Standards Column-Taming the World of Data: Pressures to Improve Data Management in Scholarly Communications

Todd Carpenter
NISO, tcarpenter@niso.org

Follow this and additional works at: http://docs.lib.purdue.edu/atg

Recommended Citation
Carpenter, Todd (2010) "Standards Column-Taming the World of Data: Pressures to Improve Data Management in Scholarly Communications," Against the Grain: Vol. 22: Iss. 6, Article 44.
DOI: https://doi.org/10.7771/2380-176X.5722

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
The University of Washington Libraries conducted a three-year pilot project of Netflix services, offering access through its Media Center. However, the subscription has since been cancelled. According to John Vallier, Head of Distributed Media Services at the UW Libraries, the service has not been missed. They have been able to respond to faculty needs through robust collection development efforts. Vallier noted that a service such as Netflix could be useful for a smaller school that could not afford to purchase materials.

**Reaction by Netflix and Librarians**

The Netflix Terms of Use specifically state, “The use of the Netflix service, including movies and TV shows made available to you by us, is solely for your personal and non-commercial use.” (http://www.netflix.com/TermsOfUse) In addition, Netflix has indicated that it considers the use of Netflix subscriptions by libraries to provide access to DVDs and streaming video to patrons against its Terms of Use. Travis Kaya’s September 18, 2010 post in Wired Campus on the Chronicle of Higher Education Website stated that “lending DVDs out for faculty members to project on-screen in class is in violation of Netflix’s terms and scope since the turn of the millennium. An analysis of research data, which has exploded in both size and complexity, including processing, publication aspects of data sharing are expected to take a considerable time to come. Several organizations have begun addressing aspects of the complexity, including CODATA (http://www.codata.org/taskgroups/), ICSTI (http://www.icsti.org/documents/Numeric_Data_FINAL_report.pdf), Science Commons (http://neurocommons.org/report/data-publication.pdf), the Dataverse Network Project (http://dataverse.org/citation/standard), NISO (http://www.niso.org/workrooms/supplemental), and the UK’s Digital Curation Center (http://www.dec.ac.uk/resources/policy-and-legal/policy-tools/

In the same piece, Ciara Healy said that in setting up the Netflix program at Wake Technical Community College “she acted according to federal copyright law, which allows faculty members to share legally obtained material in face-to-face instruction.” However, the question has been raised as to whether a court of law would consider a library’s use of Netflix “personal.” Kevin Smith, Duke University Scholarly Communications officer, told Library Journal, “[T]hose terms of use may indeed trump rights that one has under the copyright law; they are part of a private agreement between Netflix and a user with can create liability only for those two parties if there is a breach of terms.” (Hadro 2010)

In response to recent articles and blog posts discussing the lending or streaming of Netflix videos by libraries, some library folk have voiced their disapproval. Meredith Farkas, Head of Instructional Services at Norwich University in Vermont and author of the “Technology in Practice” column in American Libraries, posted an entry on her blog in which she reprimands libraries. Voicing concern about the legality of library streaming of Netflix videos, she says, “Netflix does not have institutional subscriptions. Therefore, what the library is doing is in violation of Netflix’s terms of services and opens them up to legal repercussions.” In a September 18, 2010 post on his blog, Library Law blogger Peter Hirtle agrees, saying “I don’t see how a library subscription to Netflix could be considered to be “personal” — not when the purpose of the subscription is to lend the movies to others, rather than watch them yourself (as if a library could even watch a movie).”

**References**


and-guidance). As the problems have grown in complexity, the number and scope of organizations investing time and energy in this space is increasing rapidly.

This growth in interest by organizations around the world makes the issue of coordination increasingly important. A favorite joke regarding standards is particularly relevant to the current situation regarding data distribution. Connie Morella, former congresswoman and ambassador to the Organization of Economic Cooperation and Development, said during an ANSI’s World Standards Day gala, “Standards are like toothbrushes. Everybody wants one, but nobody wants to use anybody else’s.” This is especially true in the area of research data, which spans such a broad swath of the research community. What is taking place on one end of the earth in a particular discipline is often at odds with another project halfway around the globe or even next-door in a different discipline. While some of the challenges are domain-specific, many of the problems span all fields.

