Research Data Management

Practical Strategies for Information Professionals

Edited by Joyce M. Ray

Charleston Insights in
Library, Archival, and Information Sciences

Purdue University Press
West Lafayette, Indiana

Contents

6 | Copyright, Open Data, and the Availability-Usability Gap

Challenges, Opportunities, and Approaches for Libraries

MELISSA LEVINE

THE BIG IDEALS

This chapter is about copyright as one of several significant bodies of law that touches on the creation, preservation, and use of data. And yet, this chapter barely discusses copyright at all, instead approaching copyright as a matter of policy, administration, and business choices that should minimize the complexity of copyright over the life cycle of research data. In doing so, the products of research may more easily be reused and reinvested.

"Data is the new gold," according to Neelie Kroes, vice president of the European Commission responsible for the Digital Agenda (Kroes, 2011). Computing power has increased exponentially at relatively low prices, allowing us to examine the world and detect patterns on a scale never before possible. There is much being written about data, data management, big data, and strategic approaches to managing research data. Interest and commitment have accelerated in response to the combined incentives of increased computing power and the impetus of the demands of research sponsors and funding agencies for data management plans as a condition of funding.

Right now there is a gap between aspiration and reality. Nobody has solved all the complexities of making data available and usable. In fact, nobody has yet even figured out what all the questions are. Complex practical questions abound. For example, most research universities now offer a variety of resources, guides, and planning tools that take the first steps toward providing researchers with the needed infrastructure to meet the formal requirement of having a data plan for grant purposes. It can be

daunting simply to navigate all of these resources for researchers trying to write a grant proposal.

Libraries are natural hubs for services that support preparation of data management plans, fulfilling data management obligations, and citing the products that result from research projects. Librarians can help academic researchers to compete for grant funds, protect their data, and receive recognition for it. (See a list of sample guides in "Resources" at the end of this chapter.) Librarians who are considering new data services as an extension of traditional library outreach efforts can benefit from the experiences of libraries that have already established such services and that have been active participants in information access initiatives and by maintaining resolve on key principles.

PRINCIPLES

In addressing copyright—and really most legal questions in the data arena—it is useful to remember that the goal of reliable, efficient access to knowledge is not inherently new. It is foundational for scholarship, and we can tether otherwise complex problem solving by keeping these foundations in mind. Statements like the Denton Declaration maintain focus on essentials in a constructive manner. The Denton Declaration is an aspirational statement for open data that also has practical applications. It is a statement that "bridges the converging interests . . . and promotes collaboration, transparency, and accountability across organizational and disciplinary boundaries" (Open Access @ UNT, 2012). Drafted in 2012 at a meeting at the University of North Texas by stakeholders like librarians, scholars, technology experts, researchers, and university administrators, the Denton Declaration starts with the premise that "open access to research data is critical for advancing science, scholarship, and society" (Open Access @ UNT, 2012). The declarations are:

- Research data, when repurposed, has an accretive value.
- Publicly funded research should be publicly available for public good.
- Transparency in research is essential to sustain the public trust.
- The validation of research data by the peer community is an essential function of the responsible conduct of research.
- Managing research data is the responsibility of a broad community of stakeholders including researchers, funders, institutions, libraries, archivists, and the public.

By keeping these ideas at the forefront in data policy discussions, the Denton Declaration helps to fill the aspiration-usability gap as a mooring point for policy, administrative, and legal strategies (Open Access @ UNT, 2012).

OPEN ACCESS, OPEN DATA, AND FUNDING MANDATES

"Open access" and "open data" differ in complexity, though they have much in common. Open access is now a familiar concept in the research community when it comes to publications. Conceptual familiarity with open access is useful for working through analogous strategic questions for data (in contrast to publications). Open access, for the most part, assumes a text-based expressional artifact with some level of stability (versioning). The technical infrastructure and range of legal concerns is relatively simpler than those associated with data. With data, there is an emphasis on its value as evidence, reproducibility of results for validation, and utility for future research. There is a gap between data that is "available" and data that is "usable," which can be exacerbated in the absence of legal metadata. Mandates from funding agencies are incentivizing the development of needed policies and practice by conditioning funding on meeting open access and data management requirements. In doing so, these mandates are helping to bridge this gap. These mandates reflect a major shift in policy. By directing and requiring openness broadly, and with regard to data specifically here, these mandates dictate that researchers take a new approach to storing, securing, and describing data—including legal information for copyright elements.

