

## Introduction

*Murtha Baca*

Like metadata itself, the realm of online resources is constantly and rapidly evolving. Much has changed in the digital information landscape since the first print edition of this book was published in 1998 and the revised online version appeared in 2000. The time is right for an updated edition of this text, intended to give a general introduction to metadata and to explain some of the key tools, concepts, and issues associated with using metadata to build authoritative, reliable, and useful digital resources.

Metadata creation is—or should often be—a collaborative effort, as is this book. For this edition, the three contributors to the 2000 version wrote updated chapters, and I was fortunate to find a new contributor to address the crucial issue of rights metadata.

In the first chapter, Anne Gilliland provides an overview of metadata—its types, roles, and characteristics—as well as facts about metadata that belie several common misconceptions. She also addresses current trends in metadata, especially that of metadata created by users rather than trained information professionals. Activities such as social tagging, social bookmarking, and the resulting forms of user-created metadata such as “folksonomies” are playing an increasingly important role in the realm of digital information.

In the second chapter, Tony Gill discusses metadata as it relates to resources on the Web. He explains how Web search engines work and how they use metadata, data, links, and relevance ranking to help users find what they are seeking and discusses in detail the commercial search engine that as of this writing has dominated the Web for several years: Google. He explains the difference between the Visible Web and the Hidden Web and the important implications and issues relating to making resources reachable from commercial, publicly available search engines versus systems that have one or more “barriers” to access—because they are fee based or password protected or require a particular IP address, or simply because they are not technically exposed to commercial search engines. Gill also raises issues relating to open access to digitized materials and legal obstacles that currently prevent open access to many materials.

In the third chapter, Mary Woodley examines the methods, tools, standards, and protocols that can be used to publish and disseminate digital collections in a variety of online venues. She shows how “seamless searching”—integrated access to a variety of resources residing in different information systems and formulated according to a range of standard and nonstandard metadata schemes—is still far from a reality. Woodley contrasts the method of “federation” by means of the building of union catalogs of digital collections by aggregating metadata records from diverse contributors into a single database with metasearching—real-time searching of diverse resources that have not been aggregated but rather are searched in situ by means of one or more protocols. Each method requires specific skills and knowledge; particular procedures, protocols, and data standards; and the appropriate technical infrastructure. Creating union resources via physical aggregation of metadata records or via metadata harvesting is a good thing, but we should keep in mind that it does not necessarily solve the Hidden Web problem enunciated by Gill. If resources are publicly available but users cannot reach them from Google, instead having to find the specific search page for the particular union resource, we cannot say that we have provided unfettered access to that resource. Woodley also stresses the importance of data value standards—controlled vocabularies, thesauri, lists of terms and names, and folksonomies—for enhancing end-user access. She points out that mapping of metadata elements alone is not sufficient to connect all users with what they seek; the data values, that is, the vocabularies used to populate those elements, should also be mapped.

Maureen Whalen’s new chapter, “Rights Metadata Made Simple,” argues that the research and capture of standards-based rights metadata should be core activities of memory institutions and offers practical, realistic options for determining and recording core rights metadata. If institutions would commit the effort and resources to following Whalen’s advice, many of the legal obstacles mentioned by Gill in his discussion of libraries and the Web could be surmounted.

In another new section in this edition, “Practical Principles for Metadata Creation and Maintenance,” we again emphasize that institutions need to change old paradigms and procedures. They need to make a lasting commitment to creating and continually updating the various types of core metadata relating to their collections and the digital surrogates of collection materials that we all seem to be in such a hurry to create.

Our slim volume concludes with a glossary and a selected bibliography. The glossary is not intended to be comprehensive; rather, its purpose is to explain the key concepts and tools discussed in this book. The bibliography, too, is deliberately restricted to a few relevant publications and resources. The footnotes in each of the chapters provide

numerous additional references to publications and online resources relevant to the topic of metadata and digital libraries.

At the end of her chapter, Gilliland compares metadata to an investment that, if wisely managed, can deliver a significant return on intellectual capital. I would venture to expand on her financial metaphor and say that metadata is one of our most important assets. Hardware and software come and go—sometimes becoming obsolete with alarming rapidity—but high-quality, standards-based, system-independent metadata can be used, reused, migrated, and disseminated in any number of ways, even in ways that we cannot anticipate at this moment.

Digitization does not equal access. The mere act of creating digital copies of collection materials does not make those materials findable, understandable, or utilizable to our ever-expanding audience of online users. But digitization combined with the creation of carefully crafted metadata can significantly enhance end-user access; and our users are the primary reason that we create digital resources.