Jeffrey Pomerantz, **Metadata,** Cambridge, MA: MIT Press, 2015, 256 pp., $16.95 (paperback).

Reviewed by
Indrek Ibrus[1]
Tallinn University, Estonia

This review is about a book that is not immediately part of the media and communications academic domain. Author Jeffrey Pomerantz is an information scientist who studies metadata as an object of interest generally, within the information science domain. Yet the implications, the phenomena, and the developments the book covers have important insights for media and communications studies. As such, Jeffrey Pomerantz's **Metadata** can be viewed as building on a long tradition within media studies of discussing how the technical framing of media affects its operation, its affordances, and its effects. First was Marshall McLuhan (1964) with his thesis that the medium makes the message. More recently, the entire software studies domain (Fuller, 2008) has emerged to interpret the cultural effects of software. As Lev Manovich (2014) put it his article "Software is the Message." In parallel, both in humanities as well as in social sciences–driven media studies, significant bodies of work have emerged on data studies—how (big) data processing in media use could affect social and cultural dynamics. In this context, Pomerantz's introductory book reminds us that we need to become more nuanced when talking about such data—much of that data is effectively metadata, with several dynamically evolving subsets. Furthermore, the specific logic of metadata affects many walks of life but also, significantly, media and communications. That is why, although this book is not written within the domain of media and communications studies, it is highly relevant to the field: It shows how metadata pervades all communications protocols and all automated processing of media. As Pomerantz puts it, we have entered the era of uniquitous metadata (p. 3). Metadata is the message, if I may put it a bit more provocatively.

This early conclusion—that the book can be understood as provocative, if not at least inspiring—is significant, given that it is published in the Essential Knowledge series of the MIT Press. The rationale behind the series is to offer concise and accessible overviews of compelling contemporary topics. This book does all that. It is written in an accessible style, with new concepts specific to the field presented and discussed progressively—starting with more abstract and general concepts and gradually introducing more specific and complex issues, such as the "Semantic Web" and APIs (application programming interfaces). The book often returns to issues and examples introduced earlier, but from a new angle. The many examples used to illustrate complex issues are helpful, as is the glossary at the end of the book. Reading **Metadata** was, therefore, a quick but informative journey into some of the most essential and contemporary issues regarding metadata, its subsets, its uses and functionalities, and its related future challenges (especially in the context of the evolving Semantic Web).
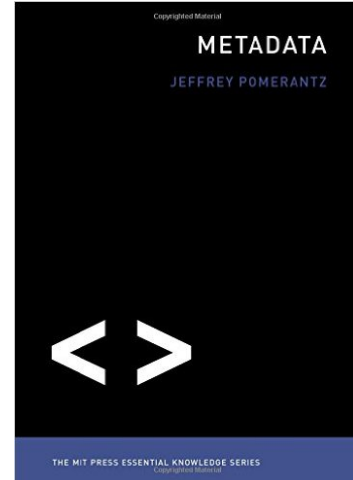
As such, the book is appropriate for anyone at any academic level who is interested in the workings of metadata. It would be useful for undergraduate studies, as well as for readers with experience working with metadata from a media and cultural studies perspective. It definitely helped to systematize some of my knowledge on the subject.

What made this guided tour especially appealing for a media scholar, however, was its interdisciplinary approach. Pomerantz builds not only on information science but also semiotics and cultural theory, some media theory, and, of course, computer science. The book establishes connections between the disciplinary perspectives, for example, when discussing metadata as a "language" in semiotic terms or when, at the end of the book, briefly discussing the politics of metadata. These links establish the relevance of the subject in relation to other disciplines and issues.

If the book has a core message, it is the nature and quality of the metadata necessary to build a universally functional Semantic Web of the future. Toward the end of the book, Pomerantz demonstrates how the Semantic Web comes together effectively as a network of subject–predicate–object triples in which each subject can be an object (the "resource," what is described by metadata) in another triple, and in which each subject/object has its own unique identifier—a URI (Uniform Research Identifier), or Web location. He then concludes that to achieve the vision of the Semantic Web once laid out by Tim Berners-Lee and his colleagues would require relying on structured data and quality metadata. All that data, he emphasizes, need to adhere to universally accepted and shared standards, as standards fragmentation would undermine the evolution of the Semantic Web.