Funding Mandates

On February 22, 2013, the White House Office of Science and Technology Policy (OSTP) issued a memo titled "Increasing Access to the Results of Federally Funded Scientific Research." The memo directs federal agencies with more than $100 million in research and development expenditures to prepare policies to make federally funded research results publicly available, free of charge, within 12 months after original publication. The new requirements ensure public access to scientific research data by ensuring that intramural researchers and researchers who receive federal grants and contracts for scientific research have data management plans in place. To support the implementation of data management plans, grant

proposals may include "appropriate costs for data management and access in proposals for Federal funding for scientific research" (Holdren, 2013, p. 5). The policies have to "include mechanisms to ensure that intramural and extramural researchers comply with data management plans and policies" (Holdren, 2013, p. 5). Further, they are to "promote the deposit of data in publicly accessible databases, where appropriate and available" and "develop approaches for identifying and providing appropriate attribution to scientific datasets that are made available under the plan" (Stebbins, 2013, p. 5).

This builds on the 2008 National Institutes of Health (NIH) mandate for open access deposit in PubMed Central, which required publications resulting from grant-funded research be open in some way as a condition of receipt of grant funds. The NIH mandate requires that an electronic version of all final peer-reviewed journal articles published as a result of NIH grant funds and accepted for publication on or after April 7, 2008, "be made publicly available no later than 12 months after the official date of publication: Provided, That the NIH shall implement the public access policy in a manner consistent with copyright law" (NIH, n.d.). Similarly, the National Science Foundation (NSF) announced on May 2010 that it would require data management plans for all grant proposals; the mandate became effective on January 18, 2011 (NSF, 2010; NSF, 2011). Other notable examples of influential mandates in this arena were implemented by the World Bank, the Wellcome Trust, and the Australian Research Council (Australian Research Council, 2013; Wellcome Trust, n.d.; World Bank, 2012).

Copyright considerations are among the first items listed to ensure authors have the needed rights to deposit. A condition of funding is that published results must be available open access. In a way, it asserts form of contract; in exchange for funding, research product must be open access. In doing so the mandate trumped copyright restrictions that occur when a researcher signed a publishing contract that transferred copyrights to a publisher. This in turn affected the ability of the general public, taxpayers, to access the product of taxpayer-funded research if the resulting article was available for a fee. The researcher is bound by that open access condition in the grant and can only pass on some subset of rights that he or she has to publishers.

ISSUES AND IDEAS

To master the challenges posed by this new data environment and to take advantage of the opportunities, librarians may wish to consider these ideas and issues as they think about how to navigate the legal implications.

Issues

Technology outpaces law

As is often the case, the pace of technological innovation has outpaced our social and legal frameworks. By analogy, in the scholarly world we are using the tools of the 17th century (that is, linear text). You can move up to the 1980s if you take word processors into account. So we are trying to retrofit the research tools of today into packages that combine technologies that are anywhere from a few decades to a few centuries old. This extends to the legal framework that we work within when it comes to copyright.

Different countries do it differently

Even with international treaties in place, copyright is handled differently in different countries, making interoperable data more challenging. Cross-border privacy norms differ, as do assumptions about contract law. There are inherent differences between public and private sector uses and needs. There are some choices that can be controlled regardless of jurisdiction, however. For example, creators can work toward limiting and reducing complexity through the terms of licenses and contracts associated with data that they generate.

Different disciplines do it differently

Very generally, science research data tends to be expressed as facts, say, numbers and measurements that are not subject to copyright protection under U.S. copyright law. While humanities data may take the form of numbers and measurements, by contrast it is often expressed as clips of sound or video or other modules of information that *are* in and of themselves subject to copyright. Use of those modules in research may indeed be subject to a limit on copyright such as fair use, but it is unclear as to whether a subsequent researcher's use also would be a fair use. Fair use requires a case-by-case consideration of facts each and every time. So in storing research data that includes elements that are subject to someone

else's copyrights, it becomes important to document and store legal metadata about the copyright information associated with each module of copyright-eligible element for the resulting body of data to be reusable.

Further, at this moment there are still highly varied comfort levels with the whole notion of research as data, as well as "openness" in concept and practice. The scientific method at its essence is evidence-based. It depends on sharing and requires access to data as evidence of results. Yet many prestigious papers are based on data that is not replicable or even accessible. In the past, it has not been worth the effort and investment to maintain bodies of data once the papers are written—the published papers are the surrogates for the data.

