



EScience in Action

Diving into Data: Planning a Research Data Management Event

Robyn B. Reed

Penn State Hershey, Hershey, PA, USA

Abstract

The George T. Harrell Health Sciences Library at Penn State Hershey initiated its participation in institutional research data management activities by coordinating and hosting a well-attended data management symposium. To maximize relevance to clinical and basic sciences researchers, a planning committee of faculty and administrators assisted in defining important topics for the event. This article describes the symposium development and outcomes. The goal is to share this information with librarians who are seeking ways to become more involved with data management in their institutions.

Correspondence: Robyn B. Reed: rbr11@psu.edu

Keywords: data management, research data management, academic libraries, collaboration, marketing



All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Introduction

The roles librarians play in data management vary depending on institutional need and support. While some libraries have established collaborations in these areas and have integrated themselves into data management activities, other libraries are in the beginning stages of assisting researchers with their data management challenges. The areas where librarians play roles also vary widely and may include consulting and writing data management plans for grant applications, assisting with determining metadata standards, data curation and archiving, and finding and citing appropriate data repositories (Tenopir et al. 2012; Soehner 2010). Additionally, many academic research libraries are planning to offer data management services but have not initiated them at this time (Tenopir et al. 2014). Since these services can be institution-specific, they can be implemented in many ways (Raboin et al. 2013).

A challenge most libraries face is in addressing the needs of a diverse clientele. The George T. Harrell Health Sciences Library (Harrell Library) supports the information, research, and education needs of almost 10,000 faculty, staff, students, and postdoctoral scholars across both the Penn State College of Medicine and the Milton S. Hershey Medical Center (Penn State Hershey). In addition to the large user population, the Harrell Library supports a wide range of research activities in clinical, biomedical, and translational areas, as well as providing support for medical and graduate education programs.

With no formal mechanism to assist researchers with data management issues, most information was scattered throughout the institution. Many people relied on “word of mouth” or did not know where to turn when faced with questions related to data management. The action taken to initiate library involvement in data management activities was to host a half-day data management symposium, with the target audience being researchers – faculty, staff, and students at Penn State Hershey and University Park campuses. The goals of this event were to assist researchers in identifying resources and information on data management and to highlight the library as a conduit of information.

Methods

When identifying the various data management and data-related activities, it was important to first understand the environment. At the beginning of the planning process, the library approached the Penn State Clinical and Translational Science Institute (CTSI) about co-sponsoring an event to increase awareness of data management issues and identify investigator needs. Coordinators of the event met with the director of the CTSI, described the symposium plans, and agreed on co-sponsorship. The CTSI’s involvement provided an opportunity to promote its services and advance its educational mission.

A small task force of librarians, researchers, and administrators from the CTSI and academic departments was formed to identify data management needs and topics for the symposium. Members of the task force had extensive expertise in working with data in various disciplines through personal research as well as with the CTSI. Although an environmental scan was not conducted prior to the symposium, the CTSI had first-hand knowledge of data management issues and researcher needs that was crucial to identifying the event topics. The group consisted of the Penn State Hershey Chief Informatics Officer, the Director of Biomedical Informatics of the CTSI, and a CTSI administrator. Additionally, the task force included two

Penn State Hershey librarians and one Penn State librarian from University Park. The librarians' backgrounds were in biomedical informatics and reference/interlibrary loan (Hershey), and digital content/scholarly communications (University Park).

The symposium, entitled "Data Management in Biomedical Research: Information Challenges and Practical Strategies" was heavily marketed. Listservs proved to be the most effective method for distributing announcements to the targeted audiences of faculty, clinical and basic sciences researchers, librarians, students, and postdoctoral scholars. Posted flyers and announcements on digital signage throughout the institution were also part of the marketing campaign. Since the symposium was held on the Hershey campus, the event was streamed live to University Park to maximize viewing opportunities for people at remote locations. At registration, attendees were given the option of in-person or online attendance, and the online registrants were provided with a link to the presentation. The streaming mechanism used allowed off-site attendees to ask questions. The registration of 120 people and attendance of 150 researchers, administrators, and librarians (in person and online) suggested very high interest in data management topics.

Results

The symposium consisted of a keynote address and two panel discussions. The panels were assembled to review data management background and policies in the first session and describe practical resources to assist in the research process in the second. Researchers and professionals from Penn State Hershey and University Park campuses served as speakers for the panel discussions. Michael Conlon, Ph.D., from the University of Florida (UF) was the keynote speaker and began the event with a thought-provoking presentation on data management advances at UF. At the time of the symposium, Dr. Conlon was the co-Director and Chief Operating Officer of the UF CTSI. Attendees learned about UF clinical data systems, emphasizing handling of personalized medicine data, and biorepository data.

The first panel began with a faculty librarian in Publishing and Curation Services, University Park, providing an overview of the research data lifecycle. She described Penn State University's (PSU) institutional repository, ScholarSphere, and the various repository services available to researchers. Next, a faculty member from the Regulatory Support and Ethics Program of the CTSI from University Park was invited to discuss data ethics in clinical and translational research. Using several examples, he demonstrated how ethics is a component of all aspects of data management. Due to the importance of protected data in biomedical research, the third presenter was a compliance manager from the Human Subjects Protection Office, Penn State Hershey, who reviewed data regulations, PSU institutional policies, and resources for researchers. The final speaker for the first panel was a faculty member from the Department of Statistics, University Park, who serves as a CTSI Privacy Officer. She emphasized the importance of privacy and security of research data across PSU campuses. She described how easy it can be to identify an individual from de-identified human subject data and discussed ways of preventing this from happening.

