EDITOR'S SUMMARY

With reproducibility of research becoming a leading issue in academia, libraries are examining their role in promoting data and information transparency. The National Science Foundation's requirement for data management plans in research projects, grant applications stressing evidence of unbiased results and scholars' demands for standards for reproducibility together highlight the need for attention to the issue. Libraries increasingly seek staff with skills to support their data repositories, curation and data management services, and library and information science programs are responding to the growing need for specialization, including courses in research reproducibility. A ready solution may be for librarians to develop best practices for research transparency and methods to document research workflow. As the demand grows for evidence that research can be replicated, libraries are well positioned not only to manage research data properly but also to enable it to be analyzed for reproducibility.

KEYWORDS

reproducibility
research libraries
professional competencies
replicative studies
research methods

Is Research Reproducibility the New Data Management for Libraries?

by Cynthia R.H. Vitale

Research reproducibility has become a hot topic among academics in the last few years. With organizations such as Retraction Watch cataloging retractions of peer-reviewed literature, replication studies finding many research outcomes to not be reproducible [1, 2] and journals signing on to transparency polices [3, 4], strategies to address these topics have been at the forefront of much academic discussion. In response, many libraries are beginning to evaluate what role they may play in improving the reproducibility of the research conducted on their campuses. Though still mostly in the exploratory phase, this interest by libraries has, in many ways, resembled the growth of research data management services. What follows is an analysis of the current state of research data and research reproducibility movements in libraries, focusing on the catalysts for services, library-staffing strategies and services provided.

Catalysts for Change

In the years before the National Science Foundation (NSF) released its data management plan (DMP) requirement, libraries and library organizations were building socio-technical infrastructure for data management services, and more broadly, E-Science support, in the information science profession. Major professional organizations, such as the Association for Information Science and Technology (ASIS&T), the Association of Research Libraries (ARL) and the American Library Association (ALA) established initiatives focused on this topic [5]. Ideologically, studies have argued, data management is similar to information management and is something libraries and librarians know much about [6, 7]. Thus, when the NSF announced the DMP requirement in 2010, university libraries took it upon themselves to develop services to support their researchers in this area.

In contrast, the federal funding requirements for reproducibility are spread across numerous notices and guidelines. One notice released in October 2015 by the National Institutes of Health (NIH) and the Agency for Healthcare Research and Quality (AHRQ) updated proposal instructions and review language under the *Implementing Rigor and Transparency in NIH & AHRQ Research Grant*

Cynthia R.H. Vitale is the digital data librarian in data & GIS services at Washington University in St. Louis libraries. In this position, Cynthia leads research data services and curation efforts for the libraries. Since coming into this role in 2012, she has worked on faculty projects to facilitate data sharing and interoperability while meeting faculty research data needs throughout the research lifecycle. She has also worked across the university to improve research reproducibility, addressing both technical and cultural barriers. She currently serves as the visiting program officer for SHARE with the Association of Research Libraries.

Applications notice. In brief, the updates ask faculty to describe the experimental design and methods proposed in the research strategy section of the proposal and to indicate how they will achieve robust and unbiased results. In complying with this requirement, the researcher establishes a trail of verifiability, which may be considered a step towards reproducibility. In December 2015, the NIH and AHRQ released *Advance Notice of Coming Requirements for Formal Instruction in Rigorous Experimental Design and Transparency to Enhance Reproducibility*. This notice, effective in 2017, will require institutional training grant and institutional career development applications to include a plan to ensure the training programs provide skills necessary to design and conduct rigorous experiments. For individual fellowship applications, this notice will require the researchers to articulate their methods for ensuring rigorous research to ensure reproducibility. In addition to these two federal examples, as mentioned earlier, a great deal has been published recently by scholars calling for greater standards for reproducibility and revealing inabilities to replicate studies in their fields [8, 9].

Staffing

Social science librarians have provided data related services for years, but the growth of librarians specifically for research data management services has been significant to say the least. Recent evaluations of data management-related job announcements have highlighted the expectations many universities have in the skills a single librarian must possess to provide data related services [10, 11]. Luckily, though, as data management services grow, many university libraries are moving more resources and staff into this burgeoning area and even retooling liaison librarians to add this skill to their toolkits [12].

Given the newer focus of reproducibility in libraries, staffing for this role specifically is still relatively limited. New York University Libraries has established one of the few known reproducibility positions, which is also split with research data management [13]. Recently, library and information science schools have also added faculty, such as Victoria Stodden at the University of Illinois, and courses in research reproducibility to master's and Ph.D. programs. Thus, the library domain may expect more librarians intentionally trained in this area in the coming years.

Services

Following a common roadmap of sorts, libraries determined what data management services to offer by first conducting surveys and data management need assessments among their faculty members [14]. As services sprang up to address the needs discovered, institutional data repositories and curatorial practices evolved and continue to develop as viable storage and discovery layers for research data created at an institution [15]. Other outcomes of this movement have been libraries offering consultations and workshops on data management planning, building databases for faculty projects and actively managing data, among others [16].