Right now, taxpayer-funded research is probably generating data that is not being fully utilized, functionally missing significant opportunities. As a result, we may waste money on the same research or miss opportunities to reuse existing data for new inquiries. As much as 80 percent of data in science, technology, engineering, and medicine are not replicable (Hartshorne & Schachner, 2012). For economics journals, data practices are in need of reconsideration (Vlaeminck, 2012). Managing copyright and related rights proactively will improve the incentive to maintain data and make it more useable.

Hidden data

Why are researchers hiding their data (Piwowar, 2011; Savage & Vickers, 2009)? Copyright is usually framed as a "problem," but the real impediment seems more diffuse: a disincentive to share, difficulty in credit and attribution, low value in the tenure process, and the essential complexity of managing data are more likely factors. There may not be a single entity that will emerge as responsible for these issues, but there is a growing network of researchers, funding agencies, universities, and corporations engaged in policy and management for research data. With the goal of making research data reusable and open, we have models for copyright management based on established open access frameworks and digital preservation and curation practices.

Citation Standards for Data and the Need for Recognition and Good Data Management

There are real challenges: evolving questions about reliable access, the need for discovery tools and citation standards, and lack of willingness to deposit and cite data that is not stable or reproducible. Some faculty members worry

that data in and of itself independently of journal articles will cause them to spend a disproportionate amount of time on data for which they do not get meaningful recognition or credit. This seems to be the case for nonjournal products generally. However, well-managed data will be citable, increasing its relevance for promotion and tenure. Data plan requirements may improve the relevance of generating meaningful data, serving as a catalyst for thinking through the workflow associated with generating and managing data. Library services to assist researchers in developing and implementing data management plans can further the relevance of data to the tenure and promotion process. By providing repository storage and services, librarians can ensure that data is properly documented for citation and reuse.

Data Citation Is Vital for Better Sharing

Standards for data citation are taking form to improve sharing and credit. Since attribution and impact are the lifeblood of scholarship, we need consistent data citation behavior. In doing so, a researcher can be credited not only for his or her direct research, but also for the reuse of the data by someone else. At this moment, the field is working toward the needs of universities and researchers to measure output and impact. For example, the Open Researcher and Contributor ID (ORCID) provides persistent digital identifiers for individual authors and researchers that allow better tracking for credit and valuable information for funders: "ORCID is an open, non-profit, community-based effort to provide a registry of unique researcher identifiers and a transparent method of linking research activities and outputs to these identifiers" (ORCID, n.d.). Basically, it allows researchers to associate unique author identifiers with research products. It is free for individuals—anyone may register for a number. Note that the ORCID identifier is distinct from the digital object identifiers (DOIs) already predominantly used for articles.

What does this have to do with copyright? This kind of citation system enhances the likelihood of having researchers remain connected to their scholarly output—to the extent copyright may exist in some aspect of the data or its structure, the more likely one can trace it back to a particular rightsholder. Currently, it may be difficult to give credit to researchers for their data because the relationship between the data and the researcher is not well documented. By tackling data citation, some elements of the copyright question can be addressed proactively (Mooney & Newton, 2012).

Good Metadata is Expensive

Metadata is as significant as the data itself. Producing meaningful metadata and making it available requires resources and expertise. There is rightly an emphasis on acquisition, preservation, and integrity of content, but without consistent metadata, the data may end up being of variable quality or utility. Data repositories are (happily) not static, but legal metadata (such as rights and access) may or may not be static (not so happily). The ever-changing nature of a repository makes it both valuable and problematic. Partnerships like DataCite are providing an important forum for having meaningful metadata for future utility. Because of their expertise with describing information and awareness of information rights and access issues, librarians are well positioned to extend their activities to include assistance with documentation of legal metadata.

Data Results as Complex Artifacts

What is the legal status of a collection of research data built from a variety of data that is generated by others or that generates other products? The value and emphasis of scholarly output has grown beyond articles. "Publications" might mean products like datasets, software, patents, and other forms of expression or documentation in addition to text-based material. All of these products are eligible for some kind of intellectual property protection; at the same time, the use of licenses can ensure that data products of research are maintained in an open fashion. If datasets are released without licenses, someone else could assert copyright in the dataset or claim credit. If the original dataset includes rich metadata, then subsequent uses can relate back to the source. As long as underlying content is retained as open, is there a reason to prevent commercial use? Intellectual property could be negotiated in a way to maintain usefulness and engage commercial productivity, perhaps with financial support or royalties back to the research endeavor, and provide insulation from liability to researchers and their institutions for subsequent uses.

