

## Metadata: Transforming Complexity to Plenty

The picture of metadata today is one of complexity, variety, and apparent overload. And this makes sense because metadata, as the markers of information, should mirror the characteristics of the information marked. Because of the Internet and the Web, the quantity of information is overwhelming. Is this complexity bad? No. It can be considered rich, a cornucopia instead of a deluge. However, the way to ensure that the amount of information in existence *is* a cornucopia is to use metadata to make specific information findable. This, of course, is the goal that libraries, traditional information retrieval systems, and search engines have always had. The view from 1998, saw a prominent and vital place for libraries in this endeavor: “[Metadata] is obviously an area that institutions will be expecting librarians to jump into; it’s something librarians have been doing for centuries” (Dave Hartland, 1998, cited in Anonymous, 1998, p. 150). The view from roughly a decade later encourages the equality of both professional, top-down, hierarchical, controlled metadata and crowd-sourced, bottom-up, faceted, uncontrolled metadata (Avery, 2010; Morville, 2005; Bengston, 2012; Weinberger, 2005; Wright, 2007).

### **Why is metadata valuable and what is it exactly?**

The value of metadata lies in its ability to mark information in a way that allows the information to be organized and retrievable for a user (Bengston, 2012; Dempsey & Heery, 1998; Morville, 2005; Rathemacher, Sanders, & Cerbo, 2010). The key idea is that metadata enables the information it marks to become useful to the user. A unique and thought-provoking definition is provided by Dempsey and Heery: “metadata is knowledge which allows human and automated users to behave intelligently” (1998, p. 148-149). Importantly, given technology’s constant state of flux and evolution, another valuable aspect of metadata is “to ensure that

reliable, authentic, and usable records are carried through time and across domains” (Wilson, 2010, p. 209).

“Traditionally, librarians and archivists have used the term metadata for ‘descriptive information used to index, arrange, file, and improve access to a library’s or museum’s resources’” (Morville, 2005, p. 125). Today, metadata comes in at least four general types including descriptive, structural, administrative, and use (Morville, 2005, p. 125; Rathemacher, Sanders, & Cerbo, 2010, p. 196). Descriptive metadata includes things like title, author, and format (Dempsey & Heery, 1998, p. 146; Rathemacher, Sanders, & Cerbo, 2010, p. 196). Structural metadata enables navigation through a complex digital object by explaining relationships between its various parts. Administrative metadata includes technical, preservation, and rights information (Morville, 2005, p. 125; Rathemacher, Sanders, & Cerbo, 2010, p. 196). Use metadata provides tracking statistics regarding circulation, web hits, and popularity (Rathemacher, Sanders, & Cerbo, 2010, p. 196). Another important way to view metadata is elucidated by Elings and Waibel (2007). They explain how different metadata standards interact to create and communicate information about a digital or analogue object. Specifically, there are metadata that create a structure for data (e.g., MARC, Dublin Core), metadata that describe content which fits into the structure (e.g., AACR2, LC Subject Headings), ways to format metadata (e.g., XML), and finally ways to communicate or exchange metadata across various systems (e.g., OAI) (Elings & Waibel, 2007, A grid of standards, para. 1 and table 1). For each category, there are multiple, different standards which might be employed, and “there is some tension between a requirement for simplicity and a requirement to recognise the diversity of resource types and the descriptive demands they raise” (Dempsey & Heery, 1998, p. 150).

## **From controlled to uncontrolled and points between**

There are several schools of thought on who should create metadata and how it should be created. The first requires the input of professional catalogers from institutions like libraries and museums. These institutions have developed metadata standards to categorize information; these standards are hierarchical. They provide controlled vocabulary to describe a singular vision of how information is best organized. This single vision, however, can be biased. For example, the Dewey Decimal System has been accused of having a particularly Western and Christian perspective (Rubin, 2010, p. 130). Yet, “the history of metadata is inextricably interwoven with hierarchy, for the organization of ideas and objects into categories and subcategories is fundamental to human experience” (Morville, 2005, p. 127). Prior to the digital age, such organization worked reasonably well because information in the form of a physical object could only be stored in a single location.

