"Librarian's Role in Reproducibility of Research" 2017 Symposium: Research Section Stipend Winner Reflection

Kimberly R. Powell, MLIS
Life Sciences Informationist
Woodruff Health Sciences Center Library
Emory University
Atanta, GA
krpowel@emory.edu

Introduction

At the 117th Annual Meeting and Exhibition of the Medical Library Association (MLA) in 2017, eight sections of the association jointly sponsored a symposium entitled Librarian's Role in Reproducibility of Research. The four hour symposium was held Saturday, May 27th as part of the MLA pre-conference activities. Shona Kirtley, Knowledge and Information Manager for the EQUATOR Network at the Centre for Statistics in Medicine at the University of Oxford, served as the keynote speaker. Invited panelists discussed their roles in initiatives aimed to reduce the research waste caused by irreproducible reporting of scientific efforts within the published literature. The four-hour symposium concluded with a hands-on brainstorming activity that asked each of the attendees to propose and reflect on increasing the reproducibility of science. The following is a summary of the information shared and a reflection on the brainstorm suggestions made at the Librarian's Role in Reproducibility of Research Symposium. A LibGuide for the event, including agenda and speaker slides are located at http://mlasymposium.libguides.com/c.php?g=584462&p=4036194.

Defining the Reproducibility Crisis

In March 2012, a commentary on the reproducibility of preclinical cancer studies was published in *Nature*. This report by the company Amgen highlighted the disappointing success rate of translating basic science findings into clinical therapeutics, especially with regards to cancer studies. Amgen researchers conducted a review of 53 published studies finding only six (11%) of the results could be replicated [1]. This endorsed an earlier report from the pharmaceutical

company Bayer that from a sample set of 67 published drug reports, only 20-25% were reproducible [2]. In response to these concerning observations, T*he Lancet* launched a series of papers that discuss increasing the value and reducing research waste in the published literature, ultimately becoming *The Lancet*'s Reduce Research Waste and Reward Diligence (REWARD) Campaign in 2015. In 2016, *Nature* published survey results from 1,576 researchers, 90% of whom affirmed a reproducibility crisis within the published literature [4].

In 2015, the U.S. National Science Foundation (NSF) released recommendations authored by the Subcommittee on Replicability in Science. Among other findings, this report noted that many terms used to discuss the ability to replicate or generalize a study were applied inconsistently. The report thereby offered the following definition:

> "Reproducibility refers to the ability of a researcher to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator (…) Reproducibility is a minimum necessary condition for a finding to be believable and informative." [3]

This differs slightly from replicability, defined by the report as "the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected" [3]. In contrast, the *Nature* survey of researchers asked if respondents were able to reproduce results in a "similar experimental system" which "may include slight variations in methods or materials" [4]. Moving forward, as funders, publishers, researchers, librarians, and other stakeholders work to formulate strategies to address the concern, it will be important to work from a standard definition of the problem. Many approaches would conceivably impact both reproducibility and replicability; however, when discussing the scale of the crisis, and proposing focused solutions, it will be important to note the difference.

Policies: What are funders and publishers doing to support reproducible research?

Proposed solutions for increasing the reproducibility (and replicability) of the published literature can typically be broken into two main approaches: those that target data reporting and those that target process reports, or methodologies. Selective data sharing was cited as one of the top factors contributing to the crises by respondents of the *Nature* survey. Selective reporting may occur when authors publish a clean story, leaving out factors such as replicates that did not meet expectations, outliers, or statistical tests that did not show desired

results. Data sharing initiatives help address these selective pressures by requiring all underlying data of published summaries and visualizations be made available. This creates options for timely, independent verification and may limit exaggerated reporting. However to address the reproducibility crises rather than just re-use initiatives, more rigorous sharing of research design and data collection methods are also required.