Different Values, Common Interests: The False Dichotomy of Public and Private Sector Interests

This chapter focuses on ways to "keep data open," but we could reframe the dichotomy that has developed in the distinctions between public vs. private

or commercial vs. noncommercial to think about how these can be complementary. Technology transfer offices are an established feature of research universities with the assumption that the private sector is well suited to bring productive research products to market. If publicly funded data is "open" for further research and is kept open, then perhaps it should be equally available to the private sector for use and investment (provided there is no legal impediment such as privacy concerns). This is a practical and philosophical observation. Businesses can and do gather lots of government data because it is open, and in doing so, they are able to expose other data and develop useful products or services. Investment in editorial staff may be part of the enhancement and value. There is a high level of work involved in taking a content set and developing valuable applications, requiring the right tools, such as subject matter experts. There may be cases where the research agenda does not need or have interest in this kind of enhancement, and where the private sector can make productive use of reliable metadata and rich data generated from publicly funded projects. This is actually a further validation of good data practices and a different kind of enrichment.

Ideas

The copyright questions are a component of the kinds of issues just discussed. Data are not reliable if you do not know where they are located, how to refer to datasets, or if you cannot assess the status of the components that make up the data, which may include elements that are subject to copyright (or indeed some other legal concern, such as privacy). In turn, it is not possible to validate copyrightable outputs based on the data, such as visualizations or articles. Projects and products like DataCite, Data Observation Network for Earth (DataONE), and Databib will help with copyright and other legal matters by providing ways for data to remain associated with metadata.

DataCite (n.d.) explains what the organization does on their website:

> We bring together the datasets community to collaboratively address the challenges of making research data visible and accessible. Members of DataCite meet in person every six months at summer and winter conferences, and collaborate in established working groups. Through collaboration, we support researchers by helping them to find, identify, and cite research datasets with confidence; support data centres by providing

> persistent identifiers for datasets, workflows and standards for data publication; support journal publishers by enabling research articles to be linked to the underlying data. Currently we are working primarily with organizations that host data, such as data centres and libraries.

The website for DataONE (n.d.) describes the organization as: "the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data." DataONE provides a substantive example of well-thought-out metadata practices, and provides a "Best Practices Primer" with an explanation of the data life cycle and related tool kit. Thinking through the life cycle will help identify points at which copyright could or should be addressed.

Finally, Databib (n.d.) describes itself as "a tool for helping people identify and locate online repositories of research data. Users and bibliographers create and curate records that describe data repositories that users can search."

Creative Commons and Open Data Commons

Creative Commons licenses may be used to describe rights and permitted uses in an internationally recognized manner, reducing complexity and encouraging global use and reuse. The Creative Commons framework is constructed to work globally. At its most basic:

> Creative Commons is a non-profit organization that created a set of simple, easy-to-understand copyright licenses. These are legally enforceable licenses that allow creators to mark a work with permission to make a variety of uses, with the aim of expanding the range of things available for others to share, quote, adapt, and build upon. Creative Commons licenses do two things: They allow creators to share their work easily, and they allow everyone to find work that is free to use without permission. As long as you obey the terms of the license attached to the work, you can use Creative Commons licensed material without fear of accidentally infringing someone's copyright.

> We encourage the use of Creative Commons licenses because they effectively help communicate information about copyright holders' intentions and thus help everyone know with clarity what may be used and how – and what requires permission. They help authors and creators manage their copyrights and share their creative work without losing control over it. Further, Creative Commons licenses facilitate creators' rights by communicating clearly a contact for permission when appropriate. (University of Michigan Copyright Office, 2013)

Open Data Commons (ODC) provides a Creative Commons-style framework for managing data based on the principle that "open data is data that anyone is free to use, reuse and redistribute without restriction (except, perhaps the requirements to attribute and share alike)" (Open Knowledge Foundation, n.d.a). ODC provides legal solutions for open data. In March 2008, it launched the first open data license: the Public Domain Dedication and License (PDDL). ODC provides licenses tailored to the data environment that are significant tools for expressing research intentions regarding copyright elements of datasets. (These are described succinctly in "Introduction to Intellectual Property Rights in Data Management," in Hirtle, 2011, and Open Knowledge Foundation, n.d.b.)