CODATA is one organization that is stepping up to the coordination question and some of the thorny coordination. CODATA is an interdisciplin ary Scientific Committee of the International Council for Science (ICSU) that works to improve the quality, reliability, management, and accessibility of data of importance to all fields of science and technology. Last October during their biannual conference in South Africa, a Task Group on Data Citation was launched. This international group, organized jointly by several CODATA committees and the International Council for Scientific and Technical Information (ICSTI), will explore the technical, scientific, socio-cultural, institutional, legal, and sustainability questions regarding data use, including references to portions or subsets of data. They are also quite aware that citing a dataset has further implications regarding the ability to reliably identify, locate, access, interpret, and verify the version, integrity, and provenance of the dataset. The goal is to help coordinate activities in this area internationally and promote common practices and standards in the scientific community. The CODATA group hopes to organize a summit next fall to build awareness and to promote better cooperation among the various leading organizations at work on these topics.

The joint NISO-NFAIS project on Supplemental Journal Article Materials is another project that touches on this space. In scope, how-ever, it is both larger and more tightly focused than the CODATA effort. It is larger from the perspective that it covers any type of supplemental material — not only research data, but also digital notebooks, textual supplementary data, software applications, audio, video, or any of the other supporting content that authors submit along with their articles for publication in scholarly journals. Nevertheless, it is also much more tightly focused on the publication-related questions, avoiding the more complex questions of provenance, copyright, security, data integration, packaging, and sharing. The project has begun with defining terms such as what content is supplemental, ancillary, and core. It is also looking at metadata questions and how to effectively link journal content and supplementary component elements. By working with the publishing community, the Supplemental Journal Article Materials project can help to codify and promote recognition of and use of these materials in the publication stream, as well as to ensure that libraries and researchers can effectively access and use them.

The Science Commons group, a sister organization aligned with the Creative Commons, is another organization with work underway. Their project, led by John Wilbanks, is looking at the legal structures necessary to share data among domain-specific research and development and legal protections regarding intellectual property are often among the most challenging issues for distribution of content. While U.S. Copyright Law doesn’t protect factual items, there are protections for the organization and representation of data forms. Where the lines are drawn in scholarly data has not yet been determined by case law or regulation and will likely not be easily decided. In addition, different laws or regulations apply outside the U.S., where, in some cases, copyright in data can be asserted. If data is shared across international boundaries, a case can be made that the data that is returned will retain the more stringent legal structures. Science Commons hopes to promote an open license solution to data sharing based on a similar structure to the Creative Commons licenses for publications and other creative works.

Existing work conducted by the Open Archives Initiative on Object Reuse and Exchange (ORE) (http://www.openarchives.org/ore/) could play a significant role in the packaging and distribution of datasets. The OAI-ORE specification presents a model for describing how elements within a compound digital object are identified, described, and related to one another. Although originally developed to deal with aggregations of Web resources, such as Web pages or whole Websites, the specification has potential to be applied to scientific datasets. ORE has seen implementation in a few testing environments such as the Chroniling America Historic American Newspapers project (http://groups.google.com/group/oai-ore/browse_thread/thread/4a71d109b665a6f?pli=1) and the oreChem project (http://www.openarchives.org/oreChem/). While ORE provides a semantic and logistical framework for packaging and distributing datasets, significant work remains before it can provide the needed tools for the scientific community.

One of the most critical success factors for the rapid adoption of the standards that are developed is making changes within the social and political environment. In the early-to-mid 20th century, the publication of scholarly journal articles took off as tenure systems were developed that required the publication of research results for promotion consideration (the “publish or perish” mantra). The new government and non-government requirements for sharing of data, mentioned earlier in this article, are having a similar impact. However, these sharing mandates are only the beginning of what is needed to support a long-term infrastructure for data management. Along with legislation and policies, where the funding will come from for all of this data management is a major concern. The recognition of the value of data sharing is of course social, not technical or political. Some researchers are reluctant to share data and some of their organizations have created restrictions on sharing or developed incentives (like the promotion and tenure system) that could result in a mind-set of hoarding one’s data. Both these organizational and individual tendencies to limit sharing will need to be overcome to succeed in large-scale data projects.

Each of these elements: legislation, organizational policy, individual behaviors, intellectual property, funding, technical infrastructure, technology, and information management standards will need to be addressed for the data sharing vision to be realized. These issues are large and interwoven and cannot be solved without significant collaboration between the affected parties and the many organizations that represent them. But the recognition of the value of research data seems to have become pervasive enough that now is the right time to facilitate this collaboration. The new government and non-government requirements for sharing of research data may just be the “tipping point” that is needed to ensure that standards are developed and adopted for the identification, citation, cura-tion, and provenance of datasets.