual writing effort was somewhat limited, as I could do the recruiting of contributors (and the inevitable herding) by telephone.

With that as background, I was really happy to hear that Steve was writing the book I hadn't the fortitude to attempt back then, and particularly pleased that he's done such a nice job of it. The metadata community needs this book—at all levels, from the beginner to the practitioner to the teacher. As it is, when those of us who teach about metadata these days go about gathering material, it's something of a treasure hunt, and each discovered resource needs to be presented with a map of where it fits in the pantheon, a bunch of caveats about age and suitability for particular purposes, and apologies for the fact that there is not a single resource that covers it all. Steve has relieved us of all that, for which we are very, very grateful. He's also included in the book the kinds of pointers to the important work of others that expands on his more comprehensive approach.

The other great news about this book is that Steve has been paying close attention to the Semantic Web, and he understands well that though we may be building our metadata using current technology we *must* pay attention to where our world is shifting. I predict that in the future we will look upon that shift as the most important change in our corner of the profession since Henriette Avram started thinking about automating the printing of catalog cards. Steve's approach—sensible and accessible to his audience—is to include that information in the relevant beginning portions and as a separate chapter. The reality is that we are on a moving sidewalk of transition that will be a part of our lives for most of (if not all of) our careers.

There is no shortcut—no one-size-fits-all template—available to us as we plan today for a tomorrow that will be rife with change, not all of which we can predict. But readers of this book will be as prepared as

it is possible to be for whatever the future hands us. Thanks, Steve; we're in your debt.

Diane I. Hillmann
Director of Metadata Initiatives
Information Institute of Syracuse
Syracuse, New York;
Partner
Metadata Management Associates

Preface

Good quality metadata is critical for providing intellectual access to the ever-increasing number of digital collections being created by libraries, archives, museums, and other organizations today. Without good metadata, the digital resources would be sadly underutilized because most potential users would never discover their existence. Information professionals charged with organizing these collections need guidance. *Metadata for Digital Collections: A How-To-Do-It Manual* and its companion website (http://www.neal-schuman.com/metadata-digital-collections) introduce readers to fundamental concepts and practices in a style accessible to beginners and LIS students, as well as experienced practitioners with little formal metadata training.

Metadata for Digital Collections does not presume that readers come to the book with any cataloging experience. Instead, it guides students and working professionals through the basics of digital resource description, raising their awareness of commonly encountered challenges, along with common solutions, regardless of the specific metadata scheme being used, and it suffuses every chapter with numerous illustrative examples.

When turning to specific metadata schemes, rather than give a cursory survey overview of a large number of them, as many texts do, this book covers only three of those most commonly used for general digital resource description—Dublin Core, MODS, and VRA—allowing for each to be addressed in greater depth. Because Dublin Core is by far the most widely used metadata scheme for digital collections today, this book provides detailed, practical guidance on applying each of the Dublin Core elements and qualifiers, taking special care to clarify those most commonly misunderstood in order to assist readers in creating better quality, more functional, sharable, and interoperable metadata. The book uses MODS in large part for purposes of giving readers some practical familiarity with a hierarchical XML-based scheme, as a basis for informed comparison and contrast with Dublin Core, and for gaining hands-on experience with mapping between diverse element sets.

The book culminates in a step-by-step guide on how to design and document a metadata scheme or application profile for local institutional needs and digital collection projects. Unlike many other metadata texts, this book takes into account the widespread use of digital collection

management systems such as CONTENTdm. *Metadata for Digital Collections* also covers such topics as XML encoding, OAI harvesting, metadata sharing and aggregation, metadata quality control, and the emerging environment of Linked Data and the Semantic Web, explaining their relevance to current practitioners and students.

Two central themes run throughout the text: the primary theme of metadata *functionality* and secondary theme of metadata *interoperability*. Each chapter emphasizes that practitioners design and create metadata to perform specific *functions* for their users and that they also often need to merge metadata from diverse sources and make it *interoperate* and function together effectively for those users. The book covers the topics of resource description, application of standardized metadata elements and controlled vocabularies, and consistency in metadata creation all from the perspective of how they function to serve end users' information discovery needs and how they facilitate metadata interoperability for the same purposes.