All Things Open and Commercial Use of Public Domain Data: Rethinking Assumptions

Creative Commons, Open Data Commons, and other open approaches to licenses all provide tools for describing and managing copyright in research data. The Creative Commons model is well established, with a range of standard, easily expressed and understood, legally enforceable licenses (Carroll, 2013). While data as fact is not subject to copyright under U.S. copyright law, datasets and databases may be protected as compilations. With the goal of maintaining an open research framework, one would not want to assert such rights in a way that encumbers the underlying data. One option is to apply a public domain license to bodies of work. The problem with doing this is that the underlying data may not be in the public domain or may be affected by legal duties other than copyright. Assuming all of the elements are in the public domain or are yours to give, placing work in the public domain may actually result in *loss* of control.

The data may become disassociated with the creator and relevant metadata. Another approach: follow the model of open-source-style licenses to data and datasets to facilitate new and flexible uses of the underlying material without risking that it will be functionally removed from the public domain (made proprietary by someone else) or disassociated from citable information.

Getting Legal Advice

Information professionals need the right lawyers, people who understand the different areas of law and the scholarly endeavor. Because copyright is so pervasive, it is treated increasingly as a do-it-yourself subject. Free videos and a plethora of "copyright education" are available from a growing range of sources. Geared for copyright nonspecialists to make them "aware" of copyright basics, many of these resources may emphasize or deemphasize different aspects of the law for any number of reasons. The idea that nonspecialists rolling up their sleeves can address complex copyright or other legal questions is problematic; I did not always have this view. But one would not suggest skimping on the expected training and experience of programmers or other technical expertise. One of the great problems is the way copyright is now ubiquitous and yet not at all intuitive. Ideally, you can or should nurture good relations with your general counsel's office or other legal expertise on your campus. It may be that there are different attorneys in that office with a variety of related expertise from copyright to privacy. Encourage them to join the discussion at your library and make the case on their behalf that it is worth the time for them to develop practical understanding of the issues so they can participate in a real, ongoing conversation with you about data issues. Ideally, they will be in the role of traditional counselors to help you think through options in a legal context and constructively move your work forward. Managing research data and data policy requires a team with sufficient range of expertise and experience; there should be an emphasis on continuous learning and problem solving. An ideal team will include people at staggered levels of experience that reflects a diversity of skills and knowledge. Think of the legal experts as an integral part of the team and assume everyone has something to learn from each other.

NEW LIBRARY SERVICES

A growing number of universities offer resources and expertise for data management support for researchers. These resources often take the form of consulting services through libraries or information provided from grants or sponsored research offices. Grant requirements set baselines for data plans. These plans typically expect the applicant to describe how they will address intellectual property matters, constructively forcing copyright questions to the forefront.

The multiple areas of legal responsibility are a significant issue in managing research data. There are a growing number of librarians stepping into the needs gap to bring together relevant information to assist researchers. The scale and complexity of the legal issues make them difficult to tackle, so in many cases basic checklists are a first step. In discussing this chapter with colleagues, a librarian said, "One of the common problems is that managing research data—in addition to managing one's research—is complicated. Universities in their effort to facilitate new requirements are developing matrices and tools for obligations, say, for storing sensitive data with an emphasis on providing the information and on university legal obligations." This is a reasonable place to start, but individual researchers will need help to follow and implement those resources. Researchers are unlikely to be aware of specific terms of service or security requirements and may need help to ensure that sensitive data is appropriately maintained. Beyond copyright, researchers and universities are responsible for meeting university policies, privacy, and other requirements. For example, a researcher may need to store sensitive data. Librarians can advise on how to store data in a manner that does not compromise security or other applicable policies. (See, generally, Sample Data Management Plan Resources on the following page.)

In trying to make sense of these responsibilities, we also see perceived concerns about the difference between the researcher's responsibility and the legal responsibility of the university. "Tools" that are essentially lists of laws, extraordinarily complex laws at that, fail to help the researcher, the university, or the utility of research data. The advice often comes in the form of web-based lists of information, often from a sponsored research office or an IT department. These kinds of guides are essential steps, but there is a need for human expertise—not just in the area of research—but also in application of law and administrative concerns.

SAMPLE DATA MANAGEMENT PLAN RESOURCES

All sites last visited July 8, 2013.

