Jeffrey Pomerantz, **Metadata**, Cambridge, MA: MIT Press, 2015, 256 pp., $14.60 (paperback).

Reviewed by
Marcia Mardis
Florida State University, USA

If you've ever located a book on a library shelf by the call number from a library's online catalog, selected a periodical article by reviewing database search results, or bought a car from an online retailer based on its description, you've used metadata. Although it is largely invisible or in the background, as Jeffrey Pomerantz points out, metadata is more than "data about data"; it is a representation, or "map," of a resource, standing in for the actual resource itself. Metadata is all around us and, as Pomerantz proposes in chapter 1, "It's metadata's world and we're just living in it" (p. 4).
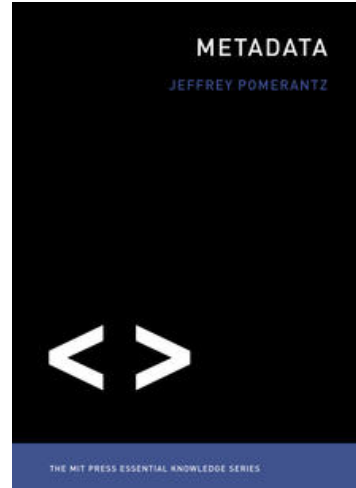
Through a thorough overview of metadata definitions, types, functions, and applications, Pomerantz aims to provide readers with an explanation of "what metadata is and why it exists" (back jacket). Pomerantz, currently a senior researcher with the EDUCAUSE Center for Analysis and Research (ECAR), is well positioned to undertake this task. At the time of the book's writing, Pomerantz was an associate professor at the University of North Carolina at Chapel Hill's School of Information and Library Science and a visiting professor at the University of Washington's Information School. I became aware of Pomerantz's work in metadata almost 20 years ago when he was a doctoral student at Syracuse University's School of Information Studies. There he worked closely with faculty who were pioneers in the areas of educational metadata and digital libraries and contributed some of the definitions and standards that information scientists and practitioners use today. I believe that Pomerantz' long view of metadata and its role in helping information users reconceptualize libraries as digital institutions was an important underpinning of the accessible explanations and clear conceptual orientation of this work.

The book's content and presentation are adapted from a well-attended (and highly completed) Massive Online Open Course (MOOC) that Pomerantz taught through MIT Open Courseware. The course's success led The MIT Press to approach Pomerantz about distilling it into a book. In chapter 1, Pomerantz draws the reader in by recalling the events surrounding Edward Snowden's 2013 disclosure of the National Security Agency's (NSA) cell phone surveillance operation. You may recall that the NSA justified its data collection operations because the agency was not recording phone calls, only capturing call metadata such as initiating and receiving phone numbers, call duration, and call location. These descriptive elements set the context for Pomerantz' chapter 1, in which he introduces the existence, ubiquity, utility, and invisibility of metadata. This chapter is anchored by large pull quotes that highlight key concepts, a useful device that is employed throughout the book. Chapter 2, "Definitions," provides, as one might expect, in-depth definitions and explanations of key concepts that are developed further in the remainder of the book. Chapters 3, 4, and 5 address descriptive metadata, administrative metadata, and use metadata (in that order) comparatively and as individual concepts. Readers learn that descriptive metadata reflects what the

resource is while administrative metadata captures information about the resource's creation, versioning, and life cycle. Use metadata includes what users have done with the resource, such as when the resource was last downloaded. While descriptive and administrative metadata are governed by the various standards and vocabularies, Pomerantz positions use metadata as an emerging set of ideas that will enable users to identify connections between resources and assess their appeal and vitality.

In chapter 6, "Enabling Technologies for Metadata," and chapter 7, "The Semantic Web," Pomerantz provides a high-level overview of advanced concepts in metadata practices and use, such as the Resource Description Framework and linked data. He wisely does not try to provide exhaustive explanations of these sophisticated and emerging areas, but instead defines them, describes their importance, and provides readers with additional authoritative resources. The final chapter of the book is a futurecast for metadata in which Pomerantz foreshadows the growth of discipline-specific metadata; here he refers to "Darwin Core," the biodiversity community's riff on the established and widely used Dublin Core descriptive scheme. He also points to application programming interfaces as enabling metadata to be embedded in a variety of online and mobile contexts and eScience as adapting metadata approaches to the management of big data and computational endeavors. Again, Pomerantz introduces the concepts at an appropriately high level and directs the reader elsewhere for more information.