The companion website (http://www.neal-schuman.com/metadata-digital-collections) includes review questions, ideas for exercises, and additional practical and reference sources useful for educators, students, and practitioners. Those who complete this book will be well equipped for engaging in concrete metadata work and entering the professional marketplace, as well as for learning additional metadata topics and schemes such as Encoded Archival Description (EAD).

This book is intended primarily, but not exclusively, for the following audiences:

- Practitioners and students who need a practical introduction to metadata for practical implementation and a detailed guide to applying Dublin Core in practice
- Practitioners in small to medium-sized libraries, museums, archives, and other institutions, rather than the largest, most well-funded research and academic institutions
- Practitioners who are short on time, staff, budget, programming expertise, professional reading, or formal metadata education
- Users of out-of-the-box digital collection software packages, such as CONTENTdm, Insight + LUNA, or Greenstone
- Students and instructors in schools of library and information studies and continuing education courses and workshops, as an introduction to the world of metadata practice experienced by these audiences

Organization and Scope

Metadata for Digital Collections is organized into 11 chapters that progressively build on one another in order to introduce fundamental concepts and practices. The book's design also facilitates independent chapter consultation by practitioners who need on-the-run guidance.

Chapter 1 introduces basic metadata concepts, definitions, functions, and types. This chapter, and indeed the entire book, emphasize the creation of metadata to perform *functions* for users of digital collections, such as searching, browsing, navigating, identifying, and interpreting digital texts, images, and other resources. Chapter 1 goes on to give an overview of the larger digital collection creation process, of which metadata is but one piece, followed by a brief overview of the process of designing a metadata scheme for local use.

The next section of the book explores the most common kinds of information needed to describe and provide access to digital resources and gives practical guidance on understanding and applying the Dublin Core Metadata Element Set. **Chapters 2**, **3**, and **4** accomplish this and dig into the nitty-gritty challenges faced by metadata designers and creators, addressing such common questions as these:

- What is a "digital object" or "resource," and what aspects of it should a good metadata record describe and represent?
- How do I balance the meaning of local elements devised for a specific collection with the meaning of standardized elements, such as Dublin Core, to which they are mapped?
- How do I deal with information about both the digitized and the original physical versions of a resource, when each has its own creator, date, identifier, and other characteristics?
- What does each of the 15 Dublin Core metadata elements mean, what qualifiers can be used with each, and how do I correctly apply them in practice?
- What is the meaning of, and difference between, the Dublin Core *Type* and *Format* elements and the *Relation* and *Source* elements?
- How should I devise titles for resources such as local photographs that have no pre-assigned titles?
- How do I analyze the subject content of a resource, such as an image, and represent it using metadata terms?

Chapter 2 introduces fundamental resource description concepts and issues encountered when creating metadata for digital collections. This chapter also introduces the Dublin Core Metadata Element Set, including simple (unqualified) as well as qualified Dublin Core (DC).

Metadata professionals face many practical challenges in the application of elements needed to address user needs and system functionality related to titles, identifiers, dates, languages, names, responsibility, and intellectual property. **Chapter 3** delves into these issues. After looking at general needs and practices, it goes into detail on how to apply the relevant Dublin Core elements for each of these aspects. **Chapter 4** continues this approach, but turns to more complex and challenging elements and practices, including resource types and formats, subject analysis and representation by means of subject terms and descriptions, and relationships among resources. After looking at each of these in

general, it details how to apply the relevant Dublin Core elements for each.

Controlled vocabularies improve resource discovery for users. **Chapter 5** provides an overview of different types of vocabularies, such as lists, taxonomies, and thesauri. It examines some of the most commonly used established vocabularies, as well as the creation of an institution's own local vocabularies.

Chapter 6 provides a very simple introduction to the basics of XML, focusing on those aspects needed to "read" and understand an XML-based metadata record. It includes examples of Dublin Core and MODS XML records, and concludes with a guide to the anatomy of an XML metadata record.