Best Practices, DataONE: http://www.dataone.org/best-practices

Columbia University Libraries, Information Services: http://scholcomm.columbia.edu/data-management/data-management-plan-templates/

Data Plan Guidance and Examples, Digital Curation Centre: http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples

Data Management Plan Examples, University of Minnesota Libraries: https://www.lib.umn.edu/datamanagement/DMP/example

Data Management Plan Examples, Yale Digital Collections Center: http://ydc2.yale.edu/documentation/data-management-plan-examples

Data Management Plans, Data Management and Publishing, Subject Guides, MIT Libraries: http://libraries.mit.edu/guides/subjects/data-management/plans.html

Data Management Planning, DataONE: http://www.dataone.org/data-management-planning

Data Management Sample Plans, University of North Carolina: http://www.irss.unc.edu/odum/contentSubpage.jsp?nodeid=570

DataONE Example Data Management Plan, NSF BIO: http://www.dataone.org/sites/all/documents/DMP_Copepod_Formatted.pdf

Example Data Management Plans, the University of California-San Diego, Research Cyberinfrastructure: http://rci.ucsd.edu/dmp/examples.html

Generic Data Management Plan Outline and Examples, Digital Data Management, Curation and Archiving, Research Guides at the University of New Mexico: http://libguides.unm.edu/content.php?pid=137795&sid=1422879

NSF Data Management Plan Template, the University of Chicago, Division of the Physical Sciences: http://psd.uchicago.edu/NSF%20Data%20Management%20Plan%20Template.pdf

Preparing Data Management Grant Plans for NSF Grant Applications, the University of California-Berkeley, Science Libraries: http://www.lib.berkeley.edu/CHEM/data/nsf/nsf_dmp.pdf

Research Data Management and Publishing Support, the University of Michigan Library: http://www.lib.umich.edu/research-data-management-and-publishing-support

University of Pittsburgh—NSF Data Management Plan—Example 1: http://www.pitt.edu/~offres/policies/NSF-DMP-Examples.pdf

What Should Be in a Data Management Plan? Research Data Management, the University of Oxford: http://www.admin.ox.ac.uk/rdm/dmp/plans/

Libraries and librarians are stepping in as a natural support for data-related needs. As information specialists, research libraries already are a source for help with data citation standards and metrics. They are extending their outreach to include workflow solutions and data management advice; data storage solutions; and consulting services. There is a growing role for librarians as advisors and liaisons with information on government and business. Librarians also naturally work to complement each other's knowledge so that a single research project may functionally benefit from the collective knowledge of dozens of librarians and their networks.

THE UNIVERSITY OF MICHIGAN LIBRARY

Data management and data curation discussions on the University of Michigan campus have been building for years, and more recently, the University Library has moved toward engaging with data storage, management, and the concept of data curation. The University Library has been involved with discussions on a number of levels, from IT rationalization committees to the creation of resources for reviewing data management plans. Associate University Librarian for Research Elaine Westbrook leads the campus conversations on data management and curation. Today, the University Library offers a range of services for researchers at various stages of their research, from planning studies to the dissemination of their results. A comprehensive set of services that support the full life cycle of research data for faculty and students is at an early stage of development. Director of Research Data Services (RDS) Jen Green also serves as head of Science, Engineering, and the Clark Library. RDS is a network of people and services preparing a roadmap for library services around data and the research life cycle.

As part of developing strategies toward establishing library data services, Council on Library and Information Resources/Digital Library Federation (CLIR/DLF) Data Curation Fellows Natsuko Nicholls and Fe Sferdean are conducting internal and external environmental scans—including a campus data needs assessment—to determine the status of the University of Michigan among academic peers and in the national data conversation. "Data services are increasingly being recognized as a crucial component of library services," says Nicholls. "While researchers are grappling with a deluge of data, libraries are trying to find ways to improve data storage, management, sharing, and preservation. We are looking for

ways for researchers and libraries to work together toward the common goals of facilitating innovative, reproducible research and increasing opportunities for data sharing and re-use" (Frick, 2013).

CONCLUSION

Copyright should be considered as part of the information infrastructure at the outset of a research project, along with a data plan as part of the research proposal. Infrastructure is more than storage and computational capacity. Copyright and intellectual property policy is integral to these discussions and resulting analyses. By involving people with expertise in copyright, data, the research process, and data management in those conversations, we can ensure the best possible framework for copyright and research behaviors around data for the research enterprise. The nature of research data and the possibilities for use and reuse globally make it imperative that participants be informed and engaged in the growing body of literature. While resources are limited, it is valuable to have staff attend professional conferences and actively involve themselves in working committees (national and global) because the conversation and problem solving is personal and immediate. Staff members need to report out to their colleagues, and they need to share experiences and observations. Closing the gap between available data and usable data demands painstaking, self-aware discipline; it is a worthwhile effort to ensure that data is truly the new gold rather than fool's gold.