Turning the federal funding update on rigor and transparency, as well as the local groundswell for improved protocols for reproducibility, into library services is not hard to imagine. Librarians could collaborate locally or with non-profits such as the Center for Open Science or the Center for Scientific Integrity to create documentation on best practices for research transparency in specific domains and offer workshops on tools that help document the research workflow. Indeed, some libraries have partnered with the research office or research computing departments

on campus to bring outside speakers to campus to discuss methods and tools for improving reproducibility [17].

Discussion

Whether libraries can claim that their existing knowledge prepares them to provide reproducibility services has not been fully explored. While it is true that libraries are well positioned as neutral in the academic landscape to provide this support, to understand what makes research fully reproducible requires domain knowledge, perhaps more than a subject specialty provides. But it can also be argued that a significant portion of reproducibility has to do with proper data management and making data resulting from research widely available. Many data management librarians are well acquainted with these practices. Perhaps more than being a new stand-alone service, though, research reproducibility will develop into an extension or additional offering in the suite of services provided by research data, subject liaison or scholarly communication librarians. Undoubtedly, research reproducibility is not a topic or concern that will go away, though library support for faculty in this domain remains to be fully realized. ∎

Resources Mentioned in the Article

[1] Ioannidis, J. P. A. (August 30, 2005). Why most published research findings are false. *PLoS Medicine 2*(8): e124. doi:10.1371/journal.pmed.0020124

[2] Henderson, V. C, Demko, N., Hakala, A., MacKinnon, N., Federico, C.A., Fergusson, D., & Kimmelman, J. (October 15, 2015) Threats to valid clinical inference in preclinical research of sunitinib. *eLife.* doi:10.7554/eLife.08351

[3] Center for Open Science. Transparency and openness promotion (top) guidelines. Retrieved from https://cos.io/top/

[4] Data Access and Research Transparency (DART): www.dartstatement.org/

[5] Gold, A. (September/October 2007). Cyberinfrastructure, data, and libraries, Part 1. *D-Lib Magazine, 13*(9/10). Retrieved from www.dlib.org/dlib/september07/gold/09gold-pt1.html#1

[6] Salo, D. (2010). "Retooling libraries for the data challenge." *Ariadne.* Issue 64. Retrieved from www.ariadne.ac.uk/issue64/salo/

[7] Hey, T., & Hey, J. (2006). E-science and its implications for the library community. *Library Hi Tech, 24*(4), 515-528.

[8] Open Science Collaboration (August 2015). Estimating the reproducibility of psychological science. *Science, 349*(6251). doi:10.1126/science.aac4716

[9] Laine C., Goodman, S. N., Griswold, M. E., & Sox, H. C. (2007). Reproducible research: Moving toward research the public can really trust. *Annals of Internal Medicine, 146*(6), 450-453.

[10] Shorish, Y. (July 16, 2015). Data, data everywhere…but do we want to drink [blog post]? *ARCL TechConnect.* Retrieved from http://acrl.ala.org/techconnect/post/data-data-everywherebut-do-we-want-to-drink

[11] Johnson, A. (May 12, 2015). Hiring data librarians. Retrieved from www.scribd.com/doc/265015825/Hiring-Data-Librarians

[12] Cox, A, Verbaan, E., & Sen, B. (November 30, 2012). Upskilling liaison librarians for research data management. *Ariadne.* Issue 70. Retrieved from www.ariadne.ac.uk/issue70/cox-et-al#sthash.ZZFfhJMi.dpuf

CONTENTS   TOP OF ARTICLE   < PREVIOUS PAGE   NEXT PAGE >

Resources Mentioned in the Article, continued

[13] Steeves, V., & Wolf, N. (2015). New services in research data management and planning at NYU Libraries. *Connect: Information Technology at NYU.* Retrieved from https://wp.nyu.edu/connect/2015/10/20/new-services-data-research-management/

[14] Kouper, I., Akers, K. G., Nicholls, N. H., & Sferdean, F. C. (2013). A roadmap for data services. *Proceedings of the ACM/IEEE-CS Joint Conference on Digital Libraries, 13.* 375-376. doi:10.1145/2467696.2467763

[15] Johnston, L. R. (2014). A workflow model for curating research data in the University of Minnesota Libraries: Report from the 2013 Data Curation Pilot. *University of Minnesota Digital Conservancy.* Retrieved from http://hdl.handle.net/11299/162338

[16] Flores, J. R., Brodeur, J. J., Daniels, M. G., Nicholls, N., & Turnator, E. (2015). Libraries and the research data management landscape. In J. C. Maclachlan, E. A. Waraksa, & C. Williford (Eds.). *The Process of Discovery: The CLIR Postdoctoral Fellowship Program and the Future of the Academy* (pp. 82-102). Washington DC: Council on Library and Information Resources. Retrieved from www.clir.org/pubs/reports/pub167/pub167.pdf#page=88

[17] University of Kansas Libraries. (2015). Workshop for increasing openness and reproducibility in quantitative research. Retrieved from https://lib.ku.edu/reproducible-research