

Reimagining the Institutional Repository as an Open Data Archive

Lisa Johnston, Research Services Librarian and Co-Director of the University Digital Conservancy (ljohnsto@umn.edu) University of Minnesota – Twin Cities, Minneapolis, Minnesota, USA

Problem

Institutional repositories (IRs) have sprung up in academic institutions over the past decade to provide archival storage and dissemination services for locally-authored digital scholarship, primarily in the form of the traditional peer-reviewed article. However the implementation of IRs has not rapidly changed the landscape of scholarly communication as expected¹ and, without institutional mandates for deposit, many remain underused for their primary purpose. Simultaneously, a shift has occurred in academia that signaled an increased need for the stewardship of digital research data, for example, the requirements by national funding agencies that researchers share their data and plan for preservation and long-term access. This poster will describe the user-needs assessments, policy decisions, and technical infrastructure plans for reimagining the IR to meet data archiving needs across campus.

User Needs

In addition to national and disciplinary sea changes, local user-needs surveys at the University of Minnesota (UMN) in 2006², 2008³, and ongoing informal assessment of researchers who attend data management training sessions⁴ have demonstrated the need for services and support of research data management, sharing data and preservation for long-term access (see Fig 1); needs that the library can meet with the existing IR.

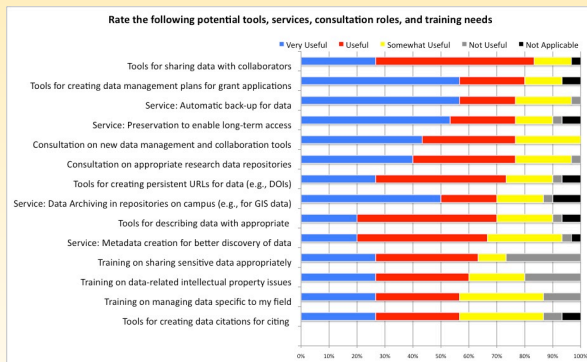


Figure 1. Responses by UMN researchers in a spring 2011 email survey on potential library services for managing, archiving, accessing and preserving research data.

Data Archiving Principles

The IR provides academic libraries a ready opportunity to assist researchers with research data sharing and preservation using our established repository services⁵, particularly where national and disciplinary data archives are not available. Recommendations include:

1. Open access is our thing

Accept open, publicly accessible data that does not contain any private, confidential, or personal identifiable information.

2. File formats for long-term access

IRs can accept files in all formats, but give simple recommendations on what formats can be persevered. Often, data might be archived in multiple formats for long-term access.

3. Upload file size limits

Large data files may not be possible in a traditional IR for many reasons (software limitations, cost to store). Set system upload limits and require that larger requests be mediated and/or fee-based.

4. Not storage, but long-term archive

IR archives are permanent and should include the final, or published, version of the data. Short-term, temporary storage needs should be handled in other ways.

6. Metadata and documentation required

Data must be supplemented by adequate documentation describing the nature of the data at an appropriate level for reuse and discovery. Set minimum level metadata requirements for data.

7. Permanent URL = easy citation

Data should have a unique, permanent URL separate from the main research article to encourage data citation.

8. Ingest: easy, but not too easy

Not all data is important enough to keep in an IR. Researchers must self-select what data is original, unique, and not easily reproduced or acquired elsewhere. The ingest process might be a necessary barrier.

9. Data download and reuse metrics

Just as with research articles, our IRs can demonstrate research-impact of open data by tracking download statistics and monitoring reuse.

10. Subject-based repositories

If an appropriate discipline-specific data repository for the data exists, that should be the first choice for long-term archiving of data sets.

Reimagining the IR for Data

At the University of Minnesota, our IR is undergoing a software replatforming from DSpace to Fedora. This transition gives us opportunity to reimagine how our IR might better address data archiving needs on campus.

The screenshot shows the digital conservancy website interface. It includes a search bar, navigation links, and a detailed metadata record for a file named 'Loki Flight 40'. The record includes fields for Title, Authors, Issue Date, Description, and Permanent URL. Below the metadata, there is a table of files in the item, listing file names, descriptions, sizes, and formats. The interface is clean and professional, with a focus on providing clear information about the digital objects.

Figure 2: Many researchers have small, manageable data that requires long-term data archiving and sharing. The IR provides a service when subject data repository don't exist.

Conclusions

Academic researchers are required to share and maintain access to their data, however, many disciplines have no viable archiving options. For many small, open data sets, the institutional repository can be an important partner to preserve and maintain research data for future use and discovery (see Fig 2). Beyond this first step we can expand our IR to anticipate new needs⁵ such as: access restrictions (enclaves/staging repositories), fee-based archiving for large data files, visualization capabilities, portability to future data archives, and custom metadata standards for interoperability and linked open data.

References

- [article] Hahn, T.B., Burrig, M., and Duggan, H. N. (2011). Has the Revolution in Scholarly Communication Lived Up to Its Promise? *ASIST Bulletin*, June/July 2011. http://www.asis.org/Bulletin/Jun-11/JunJul11_Hahn_Burrig_Nicksch.html
- [report] University of Minnesota Libraries (2007). Understanding research behaviors, information resources, and service needs of scientists at the University of Minnesota. <https://conservancy.umn.edu/handle/5546>
- [report] Anderson, T., Gjerdingen, C., Herrmann, B., Himes, K., & Johnston, L. (2009). Implementing cyberinfrastructure for 21st century research. <https://conservancy.umn.edu/handle/91772>
- [poster] Johnston, L. (2010). Creating a Data Management Plan and other Training Opportunities in the Library. International Digital Curation Conference, Chicago, Illinois, December 3, 2010.
- [web site] University of Minnesota Libraries (2011). Managing Your Data – Data Preservation and Archiving. <http://www.lib.umn.edu/datamanagement/archiving>
- [blog] Neylon, C. (2011). Building the perfect data repository...or the one that might get used. *Science in the Open*. Oct 27, 2011. <http://cameronneylon.net/blog/building-the-perfect-data-repository-or-the-one-that-might-get-used/>