

Building Outreach on Assessment: Researcher Compliance with Journal Policies for Data Sharing

by Kathleen Fear

Research Data Access and Preservation

EDITOR'S SUMMARY

Discouraged by researchers' low response to informational offerings on data sharing, librarians at the University of Rochester's River Campus Libraries refined their outreach strategies. To understand the support that faculty members may need, they checked the data sharing policies of journals that faculty members publish in and the level of compliance. A review of journal policies and evidence of data sharing showed both requirements and compliance varied widely. The university librarians plan to offer workshops specifically on data sharing requirements for PLOS One, the journal researchers published in most in 2014. Researchers with strong data sharing practices will be encouraged to mentor colleagues. Ongoing compliance assessment will point out where further library support may be effective.

KEYWORDS

editorial policy
information sharing
data sets
evaluation
outreach services
educational activities

The University of Rochester's River Campus Libraries have been actively developing services to meet researchers' data management needs. Many of these initiatives have been very popular, with the marked exception of outreach around data sharing. Workshops focusing on how to share data have gone unattended. Other efforts, like a series of articles in a campus newsletter to researchers, pose a challenge to assessment: the information is out there, but there is no easy and effective way to measure its impact. This lack of uptake was both disappointing and perplexing, given the increasing emphasis on data sharing from publishers and funders.

Conversations with faculty provided some insight into their lack of interest in guidance on data sharing. Many believed that they already shared their data effectively, while others were adamant that data sharing does not apply to them. The root of the problem seemed to be that, when it came down to it, *data sharing* was too broad a category to be of interest to these researchers. A topic like developing a data management plan is concrete, and our outreach materials use the same language as the funders who require such plans. This specificity makes it easy for researchers to recognize that they have a need and that the library is trying to help them meet it. *Data sharing*, by contrast, is nebulous, and different entities talk about sharing and openness in different ways, which poses a challenge in connecting with researchers.

As a result, we changed our concentration from education and outreach on data sharing generally to trying to find opportunities for targeted outreach. At the same time, our library was making a shift: the librarians transitioned from subject librarians to outreach librarians, reflecting the libraries' focus on building relationships with faculty, students and

Kathleen Fear is the data librarian at the University of Rochester. She can be reached at kathleen.fear@rochester.edu.

departments. These two factors together prompted a collaborative project among the data librarian and the outreach librarians for computer science (Fang Wan); brain and cognitive sciences, linguistics and public health (Judi Briden); chemistry (Sue Cardinal); physics, optics and astronomy (Tyler Dzuba); and biology, math and statistics (Diane Cass). The project came together around two goals: develop effective strategies for getting data sharing information to researchers and strengthen outreach librarians' relationships with their respective departments as well as boost their expertise in data support.

Methods

We identified journal data sharing policies and researchers' compliance with those policies as a promising area of study. First, the outreach librarians already field questions from faculty about journals and their author guidelines, so examining their data policies represented a natural extension of work they commonly do. Additionally, assessing compliance with journal data policies is relatively straightforward. Looking at compliance with data management plans is complicated both by the long timespan between when the plan is put into place and when compliance can be assessed as well as by the breadth and diversity of those plans. Journal data policies are comparatively much simpler to assess, since typically authors must comply with the policy at the time of publication, and the same policy applies to all papers published in the journal, rather than being fine-tuned to individual projects.

We carried out the project in three stages: identifying the journals our faculty commonly publish in; reviewing those journals' data sharing policies (or lack thereof); and evaluating whether and how well the authors met the requirements of those policies. To build the list of journals, we extracted from *Web of Science* all publications with at least one University of Rochester-affiliated author between January 1 and December 31, 2014. The University of Rochester has a heavy science and engineering focus, so while *Web of Science* does not have perfect coverage, we believed that the 2,784 articles across 1,181 journals it identified represented a reasonable sample set for this exploratory project.

From the 1,181 total journals represented in the article sample, we focused our policy review on the 109 journals in which Rochester researchers published five or more times in 2014. Of those journals, 43 required or explicitly encouraged data sharing. The compliance review narrowed the sample further, examining 161 articles across 13 journals with data sharing policies.

Good Policy, Good Sharing

In order to carry out the review we planned, we needed to define what we were looking for in both the policy review and the compliance review steps. In other words, we needed to determine, for our purposes, what a good data sharing policy looked like and what good compliance looked like.

Because the eventual goal of the project was to develop outreach strategies, we assessed journals' data sharing policies through the lens of outreach and instruction. We were especially interested in how much assistance researchers might need to comply with the plans as written – and, in turn, how the library might provide that assistance – so our review criteria focused on the comprehensiveness of the policies. We evaluated whether the policies specified how to share the data, how to indicate in the paper that the data were available and when the data should be accessible to others. This approach is a departure from other studies of data sharing policies, which focus on how enforceable those policies are (Piwowar and Chapman, 2008, [1] and Sturges et al., 2014 [2]), particularly whether the journals require authors to submit an accession number or other identifier as proof of deposit.

