

December 2010

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

Todd Carpenter  
NISO, [tcarpenter@niso.org](mailto:tcarpenter@niso.org)

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

 Part of the [Library and Information Science Commons](#)

---

### 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](mailto: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 or allowing students to watch streaming video from a **Netflix** account is something the company 'frowns upon,'" according to **Netflix's** Vice President of Corporate Communications, **Steve Swasey**. **Swasey** told the *Chronicle* "We just don't want to be pursuing libraries. We appreciate libraries and we value them, but we expect that they follow the terms of the agreement."

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)

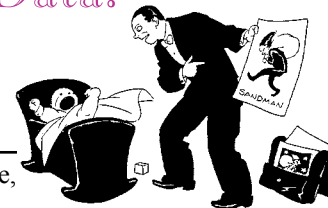
In response to recent articles and blog posts discussing the lending or streaming of **Netflix** videos by libraries, some library folk have voiced their disapproval. **Meredith Farkas**, Head of Instructional Services at **Norwich University** in Vermont and author of the "Technology in Practice" column in *American Libraries*, posted an entry on her blog in which she reprimands libraries. Voicing concern about the legality of library actions, **Farkas** 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, **LibraryLaw** 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

- Blumenstein, Lynn**. "Small Libraries Start using Netflix." *Library Journal* 133, no. 7 (April 15, 2008): 18.
- Farkas, Meredith**. "Netflix in libraries and hypocrisy." *Information Wants to be Free* (September 18, 2010), <http://meredith.wolfwater.com/wordpress/2010/09/18/netflix-in-libraries-and-hypocrisy/>.
- Fitzgerald, Rebecca**. "Using Netflix at an Academic Library." *Tame the Web* (September 9, 2010), <http://tametheweb.com/2010/09/09/using-netflix-at-an-academic-library-a-tw-guest-post-by-rebecca-fitzgerald/comment-page-1/>.
- Hadro, Josh**. "If Netflix Use Isn't Kosher, Can Libraries Find an Alternative?" *LibraryJournal.com* (September 21, 2010), [http://www.libraryjournal.com/lj/tools/articlesandawards/886889-306/if\\_netflix\\_use\\_isnt\\_kosher.html.csp](http://www.libraryjournal.com/lj/tools/articlesandawards/886889-306/if_netflix_use_isnt_kosher.html.csp).
- Healy, Ciara**. "Netflix in an Academic Library: A Personal Case Study." *Library Trends* 58, no.3 (Winter 2010): 402-411.
- Hirtle, Peter**. "Using Netflix in a Library." *LibraryLaw Blog* (September 18, 2010), <http://blog.librarylaw.com/librarylaw/2010/09/using-netflix-in-a-library.html>.
- Kaya, Travis**. "Academic Libraries Add Netflix Subscriptions." *Wired Campus. Chronicle of Higher Education* (September 18, 2010), <http://chronicle.com/blogs/wiredcampus/academic-libraries-add-netflix-subscriptions/27018>.
- Kaya, Travis**. "Netflix 'Frowns Upon Use by Academic Libraries.'" *Chronicle of Higher Education* 57, no. 6 (October 1, 2010): A8. 

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

by **Todd Carpenter** (Managing Director, NISO, One North Charles Street, Suite 1905, Baltimore, MD 21201; Phone: 301-654-2512; Fax: 410-685-5278) <[tcarpenter@niso.org](mailto:tcarpenter@niso.org)> [www.niso.org](http://www.niso.org)



There is a giant and rapidly growing wild-west-like expanse in scholarly communications. It has few boundaries, few rules, and appears as expansive as the Big Sky country of Montana. I'm speaking of the world of research data, which has exploded in both size and scope since the turn of the millennium. An often-quoted report by IDC in 2008 (<http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>) concluded that the pace of data creation had exceeded the capacity to store that information and with the rapid implementation of sensors and data creation tools of every type, this trend is unlikely to abate. Diverse and complex problems exist in managing all this data.