A major strength of this book is that it is derived from a successful course; its structure makes sense and it builds conceptually. I could almost imagine each chapter as a learning module accompanied by a suite of explanatory lectures. This flow is logical and easy to follow. With an educator's frame of mind, Pomerantz frequently uses relatable examples to illustrate new ideas. He references the Simchat Torah Jewish holiday, a celebration to mark having reached the end of reading the Torah, as an example of the drudging linearity inherent in physical book (or, in this instance, scroll) reading that online searching liberates. His use of Goodreads community members' subject tags of as an example of uncontrolled vocabulary is also an especially relatable example. In a couple of instances, the examples seem a bit incomplete or obscure for someone new to a concept, such as "If you've ever lost your keys . . . you understand the usefulness of metadata" (p. 13). How metadata fits here is never explained. Pomerantz also provides useful operational definitions for terms such as "data," defined as "potentially *[sic]* information" and a "resource" as a "potentially informative object" (p. 21, p. 26). These contentions are complemented by embedded tutorials like the brief but effective introduction to HTML provided in chapter 3 as a way to understand the function of different metadata fields.

The book does an excellent job of providing a solid, understandable, and thorough introduction to metadata and its applications. It also leaves open directions for further thought and investigation. For example, the question of how metadata is used beyond enhancing resource findability is explored in a cursory but provocative manner. Pomerantz leads the reader to, but does not tackle, ideas such as our time of burgeoning usage analysis; readers will naturally wonder whether and how metadata functions to enable large-scale data analytics. A related topic Pomerantz raises but intentionally does not attempt to thoroughly address is that most Web users do not question their searching beyond their satisfaction with the results; they may not realize that their search is not always being done on the full text of a resource and is often just done on the metadata for a resource. At a time when video and other multimedia resources are outpacing text resources on the Web, searches done on just metadata are increasingly the

norm. When metadata is driving the search experience, metadata creators have a tremendous amount of influence over what users find simply by controlling what is findable and the vocabularies they use for fields such as subject classification. Description control is access control, as Foucault (1970, 1972) and others (e.g., Bowker & Star, 1999; Rosenbaum, 2011) have posited, and understanding the motivations of the people and organizations tasked with creating descriptions is vital to understanding how access to the resource is shaped by its description. Metadata is only purported to be, but cannot possibly in actuality be, as Pomerantz argues in the introduction (pp. 16–17), completely objective.

A second question arising from this book's content relates to who creates the metadata. While Pomerantz does introduce the concept of provenance metadata to document a record's chain of custody, those metadata shed no insight into the person who created the record in the first place. Some metadata are a combination of automatically and human classified fields, but thus far automatically generated metadata has been a bit of a nonstarter. To address barriers to scaling collections, automatic generation provides, high quality metadata is usually mediated with some human involvement.

Pomerantz also introduces the idea of the "data exhaust" generated by social media, and this is an area ripe for further development and scholarship. Although the data exhaust discussion is entry level, it is detailed enough to spur readers to explore ideas such as whether Twitter hashtags function as controlled vocabulary, uncontrolled vocabulary, or something new entirely new. Related to social media applications, Pomerantz passed on providing a detailed discussion but opened an opportunity for readers to apply the book's concepts to consider how undercontributed data such as ratings, reviews, shares, and re-posts articulate with metadata. For instructors who may be using this book as a course text, Pomerantz leaves room for student questions, exploration, and complementary knowledge construction.

A notable omission in the book is a discussion of resource granularity and its relationship to metadata. That is, the size, or "grain," of a resource has a relationship to how it is described. For example, a whole video will be described at a different level of detail than a chapter or single frame of the video. As online personal content curation tools such as Pinterest or Pearltrees proliferate, the extent to which users are able to locate, isolate, and share or save a single item from a website instead of a link to an entire website is important for customizing user experience, enabling clear communication, and promoting resource remixing and customization.

I encountered other issues in the book, such as the lack of figure labeling. Because the figures are usually not explained in the chapter text and most (except those in chapter 6), lack captions, it is sometimes unclear what a figure means or why the figure is placed in a particular location. For example, figure 7 looks very valuable, but is not accompanied by a caption or an explanation, and figure 16 seems to bear no relationship to the text surrounding it.

Even with these minor drawbacks, this book is a necessary, unique, and accessible introduction to metadata for college students and faculty, new information professionals, or anyone interested in learning a bit more about an influential yet invisible driver of the information we access and share. Because metadata is "the map, not the territory," or a representation of instead of an actual resource,

understanding how metadata is created, used, and will grow is vital for information consumers, creators, and curators.

References

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Foucault, M. (1970). *The order of things*. New York, NY: Vintage Books.

Foucault, M. (1972). *The archaelogy of knowledge and the discourse of language*. New York, NY: Pantheon Books.

Rosenbaum, S. (2011). *Curation nation: Why the future of content is context*. New York, NY: McGraw-Hill.