REFERENCES

Australian Research Council. (2013, January 10). ARC open access policy. *Commonwealth of Australia*. Retrieved from http://www.arc.gov.au/applicants/open_access.htm

Carroll, M. W. (2013). Creative Commons and the openness of open access. *New England Journal of Medicine, 368*(9), 789–791. http://dx.doi.org/10.1056/NEJMp1300040

Databib. (n.d.). About Databib. Retrieved from http://databib.org/about.php

DataCite. (n.d.). What do we do? Retrieved from http://www.datacite.org/

DataONE. (n.d.a). Best practices. Retrieved from http://www.dataone.org/best-practices

DataONE. (n.d.b). Data management planning. Retrieved from http://www.dataone.org/data-management-planning

DataONE. (2011). *Example data management plan: NSF bio.* Albuquerque, NM: DataONE. Retrieved from http://www.dataone.org/sites/all/documents/DMP_Copepod_Formatted.pdf

Digital Curation Centre. (n.d.). Data plan guidance and examples. Retrieved from http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples

Frick, R. (2013, January 24). A roadmap for data services [Web log post]. Retrieved from http://connect.clir.org/clir/blogs/blogviewer?BlogKey=d2df07ab-5587-49d0-8805-a9cd5dc83d87

Hartshorne, J. K., & Schachner, A. (2012). Tracking replicability as a method of post-publication open evaluation. *Frontiers in Computational Neuroscience, 6*(8). http://dx.doi.org/10.3389/fncom.2012.00008

Hirtle, P. (2011). Introduction to intellectual property rights in data management. *Cornell University Research Data Management Service Group.* Retrieved from https://confluence.cornell.edu/display/rdmsgweb/introduction-intellectual-property-rights-data-management

Holdren, J. (2013, February 22). Memorandum for the heads of executive departments and agencies. Executive Office of the President, Office of Science and Technology Policy. Retrieved from http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

Kroes, N. (2011, December). Data is the new gold. Opening remarks given at the Press Conference on Open Data Strategy, Brussels, Belgium. Retrieved from http://ec.europa.eu/information_society/newsroom/cf//itemdetail.cfm?item_id=7668

MIT Libraries. (n.d.). Data management plans. Retrieved from http://libraries.mit.edu/guides/subjects/data-management/plans.html

Mooney, H., & Newton, M. (2012). The anatomy of a data citation: Discovery, reuse, and credit. *Journal of Librarianship and Scholarly Communication, 1*(1), eP1035. http://dx.doi.org/10.7710/2162-3309.1035

National Institutes of Health. (n.d.). NIH public access policy details. Retrieved from http://publicaccess.nih.gov/policy.htm

National Institutes of Health (NIH). (2013). Frequently asked questions about the NIH public access policy. Retrieved from http://publicaccess.nih.gov/FAQ.htm

National Science Foundation (NSF). (2010). Scientists seeking NSF funding will soon be required to submit data management plans (NSF Press Release 10-077). Retrieved from http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928

National Science Foundation (NSF). (2011). NSF-ENG: COV reports and annual updates. Retrieved from http://www.nsf.gov/eng/general/dmp.jsp

Open Access @ UNT. (n.d.). The Denton Declaration: An open access manifesto. Retrieved from http://openaccess.unt.edu/denton_declaration

Open Knowledge Foundation. (n.d.a). Open Data Commons Open Database License (ODbL). Retrieved from http://www.opendatacommons.org/licenses/odbl/

Open Knowledge Foundation. (n.d.b). Open Data Commons. Retrieved from http://opendatacommons.org/

ORCID. (n.d.). ORCID: Connecting research and researchers. Retrieved from http://orcid.org/

Physical Sciences Division. (n.d.). Template for NSF data management plan. *University of Chicago*. Retrieved from http://psd.uchicago.edu/NSF%20Data%20Management%20Plan%20Template.pdf

Piwowar, H. A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLOS One, 6*(7), e18657. http://dx.doi.org/10.1371/journal.pone.0018657