MODS, the Metadata Object Description Schema, is the subject of **Chapter 7**. This chapter gives an overview of the MODS elements, subelements, and attributes; illustrative examples of MODS records; and issues in mapping from Dublin Core to MODS. The chapter makes points about the value of learning something about MODS even if not using it in practice. Studying MODS provides an opportunity to compare Dublin Core with a more complex, XML-based general resource description scheme, and to directly experience the complexities of mapping between different element sets, among other uses.

Chapter 8 surveys the Visual Resources Association (VRA) Core Categories for works of art and architecture, including overviews of both the relatively DC-like VRA 3.0 and the relatively MODS-like VRA 4.0, with record examples of each. VRA is covered in much less detail than DC or MODS and is included primarily for purposes of further comparison and contrast with those two schemes.

In the current and future metadata environments, practitioners need to be concerned about the usability of their local metadata outside of its original context. **Chapter 9** investigates a set of interrelated topics having to do with interoperability, including viability of metadata for future system migration; sharing metadata within an institution or with a consortium or a third-party aggregator; issues of metadata harvesting, especially the use of the OAI harvesting protocol; metadata ingestion, processing, and conversion; crosswalks and mapping among different element sets; and metadata quality indicators and assessment methods. The chapter concludes with five concrete practices that readers can follow to improve their metadata quality and interoperability.

Metadata project managers often need not only to create metadata, but also to design and document their own local metadata schemes or application profiles. **Chapter 10** details a step-by-step process of assessing the context, content, and users of the collection, developing a set of functional requirements, selecting or creating a set of metadata elements and determining the element and database field specifications to meet those functional requirements, as well as examples of, and best practices for, documentation of one's scheme. The chapter looks at two basic models of metadata design: (a) selecting and adapting an established scheme such as Dublin Core to serve for multiple collections within an institution or consortium, and (b) creating collection-specific elements

and mapping them to an established scheme such as Dublin Core (the typical CONTENTdm method).

Chapter 11 serves as a beginner's-level, step-by-step introduction to the *Resource Description Framework* and other aspects of metadata in the context of the currently developing *Linked Data* and *Semantic Web* environments. This includes the formal registration of metadata elements on the Internet using URIs, some basics of the Resource Description Framework (RDF), selected concepts from the DCMI Abstract Model, and recent developments in formalized Dublin Core Application Profiles. The working assumption is that most readers of the book will not be directly working with metadata in this context, but that some familiarity with these topics is valuable both for current awareness of new developments in the field and for possible future directions in which many current practitioners may eventually work.

Not all of the information in all chapters will be relevant to all practitioners. For example, the chapters on MODS and VRA may be of little interest to practitioners using only Dublin Core. But a study of a hierarchical, XML-based scheme such as MODS has great value for better understanding a simpler, flatter scheme such as DC for getting a better sense of the strengths and limitations of DC, for better understanding hands-on issues of mapping from one scheme to another, and for gaining insight into the types of metadata schemes that could possibly supersede DC in prominence for digital collections in the future. Learning some basics about topics such as XML, interoperability, harvesting, and aggregating will help broaden and deepen metadata practitioners' knowledge of their field of practice and might also suggest unforeseen practical applications.

One aspect of the book's organization deserves special note. When creating a digital collection, the first step is to design a metadata scheme or application profile. Yet *Metadata for Digital Collections* covers this topic in the second-to-last chapter rather than in the second chapter. Experience has shown that, in order to design a well-developed and effective metadata scheme, the designer needs a solid foundational knowledge of resource description and controlled vocabulary issues, the meaning and application of the standard scheme (such as Dublin Core, MODS, or VRA) selected as the basis for the local scheme, and some familiarity with issues of interoperability, harvesting, and mapping for metadata shareability and long-term usability.

Metadata for Digital Collections provides a practice-oriented approach to learning about and applying metadata based on the author's many years of practical experience and of teaching both students and working professionals. Readers will come away with a solid working knowledge of metadata for digital resources that they can put to use in their jobs or take with them into today's professional marketplace.