The M Word:
The Good, the Bad, the Ugly

For the past few months, I have been working with many wonderful colleagues to construct an overarching theme for the 2016 E-Content department in EDUCAUSE Review. We eventually settled on Libraries, the Academy, and Data: A Renewed Focus on the “M Word.” We chose this topic for good reason. Recent years have seen the primacy of data in the research process explode across the academy—from sciences, social sciences, and traditional STEM areas to the digital arts and humanities. Libraries, as the stewards of knowledge, have embraced new programs to mine scholarship in massively digitized collections and in massively parallel ways to enable the archiving, preservation, and publication of data to support open and replicable research in all disciplines. Although not all of the data that libraries seek to provide is open access (at least not yet), libraries as a whole are starting to deploy new areas of data support for higher education. For this first column in 2016, Julie Hardesty and I will focus on some new uses of data around the threads of the “M word”: Metadata. ~RHM

People have a tendency to label, well, everything. Giving things labels and describing things is a major tenant of discovery in the sciences and is how we move through the world. These labels and descriptions, often called metadata, flow through the academy from all aspects of the research process and are changing in higher education just as rapidly as they are in society overall. The expectations are that metadata will be clean and understandable, secure and accessible when appropriate, and easily shareable. The reality is that although this is all possible, it certainly doesn’t happen naturally or without concerted effort and cooperation within areas of information policy, design, and practice. Libraries know the potential fallibility of metadata created by hand, and as a result, the academic research library has a long history of working with metadata to ensure good storage, maintainability, shareability, and most importantly, accessibility.

Altmetrics and Their Impact on the Academy
Altmetrics, the practice of tallying online activity around a scholarly publication, is one area affecting the academy and its perception of itself. The walled-off print world of journals contained only in physical academic library buildings does not exist anymore. Online publications not only are more widely available but also are under pressure to be “open access” as grant-funding agencies and public pressure require more quantifiable evidence. Libraries are becoming more engaged in the scholarly publishing process by teaming up with university presses and by facilitating open-access and “new model” (i.e., data, software) journals to enable researchers to make their findings more openly available on a faster timeline than through traditional publishing. As Stacy Konkiel and Dave Scherer have noted, altmetrics can help supplement information about the impact of scholarly publications through online and social media connections that regular usage statistics from journals do not track. Being able to show the impact of research through online use and distribution means altmetrics are not so much “alt” anymore.

Full-Text and Metadata Mining
Now that libraries have access to massively digitized collections such as the HathiTrust Digital Library (https://www.hathitrust.org) and other collections of society and journal literature, we are seeing a research trend of wanting to use these collections in new and novel ways. Many researchers call this a “meta-use” of the collection. Much as in the realm of pharmaceutical research, many academics are now wanting to utilize library collections from a machine-oriented perspective, processing massive collections of full-text and metadata through the lens of the application programming interface (API) while using the power of institutionally based, high-performance computational instruments. Key to enabling this use is the capability to reuse data in a scientific workflow and the policy support to overcome such issues as intellectual property rights and new cost models for information access at this level.

Data, Society and Libraries
An increasingly popular source for large research datasets (i.e., Big Data) can be found in data produced by individuals on social media. danah boyd, founder of the research institute Data & Society (http://www.datasocietynet), and Kate Crawford have raised questions of ethics in gathering and using these Big Data sets. In order for Big Data to be considered a reliable source of research data within the academy, details about its provenance need to be as transparent as possible. Academic libraries, in cooperation with college/university central technology infrastructure, are often at the center of caring for, maintaining, and preserving research data sets, through institutional and other types of digital repositories. Though generally time-separated from the research process for gathering data, libraries are interested in data-set provenance for preservation
and reusability. Clearly there is potential for misuse of data when it encompasses potentially personal information and location-tracking through social networks. So how can libraries offer new services to support educational opportunities for the ethical use of societal data?

Creating Open Data for Instructional Opportunities

As James L. Hilton wrote in a 2014 EDUCAUSE Review E-Content column, learning management systems—used in nearly every academic learning environment in higher education—have a problem similar to that encountered with journal publishing, in that the academy “buys back (or, more often, rents back) the content that its members produce.” Libraries see a growing trend toward open data and open educational resources (OER) for use in instruction and are actively developing models for open content subvention. This encompasses open electronic textbooks as well as new forms of collaborative textbooks that serve the central need for core courses across an academic curriculum. Some of the novel uses for this type of collaboration can be found in the work of the Open Textbook Library (http://open.umn.edu/opentextbooks/) and MOOCulus (https://mooculus.osu.edu/). How can libraries enable these open-content initiatives to thrive in their current uses and retain that record of scholarship for the long-term archives of their institutions?

Publishing Software for Sustainability

In addition to the new role that data has enabled in the scholarly process, software has also become a critical component of that workflow. In our own library at Indiana University, we are seeing data publishing that includes entire virtual machines whose goal is to enable reproducible experimentation in cloud computational environments. This has led to the creation of scholarly journals—for example, Journal of Open Research Software (http://openresearchsoftware.metajnl.com) and SoftwareX (http://www.journals.elsevier.com/softwarex)—that exist to publish software with the hope that the publication will enable long-lived community support for the software. This publication also provides inventive researchers with a way to assign credit to those who have built the software infrastructure that enables their experiments. Libraries must find ways to become a part of this archival process that enhances software reuse among discipline-specific communities.

Conclusion

Libraries have experience with metadata and data management and are often the right agency to serve as a neutral mediator for collaborations among researchers. Innovative directions of engagement with data in the research process are creating new roles and opportunities for libraries to help in preserving, managing, publishing, and accessing data. Libraries are also housing the physical spaces used for collaborative data endeavors (e.g., visualization labs, maker spaces, and computational support) and are developing connections to scholarly social networks—such as ResearchGate (http://www.researchgate.net) and VIVO (http://vivoweb.org)—that enable new research connections for use in building research teams and models. Libraries are engaging higher education through these new data working relationships. In the coming months, the expectation is to engage you, the EDUCAUSE community, in this theme of Libraries, the Academy, and Data: A Renewed Focus on the “M Word.”

Notes


Robert H. McDonald (rhmcdona@indiana.edu), Associate Dean for Library Technologies and Deputy Director for the Data to Insight Center at Indiana University, is the 2016 E-Content Editor for EDUCAUSE Review. Juliet L. Hardesty (jlhardesty@iu.edu) is an Assistant Librarian and Metadata Analyst at Indiana University.

© 2016 Robert H. McDonald and Juliet L. Hardesty. The text of this article is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0).