It is worthwhile to consider the comprehensiveness of a policy and the quality of that policy separately. A policy can be – and indeed, many were – comprehensive, clear and straightforward to comply with, without requiring or encouraging good data sharing practices and without enabling any means of enforcement. For example, *Organometallics* requires that crystallographic data be submitted as supporting information and specifically states that depositing that data at the Cambridge Crystallographic Data Center, a widely used and well-respected repository in that field, does not fulfill that journal's data sharing requirement. Others have comprehensive guidance for

have comprehensive guidance for certain data types but not others. This complicated our analysis of compliance with the policies. Because fully complying with a journal's data sharing policy was no guarantee that the data were shared effectively (that is, openly, in a usable format or linked to the paper), we developed a three-tier rating system for reviewing articles.

Articles received the highest rating if the data were linked directly from the paper, if the data were included in the paper or supplemental information in a usable format (not just as a pdf) or if the authors provided a clear justification for why the data could not be made accessible. A second level rating flagged articles that, while technically in compliance with the appropriate journal's policy, fell short of good data sharing practice. Examples include articles that indicated that the data were available, but with no link, accession number or other connection to where the data were held; those that promised that data were available on request only; and those that included data in the paper or supplement, but in an unusable format. The lowest rating was reserved for papers with no sign that the data were available.

Findings and Putting Them to Use

Our findings confirmed that there is a need to better educate researchers on sharing their data. Half the articles reviewed received the lowest rating, indicating that there was no way to find or access their data. More positively, though, of the papers that did share data, 60% received the highest rating.

There did not appear to be any relationship between how comprehensive the journal's data sharing policy was and how well authors complied with that policy. All but two journals had both compliant and non-compliant articles. The two exceptions were *Monthly Notices of the Royal Astronomical Society* and *Earth and Planetary Sciences Letters*, both of which had no non-compliant articles. Notably, those two journals also had two of the least comprehensive data policies. This reflects that journal data policies are written with an assumption of a certain amount of disciplinary knowledge. In fields like astronomy and earth science, where data sharing practices are more well-established, journals may include less guidance in

their policies on the assumption that, if authors are publishing work in that field, they have already been socialized into the field's practices around data sharing.

These findings open several avenues for outreach. Based on our findings, we can segment departments and researchers into several categories: those who are not sharing data but are supposed to; those who share, but not as well as they could; and those who are exemplary in their data sharing practices. Currently, we are working to design and test out targeted outreach strategies for each of these populations.

Reaching Recalcitrant Sharers. The journal in which researchers published the most in 2014 was *PLOS One*. This is also a journal with a rigorous and comprehensive data sharing policy. However, nearly 60% of the articles appearing in that journal did not share their data. Partly, this lack of sharing is a timing issue: *PLOS's* data sharing policy only went into effect in April 2014, so many of the papers published last year were submitted prior to that time. But given the popularity of the journal on this campus, its interdisciplinary nature and its relatively high prestige, we plan to focus on it going forward, offering workshops on publishing in *PLOS* journals (with data sharing advice as an integral part) as well as directly contacting authors who published there previously with guidance and resources for sharing their data the next time they plan to publish there.

Helping Bare-Bones Sharers Do Better. Overall, the number of papers where data was shared well was higher than those that did the bare minimum to share their data, and all but three journals had at least one paper that received the highest rating in our compliance review. We plan to speak first with authors who do a good job sharing and enroll them to encourage their colleagues to do so as well. Working in partnership with those researchers is an opportunity to expose researchers to some peer pressure while at the same time offering examples of how to share data effectively and resources for doing so.

Learning from Exemplars. In our sample, the best data sharers were from earth science and astronomy, both fields whose journals had limited guidance on data sharing in their policies. The lack of guidance in their

FEAR, continued

journal policies reflects a seemingly correct assumption that authors know what to do with their data, so our outreach to these exemplar researchers will focus on learning from them, especially on how they become acculturated into the data sharing practices in their fields. While experienced researchers may understand what they need to do with their data, it is less clear that graduate students publishing for the first time have the same knowledge. Studying how researchers gain that knowledge may open an opportunity for the library to support that process or to foster the same kind of learning in other fields. We plan to conduct interviews with researchers in these fields with an eye toward understanding their practices as well as building the libraries' relationship with those faculty.

Future Directions

We plan to continue to monitor how well researchers are sharing their data, as well as how journals' data sharing policies evolve. Our ongoing goal is not to get directly involved in enforcing researchers' compliance with those policies, but rather to assess compliance as a way of identifying where support from the library might be welcome as well as determining in what form that support might be most effective. We expect that the more closely we can target our outreach efforts to the point of need (and, where necessary, make sure researchers are aware they have a need), the better uptake of our services we will see – and hopefully, the better job our researchers will do sharing their data. ■

Resources Mentioned in the Article

- [1] Piwowar, H. A., & Chapman, W. W. (2008). A review of journal policies for sharing research data. *Nature Precedings*. Retrieved from <http://precedings.nature.com/documents/1700/version/1/files/npre20081700-1>
- [2] Sturges, P., Bamkin, M., Anders, J., & Hussain, A. (2014). Journals and their policies on research data sharing [Web log post]. Retrieved from <https://jordproject.wordpress.com/reports-and-article/journals-and-their-policies-on-research-data-sharing/>