Metadata Capital: Raising Awareness, Exploring a New Concept

by Jane Greenberg

Economics of Knowledge Organization Systems

EDITOR'S SUMMARY

While the value of information is widely recognized, the next step is recognizing metadata as an economic asset. Generating metadata involves costs in technological and human resources, but failure to generate and use metadata can lead to lost opportunity costs. Metadata activities are ultimately motivated by a drive for return on investment. The view of metadata as capital emphasizes that it has value that can rise with reuse. Metadata capital is defined as an asset that captures contextual knowledge about any information object, is produced by human labor or automated processes and yields a product or facilitates some service and can benefit the public. Expediting reuse of data and metadata is the key to maximizing their value, and early research demonstrates value in the linked open vocabulary environment and in reusing URIs. A collaborative project of the National Consortium for Data Science, the Metadata Capital Initiative, is focusing on the value of metadata through reuse in a big data setting to document its specific contributions to technology methods and intellectual advances.

KEYWORDS

metadata

economics of information

costs

information reuse

Jane Greenberg is a professor at the School of Information and Library Science, University of North Carolina-Chapel Hill, director of the Metadata Research Center and 2014 Data Science Fellow, National Consortium for Data Science. She can be reached at janeg<at>email.unc.edu.

nformation as an asset is a fairly well-accepted notion interconnected with our understanding of knowledge and the "knowledge economy" [1] [2]. Recognizing metadata as a valued asset or commodity is more challenging. This challenge stems from the ubiquitous definition for metadata as – "data about data" – and conflicting interpretations of data as valued information or as information without context having no value. Varying interpretations of data as it applies to metadata prompt us to ask: Can we recognize metadata as an asset? Where and how might we begin an inquiry on this topic? And more precisely, can we quantify the value of metadata in some way that will allow us to gather empirical evidence as to its value?

This *Bulletin* article considers these questions by introducing and exploring the concept of "metadata capital." The sections that follow discuss metadata and cost, present the initial idea of metadata capital in connection with metadata reuse and hone the definition of metadata capital. The last part of this article provides an overview of research targeting metadata capital that has been recently launched via the National Consortium for Data Science (NCDS) [3].

Metadata and Cost

"Data about data" is by far the most common definition for *metadata*. Although this phrase is a bit nebulous, it accurately captures the uber-level, and dare I say meta-meaning, for metadata. Metadata is a snippet of information (data with meaning) that describes or represents a larger chunk of content, and that content can be any "entity, form or mode" of information [4, p. 16]. More sophisticated definitions for metadata reference intellectual aspects such as data structuring and classification, and they

identify metadata-supported functions such as discovery, provenance tracking, rights management, authentication and other aspects of life-cycle management [5].

These more extensive definitions promote a higher order for metadata and help justify investing in metadata activities. Simply put, metadata generation, storage, maintenance and use require technological and human resources, and these resource requirements ultimately connect with financial ones. Conversely, there are costs, often hidden, associated with *not* producing metadata. For example, information cannot be discovered and system operations are non-executable without appropriate metadata. The logic is that metadata activities are motivated by a desired or calculated positive return on investment (ROI), also known as a net gain or positive yield. Viewing metadata in this context has given rise to the notion of *metadata capital* – a concept and an idea that may motivate metadata investment strategies overtime.

Metadata Capital: Honing the Definition

Metadata capital draws its meaning from a range discussions on capital. There are a wide variety of interpretations for *capital*. The most common rendering is "financial capital," commonly found in the business and operations literature. A theoretical foundation has been paved by well-known and somewhat popular works such as Max Weber's *The Protestant Ethic and the "Spirit" of Capitalism* and Adam Smith's *The Wealth of Nations*. Other examples include "intellectual capital," which connotes knowledge, and "social capital," which signifies personal and organizational networks with shared ideologies. The commonality among these expressions of capital is that a tangible, measurable object or phenomenon, such as a product, knowledge or a friendship, has a value and that investing in some activity can increase the value of that thing over time or lead to new assets.

Metadata capital, as first articulated, underscores metadata as an asset, the value of which may increase via reuse [6]. People put money in the bank, and it increases in value over time. The proposition put forward is that if good quality metadata is reused over time, there might be an increase in its value in comparison to the initial investment made to produce it. There

are limitations to this idea; for example, the price paid for a service or good may not equal the value. Additionally, if we follow simple supply/demand logic, more metadata suggests a decline in value. Notwithstanding these observations, the initial idea of reuse as a form of capital is serving to advance discussion about metadata value, and some empirical results have been gathered via study of the Dryad data repository's metadata workflow [6]. A second metadata context allowed us to pursue the idea within the linked open vocabulary (LOV) environment and through the reuse of URIs. This latter work also sought out identifying equations that might aid in measuring reuse of terms within the Helping Interdisciplinary Vocabulary Engineering Environment (HIVE) and the DataNet Federation Consortium [7] [8]. A refined definition for *metadata capital* is as follows:

1. An asset that contains contextual knowledge about content.
   a. Content is the data or information contained in any information object (any "entity, form or mode").
   b. Context is the *who*, *what*, *where*, *when*, *how*, *why*, etc., that can be captured via metadata attributes [9].
2. A product or service generated by human labor and/or machine-driven processes with value that increases over time or that enables the value increase of other assets.
3. A good (a service facilitator) supporting a range of functions such as discovery, provenance tracking, rights management, authentication, preservation and other functions associated with life-cycle management and access.
4. A public good if the product (metadata) is open, following which the services can be open.

