The History and Application of Metadata Into An Informational World: A Brief Overview and Analysis

Ariele Dashow

April 2024

LIS6711: Organization of Knowledge

Professor Jinfang Niu

Metadata, otherwise described as “data about data”, is a process of describing the information presented by a source through a series of prescribed terms, known as controlled vocabularies (Gregory et al., 2010). While not a novel concept to the informational world, its development from manual process to nearly fully automated process has taking our digital sphere by force in the last two decades. Its quick implementation into information sciences in the 1980s exponentially pushed information literacy into modern times, and ability to grow within a rapidly advancing digital world has made it a common and standard practice within all physical and digital information institutions (Elo, 2020). Whether in library sciences, or our digital spaces shared by scholars, researchers, the government, or the general public alike, metadata has transformed the availability, discoverability, and assessability of any and all resources that have been adapted to an online presence (Falato, 2020). As a widely accepted and a constructed resource amongst information institutions, its standard practices and guidelines have been curated to facet different informational needs and systems (Taylor and Jourdey, 2017). In this paper, I will provide a brief historical analysis of the development and use of metadata within digital information systems, as well as address its application into library and academic information sciences.

Despite its “advanced” sounding name, the concept of metadata is not a foreign structural system in the information science realm. Metadata saw its earliest inception through the physical process of manually describing structural information about “things”, such as images, books, journals, articles, artifacts, etc. (Gregory et al., 2010). This was the common practice, curating a storing information in physical formats until the shift to digitizing information in the 1980’s (Gregory, et al., 2010). The transition into the digital practice of metadata grew slowly as systems and standards were developed, however, informational institutions began to see pick up in the late 1990s and well into the 21st century; these quick increases laying reason to the creation of the Internet as well as uniform codes, languages, and systems becoming internationally codified (Park, et al. 2010). In 2016, an article produced in

Scientific Data established the FAIR Guiding Principles which attribute to the four key functionalities of digital metadata: Findability, Accessibility, Interoperability, and Reusability (Wilkinson, et al. 2016). These guiding principles lead to various standardizations within the practice including designating system and organizational types to information institution groupings (i.e. museums, scientific online databases, libraries, etc.), usage of controlled vocabularies across multiplatform/multi-institutional systems, and formats to present the information in accessible and digestible means (Berg-Cross and Arbor, 2022). Even in the wake of advancement and normalization within the informational field, metadata remains to be a complex and ever-evolving configuration within our digital and physical dataspheres (Taylor and Jourdey, 2017).

Metadata and the databases that store these data profiles and the information within them cover wide arrays of knowledge fields; anything from geographic information systems, online library catalogs, cultural heritage domains, and more (Edwards et al., 2011). To understand metadata's application within these information sources, it needs to be dissected from its internal processes as well as their external system usage. Metadata is procured in multiple facets, integrating into multiple layers of data based on a metadata profile is created (Elo, 2020). Information scientists take the "thing" they are digitizing, and will break it down by its important attributes, such as title, type of "thing", author, figures in an image, descriptive terms, its file size, who created the file, and more (Taylor and Jourdey, 2017). Metadata itself has different factors and types, which allow for a multitude of attributing factors to be assigned to one item (Taylor and Jourdey, 2017). Priorly, this used to be done by hand, but the advancements made in metadata-profile creation systems can allow for the "thing" to be uploaded into the institution's governing software and done automatically (Taylor and Jourdey, 2017). As described, the process sounds relatively easy, but it has taken decades of system configuration and interinstitutional cooperation to make it as seamless as it is today.

From a user's point of view, metadata profiles are presented through a "discovery-layer" system designed internally by the organization (Falato, 2020). The typical user will not know all the governing terms that could be used to reach a resource within an organization's digital profile, so they will use common knowledge about the item to find it (Falato, 2020). Based on the language used within the first search, the database is able to pull all data bits that relate to the inputted search terms, giving users their first "discovery layer", similar to a general Google search (Falato, 2020). Scholarly and general sources alike will then provide advanced search features either through a separate function or an interactive side panel that will then allow users to penetrate other layers (Elo, 2020). These additional layers will apply different date ranges, authors, themes, established groups, color, size, etc. (Payne, 2013). Depending on the information inputted into the metadata profile, the resource will become available via these searches, or it will not (Falato, 2020). Like-terms will often be associated via the controlled vocabulary system used within the profile creation, but occasionally, a source will be so loosely connected or misconstrued to be one topic over another that multiple searches using different terms is sometimes necessary (Falato, 2020). Metadata curation is far from a perfect practice, but its adaptability to a growing digital age has encompassed incredible influence on the structure, creation, and maintenance of these profiles and the accessibility of information worldwide (Berg-Cross and Arbor, 2022).

This practice is not individualized by one type of informational institution, but is used commonly throughout library, scientific, government, and other databases (Elo, 2020). Academic, public, and government databases have induced the most habitual and continual use of metadata curation (Tomasi, 2018). Academic sources, like JSTOR, Library of Congress, online journals, PubMed Central, domestic, and international agencies, and libraries will typically align in terms of the controlled vocabularies and formats used when creating metadata profiles and configuring them to their digital sources (Tomasi, 2018). The principles established by Wilkinson's 2016 article *The FAIR Guiding Principles for Scientific Data Management and*

Stewardship ensured the ability to cross-use resources on these different platforms since many of them will house the same access points either to entire databases or specific sources (Berg-Cross and Arbor, 2022). Public information institutions, such as news and social media sources, museums, and cultural heritage sites, began applying the same standards and systems to increase interoperability between the sources (Kristensen and Sørensen, 2023).

