



METADATA FOR DIGITAL COLLECTIONS: A HOW-TO-DO-IT MANUAL, SECOND EDITION

By Steven Jack Miller

ALA Neal-Schuman, 2022

505 pages, 212 figures and tables

Softcover: \$69.99, ISBN 978-0-8389-4748-7

Review by: Kate Thornhill (she/her), University of Oregon

THIS HOW-TO TEXTBOOK published by the American Library Association's Neal-Schuman imprint is the much anticipated second edition of a volume originally released in 2011. This review will focus on the new edition's scope and arrangement, and how it represents today's metadata best practices. Although this manual is primarily geared toward developers and managers of cultural heritage digital collections, its approaches to data design and interoperability are also relevant to those whose work intersects this field. That could be cartographers seeking foundational knowledge of, or advice about, object-oriented cultural projects and other digital collections stewarded by galleries, libraries, archives, and museums (GLAMs). It also includes anyone with little formal data science training, including beginner data creators and those who digitize print maps, who wants to learn how to describe and structure data so that it's accessible to others, and are looking to learn some significant good data hygiene practices, too.

The book's author, Steven Jack Miller, is an emeritus faculty member at the University of Wisconsin–Madison. He is an expert in information knowledge and organization, resource description and access, and linked data ontologies. He was a metadata and cataloging librarian from 1991 to 2006, when he led library materials acquisitions and cataloging, and he has also taught graduate courses in library and information sciences. Miller's professional and pedagogical experiences have made him an authority on a wide range of metadata-related topics. His extensive teaching experience has also heavily influenced his approach to

metadata—treating it as *data* instead of *annotation*—and it is this approach that makes his text so valuable, accessible, and applicable.

The over five hundred pages of *Metadata for Digital Collections* are divided into twelve chapters that can be followed step-by-step or cherry-picked by readers, depending on the need or predilection of the user and the particulars of the collection and data retrieval system in use. Each chapter starts with a brief introduction, followed by in-depth explanations, real-world examples, and a summary with references. The book is designed to provide a structured learning experience, building on foundational metadata practices established in the 1990s and 2000s. As the author, Steven Jack Miller, explains in the preface, every chapter has been reworked to some extent, but perhaps most noticeable are the structural changes apparent when comparing this second edition's table of contents to the earlier book's. Chapters have been reordered, and both new and updated material added to reflect recent developments in the cultural heritage field, such as the Linked Data movement.

Chapter 1, "Introduction to Metadata for Digital Collections," showcases the ups and downs typically faced by metadata designers and covers the rationale for constructing cultural heritage digital collections in ways that allow users to browse, search, navigate, identify, and interpret. It also gives a very brief overview of how digital collections are planned, developed, digitized, and hosted. The



© by the author(s). This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

purpose and functional distinctions between the various types of metadata are illustrated, as are the different metadata topologies, such as structural, content, and data value. Data encoding and exchange standards are touched upon as well. Again, it is important to note that Miller’s metadata-is-data approach to framing these concepts supports the designers of data-driven projects who are intent on using digital collection platforms like ContentDM (oclc.org/contentdm), Samvera (samvera.org), CollectionBuilder (collectionbuilder.github.io), Islandora (islandora.ca), and Omeka (omeka.org).

Chapter 2, “Introduction to Resource Description,” is about how to approach metadata creation, cataloging, indexing, and resource description, and it lays out foundational concepts and practices that GLAM digital collection metadata managers utilize to communicate and implement interoperable data. It demystifies terms like *description*, *digital objects*, *resources*, and others typically used by cultural heritage data workers, and sorts out what is involved in choosing between, say, collection-level and item-level; metadata and the data makeup for simple, compound, or complex digital objects; element repeatability and element functionality; or what constitutes a property-value pair. It also expands on strategies for dealing with content and carrier descriptions for digitized cultural objects.

Chapter 3 focuses on one of the most important, shared, and internationally used metadata standards, Dublin Core. It provides a general yet comprehensive overview of the standard, a description of how to apply the Dublin Core schema when designing a project, and instructions for how readers can align project-specific metadata elements or non-Dublin Core elements to the scope covered by either Simple or Qualified Dublin Core.

Chapter 4, “Resource Description: Identification and Responsibility,” and Chapter 5, “Resource Description: Content and Relationship Elements,” continue and extend discussions opened in the previous chapters. Both investigate and explain what information is commonly needed to give access to, and to describe, digital cultural objects, and how to make that information interoperable and retrievable from digital collection-oriented software systems. These two chapters also dive deep into the nuances of working with intellectual and artistic content descriptions such as resource type, genre, and subject. Two commonly used controlled vocabularies, MODS (loc.gov/standards/mods)

and VRA Core (core.vraweb.org), are introduced here, although each schema is covered in detail in Chapter 8, “MODS: The Metadata Object Description Schema,” and Chapter 9, “VRA Core: The Visual Resource Association Core Categories.”

Chapter 6, “Controlled Vocabularies for Improved Resource Discovery” describes how resource interpretation through categorical groups is affected in a metadata system. Such a system, like a database, needs a way to allow users to search and browse objects, and it is through controlled lists of terms and controlled vocabulary topologies that this is provided. The chapter describes different controlled vocabulary topologies and established controlled vocabularies typically used by GLAMs, as well as where to find human-readable linked open data vocabularies published by the Library of Congress.