Funders of research have put together recommendations and guidelines in both these areas as seen by the NSF data sharing policy (https://www.nsf.gov/bfa/dias/policy/dmp.jsp) and the NIH Rigor and Reproducibility web portal (https://www.nih.gov/research-training/rigor-reproducibility*)*. Additionally, organizations such as the Center for Open Science (COS) (https://cos.io/) in the U.S. and the Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network (http://www.equator-network.org) in the U.K. also provide resources to address the process of sharing scientific findings. These organizations provide frameworks and guidelines for conducting and reporting reproducible scientific efforts. Guidelines and toolkits for reporting common study types are linked on the EQUATOR Network, while COS provides a framework for project management support, training, and updates on ongoing projects looking at field specific reproducibility issues.

Journal publishers have also responded to the call for increased reproducibility, though most have focused on data sharing strategies. *Science, Nature, Public Library of Science (PLoS), and Proceedings of the National Academy of Sciences (PNAS)* have all established data sharing or data accessibility policies for their authors. Authors in these journals are expected to share the data and statistics underlying their findings, preferably in a public domain repository. The sharing of such data allows for both reproducibility and confirmatory actives as well as the reuse of data for new discoveries.

Journal publishers focusing on the increased rigor of published methodologies have done so primarily by exploring new publication types. The *Journal of Visualized Experiments (JoVE)* has been a leader in this area, developing a publication module of instructional videos aimed to better communicate experimental processes. As a panelist at the 2017 MLA Reproducibility Symposium, *JoVE* co-founder and CEO, Moshe Pritsker, noted that the classic structure of the journal article has remained largely unchanged since the first publication in 1665 and called emphatically for detailed methods as a stand-alone publication model [5]. Reinforcing his call, another panelist shared an example of the importance of robust materials reporting when a study found that variations in software version, workstation type (Mac or PC), and Mac

operating system all had a significant impact on the analytic readings of 30 MRI scans [6]. When variables such as computer type and operating system are shown to play a role in reproducible data analysis, there is a clear need for more robust reporting than just product names.

Registered reports offer another emerging publication type in support of reproducibility. These are detailed study designs which undergo the peer-review process separate from the final study analysis. Registered reports describe key methodologies such as experimental model, data collection instruments, and statistical methods used in the proposed analysis. These are then reviewed for rigor and reproducibility before data are collected. Currently, COS lists 52 journals which have adopted the publication of registered reports in some capacity. Journals such as *Royal Society Open Science* and *BMC Biology* will then provisionally accept results for publication, contingent upon adherence to registered study design, regardless of study outcomes. This slow shift in accepted publication types represents a positive cultural shift in the scientific community by placing emphasis on rigorous scientific processes rather than focusing solely on novel, positive results.

Library Services: How can librarians support a culture of reproducible research?

As Kirtley emphasizes in her 2016 *The Lancet* commentary, librarians are well posed to be part of the answer to managing the current reproducibility crisis [7]. Academic and medical librarians are familiar partners in the research lifecycle, from developing robust and comprehensive search strategies, to selecting a journal and assisting with data sharing and management plans. Librarians who work on systematic review projects are intimately familiar with the challenges of adapting complex methods to new database environments. By attuning specific knowledge and services to the language and needs of addressing reproducibility, librarians are equipped to serve as advocates and partners. Even traditional roles such as collection development and access training can be easily adapted to addressing reproducibility as it relates to growing the awareness, availability, and utilization of new publication types.

Panelists at the 2017 MLA Reproducibility Symposium shared unique ways librarians at their institutions have contributed to creating reproducible research. Cynthia Hudson-Vitale, at the University of Washington in St. Louis, works with the Institute for Clinical and Translational Sciences to establish a framework for reproducible methods when working with electronic health records (EHRs). Librarians collaborated with researchers to identify 103 variables needed to ensure the reproducibility of EHR analysis: beginning with stating a clear, focused hypothesis

through tracking query language with version notes and access dates, identifying statistical tests and packages, and reporting such specifics with standard documentation. All 103 variables would be needed in any resulting publications or data codebooks or the analyses could not be reliably replicated, regardless if the raw de-identified data is available and accessible. Bart Ragon from the University of Virginia and Kristi Holmes from Northwestern University, echoed the vital importance for librarians to offer collaborative support for open science and data management.