The Associate Director from the Office of Technology Development, Penn State Hershey, led the second panel by describing intellectual property as it relates to data. Addressed were practical issues about data ownership, the sharing of data, and the policies that govern these activities. The second presenter was the Research Electronic Data Capture (REDCap)

Administrator from Research Informatics at Penn State Hershey who provided an overview of REDCap, the types of data in the system and appropriate uses of the software. Attendees were assured that it complies with both HIPAA and PSU's data security and integrity policies. The third presenter addressed the common challenge of statistical analysis of data. A faculty member from the Department of Public Health Sciences, Division of Biostatistics and Bioinformatics, Penn State Hershey, presented statistics services available through his department. He noted that working with statisticians is a research collaboration and emphasized the importance of their involvement in research projects in its early stages.

The second panel discussion continued with the CTSI Biomedical Informatics Program Leader from Penn State Hershey explaining the Clinical Informatics Research and Bioinformatics Cores of the CTSI. A description of useful data services available in bioinformatics and clinical research were presented to researchers. The final presentations were on information technology. A Professor in the College of Information Sciences and Technology, University Park, showed how he developed a new search engine to automate data extraction and indexing for the purposes of identifying experts and specific expertise. The rapid increase in available data sets and metadata provided even more opportunity for expansion. The Director of Research Computing, University Park, described IT services for scientists provided by Research Computing and Cyberinfrastructure. Additionally, the speaker gave an overview of current equipment, capabilities, and large memory servers that can assist with big data analysis.

A 20-minute question and answer period was allotted following each panel discussion. A few of the questions that arose were clarifications from the presentations; however, most of the questions were specific to individual research projects. Researchers asked about receiving specialized assistance in areas such as building a database, or inquired about limitations they experienced using a specific resource. These interactions helped to bridge connections needed to advance ongoing research projects.

Discussion

Success of the program was evident with high attendance throughout the event as well as positive feedback from individual participants. A possible reason for the success of the program was that it addressed an unmet need. The library and planning committee was effective in understanding researcher needs and identifying topics of interest and importance to them.

Following the event, one participant remarked, "That was really good. When you said you were organizing a research data management event, I thought it was going to be a library thing." Although some would consider the comment somewhat negative, another possibility is that the library was viewed as a legitimate contributor to the data management enterprise on campus. By not focusing on traditional library services and seeking the advice of people in relevant areas throughout the institution, the event highlighted the library as a valuable partner in the research process.

Additionally, establishing a partnership with the CTSI contributed to the success of the symposium. The CTSI is easily recognized across the institution and being interdisciplinary, covers many different departments and fields of study. Furthermore, technical and financial

support for some of the data management and informatics tools at Penn State comes from the CTSI.

Conclusion

The “Data Management in Biomedical Research” symposium was important in understanding the strengths, weaknesses, and challenges in data management across a wide spectrum of researchers in our institution. The symposium provided opportunities for the library to collaborate with a number of researchers, clinicians, administrators, the CTSI, and faculty both within our institution as well as at another Penn State campus. The event helped everyone gain a better understanding of the overall research landscape in such a large institution. The wide collaboration with and inclusion of units from outside the library contributed greatly to the success of the symposium. This is a model that could be implemented by other institutions wishing to hold a similar event. The next step the library is taking is to host a smaller event focusing on changes in funder regulations and PSU policies related to data.

Acknowledgements

Thank you to all who made this event possible: the panelists and keynote speaker, the faculty and staff of the Harrell Library and members of Penn State CTSI. A special thanks to the planning committee of Thomas Abendroth, MD, Arthur Berg, Ph.D., Esther Dell, MLS, Patricia Hswe, MLS, Ph.D., and Andrea Lazarus, MS, Ph.D. Thank you to the event’s sponsors: the Harrell Library, Penn State’s CTSI,* and the Penn State University Libraries Innovation Microgrant Program. Thank you to Nancy Butkovich, MLS for comments and editing of the article.

Funding Statement

*The project described was partially supported by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1 TR000127. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Disclosure

The author reports no conflict of interest.

References

Raboin, Regina, Rebecca C. Reznik-Zellen, and Dorothea Salo. 2012. “Forging New Service Paths: Institutional Approaches to Providing Research Data Management Services.” *Journal of eScience Librarianship* 1:134-147. Accessed January 27, 2015. <http://dx.doi.org/10.7191/jeslib.2012.1021>

Soehner, Catherine, Catherine Steeves, and Jennifer Ward. 2010. “E-Science and Data Support Services: A Study of ARL Member Institutions.” *Association of Research Libraries (ARL)*. Accessed January 27, 2015. <http://www.arl.org/storage/documents/publications/escience-report-2010.pdf>

Tenopir, Carol, Ben Birch, and Suzie Allard. 2012. “Academic Libraries and Research Data Services: Current Practices and Plans for the Future, An ACRL White Paper.” *Association of College and Research Libraries (ACRL)*. Accessed January 27, 2015. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf

Tenopir, Carol, Robert J. Sandusky, Suzie Allard, and Ben Birch. 2014. "Research Data Management Services in Academic Research Libraries and Perceptions of Librarians." *Library & Information Science Research* 36:84–90. Accessed January 27, 2015. <http://dx.doi.org/10.1016/j.lisr.2013.11.003>