External factors are also driving this growth in data availability and distribution. In 2007, **President Bush** signed the **America COMPETES Act** (PL 110-69) into law, which among many other things requires civilian federal agencies that conduct scientific research to "develop and issue an overarching set of principles to ensure the communication of open exchange of data and results to other agencies, policymakers, and the public." This led various organizations,

both within and outside the federal government, to review their policies on data management. In October, the **National Science Foundation** amended its grant proposal submission guidelines to require the inclusion of a detailed Data Management Plan. This change is in support of **NSF's** new **NSF Data Sharing Policy** (<http://www.nsf.gov/bfa/dias/policy/dmp.jsp>), which states that recipients of grants are "expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing." They are not the only grant funding organization to expect awardees to facilitate and participate in data sharing. The **National Institutes of Health** has been a leader in promoting data sharing ([http://grants.nih.gov/grants/policy/data\\_sharing/](http://grants.nih.gov/grants/policy/data_sharing/)) since 2003. Other non-government sponsors of research such as the **Wellcome Trust** (<http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/wtx035043.htm>) — a global charitable foundation that sponsors research

in bioscience, medicine, and the environment — and the **Australian Research Council** (<http://ands.org.au/guides/code-awareness.html>) have implemented policies on data sharing. These are only a few examples among many throughout the world.

While the number of organizations demanding that scholars share their data is increasing, there is not yet clear understanding of how to accomplish all the sharing that is being mandated. The political, legal, technical, curatorial, and publication aspects of data sharing are problems our community will be addressing for 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](http://www.icsti.org/documents/Numeric_Data_FINAL_report.pdf)), **Science Commons** (<http://neurocommons.org/report/data-publication.pdf>), the **Dataverse Network Project** (<http://thedata.org/citation/standard>), **NISO** (<http://www.niso.org/workrooms/supplemental>), and the UK's **Digital Curation Center** (<http://www.dcc.ac.uk/resources/policy-and-legal/policy-tools>

continued on page 85

## Standards Column from page 84

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 thornier questions of citation. **CODATA** is an interdisciplinary **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 and citation, 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 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 publications. From the perspective of data, however, it is 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 to understanding. 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 researchers. As is usually the case, copyright 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 strictures. **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 **Chronicle America Historic American Newspapers** project ([http://groups.google.com/group/oai-ore/browse\\_thread/thread/4a71d09b6b5a6feb?pli=1](http://groups.google.com/group/oai-ore/browse_thread/thread/4a71d09b6b5a6feb?pli=1)) and the **oreChem** project (<http://www.openarchives.org/oreChem/>). While **ORE** provide 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 biggest inhibitor of adoption 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, curation, and provenance of datasets. 🌸

## Pelikan's Antidisambiguation from page 86

content would be what's popular to describe today as a great big "value add," for it would mean that **Amazon** would no longer be locked out of selling licensed content to owners of **Sony Readers** or **Barnes & Noble Nooks** — or the other way 'round, don'tcha see...

Then, the competition could be between makers of devices, based upon features, quality, snaziness, etc. There's room in the world for **Sears**, **Best Buy**, **The Sharper Image**, and **Hammacher**

**Schlemmer**. I mean, they all sell (or ought to sell) amazingly cunning nose hair trimmers. Why not content access devices?

And the content vendors could compete based upon the depth of their catalogs, the quality of their customer service, their ability to address the diverse interests of niche communities, and so forth.

Not so difficult from a technology point of view, really...

Well, ok, it is difficult. And you still have to empty the darned things (the nose trimmers, I mean...).

But so is every other worthwhile thing difficult

that we're all trying to accomplish in this increasingly complicated, inescapably interconnected world. And yet there must be at least fifty commercial concerns around the globe (my own wild guess, for which **Against The Grain** holds no responsibility) whose interests are focused upon perfecting the nose hair trimmer. (**Ishmael** said, "...we are all somehow dreadfully cracked about the head, and sadly need mending.") So there must be a market...

This means it's not really an argument about what can or cannot actually be done.

We just have to decide how we'd like this all to turn out. 🌸