The term "public" used to describe some of these information resources is intentional, given the social influence the general public can have on these types of information (Kristensen and Sørensen, 2023). A relating sector of information sciences includes folksonomies and taxonomies, which are other sets of organizational constructs used within the metadata world (Hedden, 2022). These managerial sectors of metadata control how the information is stored and organized based on the type of media the defined "thing" falls under, taxonomies being created by set procedures and data criteria and folksonomies being human controlled and created (Hedden, 2022). Metadata profiles are created fairly equally, regarding the type of information that needs to be inputted to establish a full profile, but it is the type of users and their relations to the different sources that draw a thin line between them (Kristensen and Sørensen, 2023). The traditionally viewed "academic" sources are used for purposes of research, education, and advancements within professional fields (Tomasi, 2018). Conversely, the more "public" sources, while still educational in nature, will tend to be more socially influenced, relevant, and stored, overall meant to draw public attention, or produce information in digestible terms (Elo, 2020). Despite these viewable differences, metadata remains pertinent to both source types as to their means of information accessibility and digital display.

In all, the functionality and producibility of metadata within information sciences has made great advances in our rapidly changing digital world. From its switch to software automated from manual, its widespread application to all digitally rendered resources, and international standardization to allow for the FAIR guiding principles to be grandly applicable, metadata is a systemic structure that has historically proved its worth and importance within

information literacy. The practice of finding, assessing, and digesting information has superseded its expectations over the past 40 years and has allowed us as global citizens to arm ourselves with knowledge that would have otherwise been restricted to its housing institution through interoperability and reusability of resources. Whether through academic or public structures, metadata has changed our digital informational landscape for the better, and will continue to develop as we do scientifically, informationally, and socially.

Resources:

1. Berg-Cross, G., & Arbor, S. (2022). Beyond Simple FAIR Principles for Ontologies and Semantic Resources: Grounding Rich, Meaningful Metadata. *Journal of the Washington Academy of Sciences*, *108*(4), 1–26. https://www.jstor.org/stable/27281297

2. Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, *41*(5), 667–690. http://www.jstor.org/stable/41301955

3. Elo, K. (2020). Big Data, Bad Metadata: A Methodological Note on the Importance of Good Metadata in the Age of Digital History. In M. Fridlund, M. Oiva, & P. Paju (Eds.), *Digital Histories: Emergent Approaches within the New Digital History* (pp. 103–112). Helsinki University Press. https://doi.org/10.2307/j.ctv1c9hpt8.11

4. Falato, B. J. (2020). Representation of Atypical Resources in the Discovery Layer: Metadata and Cataloging Aspects. In B. R. Bernhardt, L. H. Hinds, L. Meyer, & K. P. Strauch (Eds.), *The Time Has Come . . . to Talk of Many Things: Charleston Conference Proceedings, 2019* (pp. 296–297). Purdue University Press. https://doi.org/10.2307/j.ctv33t5ggk.55

5. Gregory, A., Heus, P., Ryssevik, J., & German Data Forum (RatSWD). (2010). Metadata. In *Building on Progress: Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences* (1st ed., pp. 487–508). Verlag Barbara Budrich. http://www.jstor.org/stable/j.ctvbkk43d.30

6. Hedden, H. (2022). The accidental taxonomist, 3rd Edition. Information Today, Inc.

7. Joudrey, D. N., & Taylor, A. G. (2017). *The organization of information; Chapter 5 An Introduction to Metadata*. Bloomsbury Publishing USA.

8. Kristensen, L. M., & Sørensen, J. K. (2023). Classifying the News: Metadata as Structures of Visibility and Compliance with Tech Standards. In J. K. Sørensen, J. M. Hartley, & D. Mathieu (Eds.), *DataPublics: The Construction of Publics in Datafied Democracies* (1st ed., pp. 145–164). Bristol University Press. https://doi.org/10.2307/jj.5590543.12

9. Park, J., Tosaka, Y., Maszaros, S., & Lu, C. (2010). From Metadata Creation to Metadata Quality Control: Continuing Education Needs Among Cataloging and Metadata Professionals. *Journal of Education for Library and Information Science*, *51*(3), 158–176. http://www.jstor.org/stable/40732596

10. Payne, D. (2013). Metadata: The Good, the Bad, and the Misunderstood. *GPSolo*, *30*(2), 51–53. http://www.jstor.org/stable/23630805

11. Riley, J., & Shepherd, K. (2009). A Brave New World: Archivists and Shareable Descriptive Metadata. *The American Archivist*, *72*(1), 91–112. http://www.jstor.org/stable/40294597

12. Tomasi, F. (2018). Modelling in the Digital Humanities: Conceptual Data Models and Knowledge Organization in the Cultural Heritage Domain. *Historical Social Research / Historische Sozialforschung. Supplement*, *31*, 170–179. https://www.jstor.org/stable/26533637

13. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18