To further explore actionable roles and services, participants at the Symposium completed a hands-on exercise in which each person was asked to propose a specific idea for librarian involvement and collaboration. Ideas were then anonymously scored for how well the proposal resonated for an individual's library and institution. Collected from over the 30 responses, the highest scoring suggestions are listed below. While specifics of each suggestion were not discussed at the Symposium, one possible interpretation of the proposal follows each participant idea.

- Host "Reproduce-My-Research" Events

  Such outreach or training events may take many different approaches. The idea seems to suggest giving researchers and students a formalized setting to reflect on and engage with each other specifically around how to improve the scientific reporting of their manuscripts. As this was the high scoring suggestion, it clearly resonated with many librarians who saw inspiration in this event title within their institutional outreach even without further details.

- Offer training to students and early career researchers

  Many libraries already provide training opportunities to gain a better understanding and engage with resources, publishers, and services. Refocusing or adopting specific language to target reproducibility concerns may be an easy adaptation for services already in place.

- Incorporate into the researcher workflow

  Again, many libraries already have collaborations or specific services targeting various aspects of the research lifecycle. Emphasizing these services as essential for addressing the reproducibility crisis may further campus collaborations and refine librarian roles.

- Establish data management best practices

When working with data management strategies, many initial approaches supported by libraries may have been reactive to specific funder mandates. By proactively establishing best practices in line with robust and reproducible science, librarians are natural champions of more accessible research and open data.

- Educate Librarians

  By seeking out learning and professional networking opportunities such as the 2017 Reproducibility Symposium, librarians are educating themselves about the problem, understanding researcher frustrations and emerging policies, and collaborating as a profession to explore targeted services. Opportunities to brainstorm and share challenges and success stories are invaluable for the development of robust and innovative services.

- Collaborate with institutional offices

  Collaboration is implied in many of the top-scoring suggestions. Continued outreach to researcher groups and campus partners will be essential to the recognition of librarians as part of the solution.

- Provide high-quality, reproducible search results

  For those librarians working on systematic reviews or providing in depth reference support, this is an opportunity to lead by example. Providing details of the search query along with specifics such as database coverage, search date and applied filters, raises subtle awareness of the necessity of these details. This may then provide the opportunity for discussion of similar essentials when reporting results.

The majority of these high scoring responses reflect activities and services already underway, with some variation, at many academic libraries. By educating librarians and raising awareness among researchers in the specific areas of reproducible research, many library services can be easily adapted to address the reproducibility crises. Hosting additional symposiums and workshops around the topic will encourage other librarians to share the specific services and outreach initiatives that have been successful.

References

1. Begley CG, Ellis LM. Drug development: Raise standards for preclinical cancer research. *Nature.* 2012;483(7391), 531-533.

2. Prinz F, Schlange T, Asadullah K. Believe it or not: how much can we rely on published data on potential drug targets? *Nat Rev Drug Discov*. 2011;10(9), 712-712.

3. Bollen K, Cacioppo JT, Kaplan RM, Krosnick JA, Olds JL, Dean H. Reproducibility, replicability, and generalization in the social, behavioral, and economic sciences. 2015. Retrieved from https://www.nsf.gov/sbe/SBE_Spring_2015_AC_Meeting_Presentations/Bollen_Report_on_Replicability_SubcommitteeMay_2015.pdf

4. Baker M. 1,500 scientists lift the lid on reproducibility. *Nature*. 2016;533(7604), 452-454. doi:10.1038/533452a

5. Pritsker M. The Science Reproducibility Crisis: how librarians can solve it and why they should care? Paper presented at the The Librarian's Role in Reproducibity of Research Symposium, Seattle, WA. 2017.

6. Gronenschild EHBM, Habets P, Jacobs HIL, Mengelers R, Rozendaal N, van Os J, Marcelis M. The Effects of FreeSurfer Version, Workstation Type, and Macintosh Operating System Version on Anatomical Volume and Cortical Thickness Measurements. *PLOS ONE*. 2012;7(6), e38234. doi:10.1371/journal.pone.0038234

7. Kirtley S. Increasing value and reducing waste in biomedical research: librarians are listening and are part of the answer. *The Lancet*. 2016;387(10028), 1601. doi:http://dx.doi.org/10.1016/S0140-6736(16)30241-0