To be clear, discussion and exploration on the value of metadata along with overlapping and associated topics (representation, cataloging, indexing) is not new by any stretch and predates the adoption of the word *metadata* in the information community. Moreover, this topic is one that will continue on, whether or not the word *metadata* stays in our *lingua franca*. Even so, exploring metadata capital is timely. The last section of this article presents a brief overview of the Metadata Capital Initiative and overriding goals.

Metadata Capital Initiative

The Metadata Capital Initiative is exploring the value of metadata via reuse with big data. The project is part of the data fellows program sponsored by the NCDS, a collaboration of leaders in academia, industry and government. The core mission for the NCDS is to address data challenges of the 21st century. Metadata plays a vital role in this mission. The cliché, your data is only as good as your metadata, underscores the significance of metadata for big data – or any data. The initiative is being launched through collaborations with two NCDS partners:

- National Institute of Environmental Health Sciences (NIEHS), specifically with an initiative exploring development of a common environmental health science vocabulary, spearheaded by Dr. Rebecca Boyles, data scientist, Office of Scientific Information Management NIEHS

- RTI International, specifically with a project on self-generated health information (SGHI) and the metadata associated with this data, led by Dr. Tom Caruso, health information liaison research associate, University of North Carolina School of Information and Library Science (UNC-SILS) and RTI's Center for the Advancement of Health Information Technology.

The NIEHS collaboration relates to the HIVE initiative and concept (URI) reuse, and the RTI International collaboration is exploring the value of data and metadata generated from personal health applications, such as Fitbits. Next steps include articulating relevant equations and testing their applicability for studying increase in data value. These steps support exploring metadata generation economics and potential value increases that may come from metadata propagation, although it is recognized that there are other facets to consider such as enrichment and impact. The work noted here is at an early stage, with the capital-sigma notation having been identified and modified for exploration, which allows for analysis of cost through increments defined by reuse. The team pursing this work includes research and students working at the School of Information and Library Science, Metadata Research Center, University of North Carolina at Chapel Hill, along with above noted collaborators from NIEHS and RTI International (Boyles and Caruso). The end goal is a final report documenting the assessment and recommending steps for future research. The team pursing this work includes students working at the School of Information and Library Science, Metadata Research Center, University of North Carolina at Chapel Hill, along with above noted collaborators from NIEHS and RTI International (Boyles and Caruso). The end goal is a final report documenting the assessment and recommending steps for future research.

Conclusion

Federal and international agencies, industry and many organizations have allocated substantial amounts of funding toward addressing big data challenges; however, the percentage of support directed toward specific metadata challenges is limited. The Metadata Capital Initiative, with support from the NCDS Fellows program, provides an opportunity to target metadata and advance work on metadata capital. The work underway intends to make both methodological and intellectual contributions specific to metadata. The fairly recent Snowden affair, revealing the U.S. National Security Administration's access to phone communication metadata, is perhaps not a timely event for the current government administration, but it has been a boon for the concept of metadata and the power it holds. In conclusion, there is evidence of a growing interest in metadata, and advancing our understanding of how to measure metadata value by exploring metadata capital may lead to other advances in information activities. ■

Resources Mentioned in the Article

[1] Machlup, F. (1962). *The production and distribution of knowledge in the United States*. Princeton, NJ: Princeton University Press.

[2] Drucker, Peter (1969). *The age of discontinuity: Guidelines to our changing society*. New York: Harper and Row.

[3] *National Consortium for Data Science*: http://data2discovery.org

[4] Greenberg, J. (2003). Metadata generation: Processes, people and tools. *Bulletin of the American Society for Information Science and Technology, 29* (2), 18-21.

[5] Greenberg, J. (2009). Metadata and digital information. In *Encyclopedia of library and information science* (pp. 3610-3623). New York: Marcel Dekker.

[6] Greenberg J., Swauger S., & Feinstein E.M. (2013). Metadata capital in a data repository. *Proceedings of the International Conference on Dublin Core and Metadata Applications DC2013*, 140-150.

[7] Greenberg, J. (2013). Metadata capital via a data HIVE. In G. Campbell (Ed.). *Advances in Classification Research/ASIST, SIG/CR 2013 Proceedings*, Montreal, Canada, November 3, 2013.

[8] Greenberg, J. (2014) Metadata capital via a linked data HIVE. *Advances in Classification Research Online, North America, 24*. Retrieved March 15, 2014, from https://journals.lib.washington.edu/index.php/acro/article/view/14680/12320

[9] Kunze, J. (2001). A metadata kernel for electronic permanence. *International Conference on Dublin Core and Metadata Applications, North America, DC2001*. Retrieved March 15, 2014, from http://dcpapers.dublincore.org/pubs/article/view/656