Today, “we invent new principles of organization that make sense in a world of knowledge freed from physical constraints” (Weinberger, 2007, p. 7). In terms of metadata generation, one sees the creation of folksonomies, bottom-up, crowd-sourced, faceted, uncontrolled vocabularies made by the general public (Avery, 2010, Folksonomies and Contemporary Practice, para. 3). Such tagging allows for searching on any given facet of information, and specifically the facet that is most important to that particular searcher. There is no biased hierarchy or worldview because there is no hierarchy or worldview other than the one the searcher creates (Avery, 2010, Social Indexing and Information Professionals, para. 2). As Weinberger states, “We can share orderings and build on them. Each enhances the meaning of the whole. None has to be given priority. None is more real than another” (2007, p. 233). This idea of nimbleness is key to tagging; user-generated metadata can be updated immediately and

constantly and evolves over time (Avery, 2010, Folksonomies and the Future, para. 5), while traditional metadata can stagnate or at least requires significant institutional time and resources to update (Morville, 2005, p. 138). These ideas are not exactly new but a medium (the Web) now exists that enables them. Paul Otlet's thoughts on how information users impact information is particularly in tune with today's thinking on folksonomies. For Otlet, "each user would leave an imprint, a trail, which would then become part of the explicit history of each document" (Wright, 2007, Chapter 11, The Forgotten Forefather, para. 17).

The need for a way to organize the world—thereby making knowledge accessible—is fundamental to humans (Wright, 2007). "We are all born with a deep-seated need to understand the world in terms of categories and to share that understanding with each other" (Wright, 2007, Chapter 2, Prototype Theory, para. 14). Hence, there is a middle road between the controlled and uncontrolled (Avery, 2010, Social Indexing and Information Professionals, para. 4); they are not mutually exclusive (Morville, 2005, p. 139). "Folksonomies should not be seen as opposed to traditional taxonomies, but rather as supplementary to them" (Avery, 2010, Conclusion, para. 1). Museums are leading the way by encouraging users to assist with cataloging objects by adding their own metadata (Avery, 2010, Folksonomies and Contemporary Practice, para. 3; Elings & Waibel, 2007, More Recent Trends, para. 4). Morville (2005) says it best: There is a "fundamental shift in power from author to reader and from authority to popularity that is only just beginning to make waves" (p. 152); however, "the wisdom of crowds does not negate the value of bright individuals and informed decisions" (p. 158).

### **Standardization, variety, and usability**

As noted above, there are many metadata standards to choose from; Hirwade reviewed 20 different options in her 2011 study. While several standards, like Dublin Core, have been

introduced as panaceas for all metadata ills, the reality is that different groups and different resources need different metadata (Dempsey & Heery, 1998, p. 154, 155; Hirwade, 2011, p. 18; Morville, 2005, p. 134). For example, the amount of metadata considered ideal for streaming media is incredible: “Ideally, every sound and every image needs to be indexed, tagged, and cataloged” (Siglin, 2009, p. 57). How can this variety in metadata standards be accommodated? One must ensure that they are interoperable; one must enable exchange and communication between them (Bengston, 2012, p. 351; Elings & Waibel, 2007, Introduction, para. 6; Hirwade, 2011, p. 24; Mitchell, 2012, p. 136). What has long been considered a problem needs to be recognized instead as the reality of the situation and managed accordingly.

### **Current and future solutions**

One option already being implemented is to combine top-down with bottom-up metadata (Wright, 2007, Chapter 11, *The Forgotten Forefather*, para. 21). As explained above this is already occurring in museums, and many libraries are also encouraging crowd-sourced cataloging through software that enables such social dialogue, like BiblioCommons ([www.bibliocommons.com](http://www.bibliocommons.com)).