Chapter 7, “XML-Encoded Metadata,” Chapter 8, “MODS: The Metadata Object Description Schema,” and Chapter 9, “VRA Core: The Visual Resource Association Core Categories,” are devoted to other technical topics and how-tos about encoding metadata. They offer introductions to XML and to using the MODS, Dublin Core, and VRA Core technical standards to markup resource descriptions. Similarly, Chapter 10, “Metadata Interoperability, Shareability, and Quality,” also emphasizes how the terms in the chapter’s title are essential for data exchanges and sharing, and how they facilitate the efforts of data providers and services to bring digital objects together for better discovery, use, reuse, and preservation opportunities. This chapter also highlights the concept of data quality and introduces the commonly used data cleanup and remediation software, Open Refine (openrefine.org).

Chapter 11, “Linked Data and Ontologies,” demystifies and explains linked data and ontologies—not just with textual explanation, but also lots of data modeling illustrations as well. Miller systematically breaks down the many aspects of how linked data and ontologies are configured and structured, while emphasizing how this scaffolding backstops the digital collections that exist in the foreground. In a lovely foundational and technical way, this chapter brings home the importance of what was covered in all the previous chapters about metadata schema, resource descriptions, controlled vocabularies, and XML markup.

The twelfth and final chapter “Metadata Application Profile Design,” provides a step-by-step process for assessing and documenting (in useful and usable guides) all aspects of user and digital collection system communication. Miller presents multiple examples of these types of guides from the University of Washington and the University of Wisconsin–Milwaukee. The system employed by the American Geographic Society Library and Digital Collection, which contains a browsable world mapping feature, is prominently highlighted.

What stands out most about this book is Miller’s ability to explain highly technical data concepts in straightforward and illustrative ways, and it is evident that he has a great deal of experience in systematically explaining their application. The technical nature of the text will leave readers with a solid, practical knowledge base for working with cultural heritage digital collections, and a real appreciation for Miller’s understanding that working with metadata for digital collections is not just a science but an art. Even though the book does not target digital mapmakers, cartographers seeking to use historical and contemporary cultural materials while building or adding to databases, or while creating interactive data-oriented maps, will appreciate this how-to publication’s step-by-step and reference nature.

Although this book focus on digital collections and standards deployed by library and information scientists, it prepares all researchers in the foundations of data curation and provides a conversance with how metadata is made usable and shareable. The understanding it provides about crucial concepts, like resource description and encoding data with interoperable standards, positions researchers to make reproducible and shareable projects. Those working in interdisciplinary teams striving to tell geospatial stories with cultural materials, or with research objects captured in the field, will find the educational tone, approach, and structure taken in this book convivial for learning. The author’s ability to shine light on the meaning and use of technical terms and methods gives project teams a common vocabulary and method for planning how data is to be collected, described, structured for interoperability and sharing, and reviewed for quality, all of which can ensure that their data is responsibly documented and cataloged.

Although the second edition of *Metadata for Digital Collections* is a solid technical manual, there remain some areas where its coverage could have been elaborated. Its

illustrations and examples are drawn almost entirely from academic libraries supporting digital collections, something that makes the book feel less open to, or appropriate for, other types of GLAM workers or people from other disciplines. The book could also be improved by highlighting current trends and criticisms about resource descriptions that are coming from certain GLAM communities like libraries and archives. These trends and criticisms are relevant to any data practitioner concerned with, or curious about, the way cultural, institutional, and societal oppressions impact metadata creation and reuse. For example, there is little to no discussion about the invisibility of metadata creators’ labor—although it is briefly mentioned in Chapters 1 and 2 when Miller writes about creating a digital collection and the need for research when describing objects. It is surprising to not see even a passing reference to the Digital Library Federation’s Digitization Cost Calculator (dashboard.diglib.org), which lets anyone estimate how much time and money a project to catalog digital objects would require. Nor does Miller reference recent thought pieces or scholarship such as Stacie Williams’s 2016 “Implications of Archival Labor” (medium.com/on-archivy/implications-of-archival-labor-b606d8d02014), which sparked critical conversations in the digital libraries community about who supports resource discovery and does archival labor.

Beyond his mention of efforts by the Library of Congress (using Flickr) over 15 years ago, Miller gives little to no attention to contemporary discussions about the implications of resource descriptions written by white heteronormative metadata practitioners who work at predominantly white institutions like academic libraries. For example, when discussing how metadata practitioners value end-user contributions, Miller might have also highlighted the work of Archives for Black Lives in Philadelphia’s Anti-Racist Description Resources (github.com/a4blip/A4BLiP); Violet Fox, who started the Cataloging Lab (cataloginglab.org) in 2018; and Emily Drabinski’s *Queering the Catalog: Queer Theory and the Politics of Correction* (doi.org/10.1086/669547). These people and organizations have been instrumental over the past ten years in trying to make metadata work more just, diverse, and inclusive so users feel seen and understood from their cultural and social contexts.

In conclusion, *Metadata for Digital Collections* is an invaluable resource for anyone interested in working with digital collections. Miller’s expertise in metadata management

and his ability to explain complex technical concepts in a clear and concise manner make this book a valuable asset to librarians, archivists, and other information professionals. While the book's focus is on digital collections and standards used by library and information scientists, its foundational principles of data curation and the importance of making metadata usable and shareable can be applied across disciplines. While there may be some areas

where the book's coverage could be elaborated, Miller's focus on sound management practices and his practical approach makes this a must-read for anyone involved in digital collection management. Overall, *Metadata for Digital Collections* is an excellent resource that provides a solid technical foundation for working with cultural heritage digital collections and an appreciation for the art of metadata management.