Research Cyberinfrastructure. (n.d.). Example data management plans. *UC San Diego*. Retrieved from http://rci.ucsd.edu/dmp/examples.html

Research Data Management. (n.d.). What should be in a data management plan? *University of Oxford*. Retrieved from http://www.admin.ox.ac.uk/rdm/dmp/plans/

Savage, C. J., & Vickers, A. J. (2009). Empirical study of data sharing by authors publishing in PLoS journals. *PLoS ONE, 4*(9), e7078. http://dx.doi.org/10.1371/journal.pone.0007078

Scholarly Communication Program (n.d.). Data management plan templates. *Columbia University Libraries/Information Services*. Retrieved from http://scholcomm.columbia.edu/data-management/data-management-plan-templates/

Science Libraries. (n.d.). *Preparing data management grant plans for NSF grant applications: A guide to the NSF policy for data management*

plans. Berkeley: University of California, Berkeley. Retrieved from http://www.lib.berkeley.edu/CHEM/data/nsf/nsf_dmp.pdf

Stebbins, M. (2013, February 22). Expanding public access to the results of federally funded research [Web log post]. Retrieved from http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research

University of Michigan Copyright Office. (2013, March 13). Creative Commons. *University of Michigan Library*. Retrieved from http://www.lib.umich.edu/copyright/creative-commons

University of Michigan Library. (n.d.). Research data management and publishing support. Retrieved from http://www.lib.umich.edu/research-data-management-and-publishing-support

University of Minnesota Libraries. (n.d.). Data management plan examples. Retrieved from https://www.lib.umn.edu/datamanagement/DMP/example

University of New Mexico Libraries. (n.d.). Generic data management plan outline and examples. Retrieved from http://libguides.unm.edu/content.php?pid=137795&sid=1422879

University of North Carolina. (n.d.). Data management sample plans. Retrieved from http://www.irss.unc.edu/odum/contentSubpage.jsp?nodeid=570

University of Pittsburgh. (n.d.). University of Pittsburgh—NSF data management plan—Example 1. Retrieved from http://www.pitt.edu/~offres/policies/NSF-DMP-Examples.pdf

Vlaeminck, S. (2012, December 12). Research data management in economic journals (Part I) [Web log post]. Retrieved from http://www.edawax.de/2012/12/research-data-management-in-economic-journals-part-i/#more-1245

Wellcome Trust. (n.d.). Open access at the Wellcome Trust. Retrieved from http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/index.htm

World Bank. (2012, April 10). World Bank announces open access policy for research and knowledge, launches Open Knowledge Repository. Retrieved from http://web.worldbank.org/WBSITE/EXTERNAL/NEWS/0,,contentMDK:23164491~pagePK:64257043~piPK:437376~theSitePK:4607,00.html

Yale Digital Collections Center. (n.d.). Data management plan examples. Retrieved from http://ydc2.yale.edu/documentation/data-management-plan-examples

> persistent identifiers for datasets, workflows and standards for data publication; support journal publishers by enabling research articles to be linked to the underlying data. Currently we are working primarily with organizations that host data, such as data centres and libraries.

The website for DataONE (n.d.) describes the organization as: "the foundation of new innovative environmental science through a distributed framework and sustainable cyberinfrastructure that meets the needs of science and society for open, persistent, robust, and secure access to well-described and easily discovered Earth observational data." DataONE provides a substantive example of well-thought-out metadata practices, and provides a "Best Practices Primer" with an explanation of the data life cycle and related tool kit. Thinking through the life cycle will help identify points at which copyright could or should be addressed.

Finally, Databib (n.d.) describes itself as "a tool for helping people identify and locate online repositories of research data. Users and bibliographers create and curate records that describe data repositories that users can search."

Creative Commons and Open Data Commons

Creative Commons licenses may be used to describe rights and permitted uses in an internationally recognized manner, reducing complexity and encouraging global use and reuse. The Creative Commons framework is constructed to work globally. At its most basic:

> Creative Commons is a non-profit organization that created a set of simple, easy-to-understand copyright licenses. These are legally enforceable licenses that allow creators to mark a work with permission to make a variety of uses, with the aim of expanding the range of things available for others to share, quote, adapt, and build upon. Creative Commons licenses do two things: They allow creators to share their work easily, and they allow everyone to find work that is free to use without permission. As long as you obey the terms of the license attached to the work, you can use Creative Commons licensed material without fear of accidentally infringing someone's copyright.