The use of linked open data (LOD) seems particularly compelling for both increasing interoperability of metadata (Mitchell, 2012, p. 139) and for enabling the confluence between top-down and bottom-up vocabulary. M. Christina Pattuelli (2012) completed a study that employed LOD in the creation of the Linked Jazz Name Directory. Vocabularies were created from the crowd-sourced database DBpedia, and “vocabularies lie at the core of LOD functionality. They are key tools in facilitating the integration and reuse of content because of their capability to reduce semantic ambiguity and effectively support interlinking among different datasets” (Pattuelli, 2012, p. 559).

Jason Bengston proposes some very interesting ideas for the future of metadata using emergent behavior and self-organizing behavior witnessed in complex systems in other sciences. “The idea is to create a system capable of novel connections and complexity building that is not reliant upon human direction” (Bengston, 2012, p. 348). Both hierarchical, controlled vocabularies and folksonomies could “interact in unpredictable but potentially useful ways” (Bengston, 2012, p. 356) enhancing user access to information which, after all, is “the ultimate measure of success” for metadata and information retrieval systems (Elings & Waibel, 2007, More Recent Trends, para. 3).

## References

- Anonymous. (1998). The future is metadata. *Information World Review*, December, 55-56. Retrieved from <http://www.iwr.co.uk/>
- Avery, J. M. (2010). The democratization of metadata: Collective tagging, folksonomies, and web 2.0. *Library Student Journal*, February. Retrieved from <http://www.librarystudentjournal.org/index.php/ljsj>
- Bengston, J. (2012). Imagining emergent metadata, realizing the emergent web. *Journal of Library Metadata*, 12(4), 345-361. doi: 10.1080/19386389.2012.735525
- Dempsey, L., & Heery, R. (1998). Metadata: A current view of practice and issues. *Journal of Documentation*, 54(2), 145-172. Retrieved from <http://www.emeraldinsight.com/journals.htm?issn=0022-0418>
- Elings, M. W., & Waibel, G. (2007). Metadata for all: Descriptive standards and metadata sharing across libraries, archives, and museums. *First Monday*, 12(3). Retrieved from [http://firstmonday.org/issues/issue12\\_3/elings/index.html](http://firstmonday.org/issues/issue12_3/elings/index.html)
- Hirwade, M. A. (2012). A study of metadata standards. *Library Hi Tech News*, 7, 18-25. doi: 10.1108/07419051111184052
- Mitchell, E. (2012). Leveraging metadata to create better web services. *Journal of Web Librarianship*, 6(2). doi: 10.1080/19322909.2012.671653
- Morville, P. (2005). *Ambient findability*. Sebastopol, CA: O'Reilly.
- Pattueli, M. C. (2012). Personal name vocabularies as linked open data: A case study of jazz artist names. *Journal of Information Science*, 38(6), 558-565. doi: 10.1177/0165551512455989
- Rathemacher, A.J., Sanders, M. R., & Cerbo II, M. A. (2010). Summary of Metadata 101 presented by Kelcy Shepherd at the 2010 New England Technical Services Librarians Spring Conference: Crosswalks to the Future: Library Metadata on the Move. *Serials Review*, 36(3), 195-196. doi: 10.1016/j.serrev.2010.05.005.
- Rubin, R. E. (2010). *Foundations of library and information science* (3<sup>rd</sup> edition). New York, NY: Neal-Schuman.
- Siglin, T. (2009, October/November). Metadata: What you need to know (and why you need to know it). *Streaming Media*, 52-60. Retrieved from <http://www.streamingmedia.com/magazine/>
- Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. New York, NY: Holt Paperbacks.

Wilson, A. (2010). How much is enough: Metadata for preserving digital data. *Journal of Library Metadata*, 10, 205-217. doi: 10.1080/19386389.2010.506395

Wright, A. (2007). *Glut: Mastering information through the ages* [Amazon Kindle version]. Retrieved from <http://www.amazon.com>