In this context, what the book is, perhaps, missing are the contemporary "standards wars," especially the constitutive standoffs between proprietary and open standards. In a media studies context, we could, for instance, use audiovisual content metadata as an example. While working within the broader Semantic Web framework, the World Wide Web Consortium (W3C) created the Media Annotations Working Group and has worked with the European Broadcasting Union (EBU; unites Europe's public service media institutions). The EBU had previously developed the EBU Core, a metadata set for audiovisual resources, with the W3C later building its media annotation ontology broadly on the EBU model. However, what matters here is that the EBU, as a union of public service organizations, is designed to prefer an open standard. The W3C has a similar rationale—its perspective is that the Web can only evolve as a set of consensually accepted open standards (Ibrus, 2013). Yet next to these ambitions exist various commercial service providers, like Netflix or YouTube, that use metadata with other goals in mind (e.g., increasing customer loyalty) and thus use their own custom-designed ways to annotate content, resulting in preferred proprietary standards. The lack of transparency in Netflix's operations, for example, has made the company infamous in its unwillingness to cooperate with regulators or other public service institutions, including consensual standardization activities. Instead, its metadata schemas are a guarded secret (see Madrigal, 2014). Yet, being an increasingly dominant global service, they "gatekeep" much of the access to audiovisual culture; therefore, its specific metadata system not only starts to shape contemporary global cultural dynamics, but, as one of the largest catalog holders, its impact on the diffusion of open standards (needed for the success of the Semantic Web) would be significant. It is for this reason that the standards wars matter in this domain.

Regarding the further importance of metadata studies to media and cultural studies, this can be explained by elaborating on Pomerantz's definition of *metadata*: statements about a potentially informative object (i.e., statements about data as raw, unprocessed information that may or may not be informative to its receiver). If we think about this in Foucauldian terms (Foucault, 2002), a system of statements makes up a discourse, and what we are talking about is effectively a discourse on what and under what conditions it may be informative to members of society. Pomerantz also points out that metadata, especially in the case of "controlled vocabularies," effectively constitute subjective languages. Continuing with Foucault's archaeological method, the metadata analysis could, therefore, also start asking how those statements were created, what can and cannot be said, and how spaces are created in which new statements can be made. But specific to metadata and search algorithms are these questions: How does metadata channel-access to resources, and how do constellations of metadata in aggregate model culture?

That is, as we are entering the "metadata era," it becomes crucial to start investigating the modeling and ideological effects of all kinds of metadata and accessory technologies. This holds whether we search for cultural materials from online databases or catalogs, but it also holds for the wider Semantic Web—if the Semantic Web is about structured data and metadata, then it will soon hold for all of the Web.

In this context, as Pomerantz used semiotics to define metadata as language, I would also suggest other concepts within semiotics, for instance, that of "modeling systems," initially defined by Juri Lotman (1967/2011) and later developed by others, especially Thomas Sebeok and Marcel Danesi (2000). The networks and ontologies of metadata form models of data as expected to represent the world. As such, metadata is an exceptional form of "secondary" modeling systems—the one organizing and filtering all other modeling systems. Its emergent role in cultural dynamics needs thorough investigation.

My last brief comment regarding the future developments that this book introduces is the blurred distinction among descriptive, administrative, and use metadata. Pomerantz highlights the relatively new term "data exhaust" to refer to data on the usage of digital resources that remain linked to a resource. He also suggests that it would be good to keep usage data separate from other, intentionally created metadata. Yet he does not sound optimistic about this prospect. The fact that a lot of contemporary social scientific data studies are concerned with the latter indicates that, increasingly, the different forms of metadata are perceived as mixed. Algorithms can use usage data to understand or contextualize the contents and, thereafter, also administer it, for example, to recommend to specific users, or not. As Wolfgang Ernst (2013) put it, we live in an era of dynamic archives—the usage of content generates metadata that recontextualizes resources, and thus all triples in semantic networks also can be updated and modified in the process—and the Semantic Web is dynamic due to the data exhaust.

All in all, we live in an increasingly dynamic era of metadata, and Jeffrey Pomerantz's introduction to it is a good way to learn about its inner mechanics.

References

Ernst, W. (2013). *Digital memory and the archive* (J. Parikka, Ed.). Minneapolis, MN: University of Minnesota Press.

Foucault, M. (2002). *The archaeology of knowledge*. London, UK: Routledge.

Fuller, M. (2008). *Software studies: A lexicon*. Cambridge, MA: MIT Press.

Ibrus, I. (2013). Evolutionary dynamics of media convergence: Early mobile Web and its standardisation at W3C. *Telematics and Informatics, 30*(2), 66–73.

Lotman, J. (2011). The place of art among other modelling systems. *Sign Systems Studies, 39*(2/4), 249–270. (Original work published 1967)

Madrigal, A. C. (2014, January 2). How Netflix reverse engineered Hollywood. *The Atlantic*, Retrieved from https://www.theatlantic.com/technology/archive/2014/01/how-netflix-reverse-engineered-hollywood/282679/

Manovich, L. (2014). Software is the message. *Journal of Visual Culture*, *13*(1), 79..

McLuhan, M. (1964). *Understanding media: The extensions of man*. New York, NY: Mentor.

Sebeok, T. A., & Danesi, M. (2000). *The forms of meaning: Modeling systems theory and semiotic analysis*. Berlin, Germany: Mouton de Gruyter.