

Archives and Data Management: The Purdue Story

Carly Dearborn, Digital Preservation and Electronic Records Archivist,
Purdue University Libraries

Purdue University archivists were involved in some of the earliest discussions on campus around data management, thanks to the vision of former dean of libraries, James Mullins. When Mullins first began as dean in the early 2000s, he toured campus to find out how the libraries could meet the needs of campus departments, faculty, and students. Data management was a need quickly identified by the chief information officer at the time; however, faculty and department heads struggled to see how the libraries could help with data. They viewed the library as a place where ideas and research were shared openly and freely, often unattractive concepts to research scientists prior to national data-sharing initiatives. Hearing this, Mullins realized that the archives was especially equipped to handle the concerns of these researchers. Archivists are specifically trained to handle sensitive information and work with donors from a variety of disciplinary and professional backgrounds who may need portions of their collections restricted or embargoed for personal, proprietary, or security reasons.¹ Furthermore, archivists have always been involved at every stage of the research process—from data collection to preservation and reuse—and are familiar with raw data (even if that data has historically been analog) and its challenges.²

After Dean Mullins identified data management as a space where the libraries could add value, Purdue established a series of committees to think through the issue of data management at an institutional level. These committees included librarians, computer engineers, IT professionals, and domain scientists. Conversations were led by the vice president for research, and chaired by the dean of libraries and vice president for information technology and research, a trio that would oversee the creation of the Purdue University Research Repository (PURR) and make up the PURR Steering Committee. In 2010, the steering committee created the PURR Working Group to

define and deploy the repository concept using the locally developed HUBzero software.³ The university archivist was a member of this working group along with other librarians, IT professionals, and domain scientists.

PURR was created to be a research and data management tool for Purdue researchers and their collaborators. It is currently the official data repository of the university. PURR now provides researchers

“The success of these and other collaborations at Purdue Libraries is directly related to an administratively supported environment of experimentation.”

with the tools to meet evolving data management and sharing requirements and a platform to seek help from their subject specialist librarian. PURR also provides a workspace for researchers to collaborate with their colleagues and an online publishing platform to

ultimately provide access to data. PURR publishes data sets with digital object identifiers (DOIs), which make it easier for other researchers to cite published data. Finally, PURR also provides preservation support for deposited data for up to 10 years—at which point the data set will go under review as would any other library collection.

After the PURR Working Group identified the requirements to create a repository and a service model, they put together a new team from the libraries to develop the platform’s functionality and preservation infrastructure. This team includes librarians, a repository specialist, a programmer, a data curator, and an archivist. All still actively work together on the repository to maintain content, develop new features and improve user experience.⁴

The early collaborative initiatives between the libraries and its campus partners, and within the libraries amongst archivists, librarians, and IT professionals still inform collaborative work today. Outside of involvement in PURR, for example, archivists frequently collaborate

with data librarians around data management instruction. Data management is still a growing need among graduate students and even undergraduate students. The libraries increasingly offer for-credit coursework in this area around campus. It can be challenging to talk about data management without speaking in hypotheticals. “If you don’t have a file-naming convention, X might happen.” “If you fail to export your data in this format, Y could occur.” It is difficult to teach the necessity of data management to those who have never experienced the devastation of poor data management. This is where Purdue archivists come in. Using teaching data sets drawn from actual donated data sets, archivists join data librarians in classes to demonstrate what poor data management looks like at the end of the research life cycle and how it affects reuse of the data.

Data management is often taught in the context of a research life cycle; however, preservation is still approached as an afterthought or a step to take at the end of a research project. Involving archivists and their expertise in data management instruction can demonstrate how preservation informs good data management practices at the outset. Collaborations at Purdue in this area have resulted in additional workshops and documentation to help students (and faculty) better prepare their data for long-term preservation or access. The success of these and other collaborations at Purdue Libraries is directly related to an administratively supported environment of experimentation.

“While data management is certainly a space where archivists belong, it does not mean the sense of belongingness comes easily.”

More and more institutions are enveloping archivists into their data management services. While data management is certainly a space where archivists belong, it does not mean the sense of belongingness comes easily. Still, it is important to remember that everyone is struggling with tough issues in the area of data management and

the perspective of an archivist can have a lot of impact in how a preservation strategy is developed or how repository services should support researchers. Purdue Libraries certainly is not done grappling with difficult data management issues, but the diverse makeup of the team makes some of the problems appear a little less daunting.

Endnotes

1. James L. Mullins, personal communication, July 9, 2015.
2. James L. Mullins, “Bringing Librarianship to E-Science,” *College & Research Libraries* 70, no. 3 (May 2009): 212–213, <https://doi.org/10.5860/0700212>.
3. Michael Witt, “Co-Designing, Co-Developing, and Co-Implementing an Institutional Data Repository Service,” *Journal of Library Administration* 52, no. 2 (2012): 172–188, <https://doi.org/10.1080/01930826.2012.655607>.
4. “The PURR Team,” Purdue University Research Repository, accessed December 14, 2018, <https://purrr.purdue.edu/team>.

© 2018 Carly Dearborn



This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>.

To cite this article: Carly Dearborn. “Archives and Data Management: The Purdue Story.” *Research Library Issues*, no. 296 (2018): 33–36. <https://doi.org/10.29242/rli.296.4>.