The Evolving Role of the Library in Research

Paper Submitted to ARN Envisioning the Future as Critical Partners in Data-Driven Science Workshop
April 18-19, 2013
Washington, D.C.

Prepared by

Stephen Marine
Associate Dean of Special Collections
University of Cincinnati Libraries

With

Marcio Faerman, Ph.D.
Senior Scientist
Ohio Supercomputer Center

Mark Faulkner
Senior Associate VP for Innovations & Partnerships
University of Cincinnati Information Technologies (UCIT)

Linda Newman
Head, Digital Collections and Repositories
University of Cincinnati Libraries

UNIVERSITY OF Cincinnati

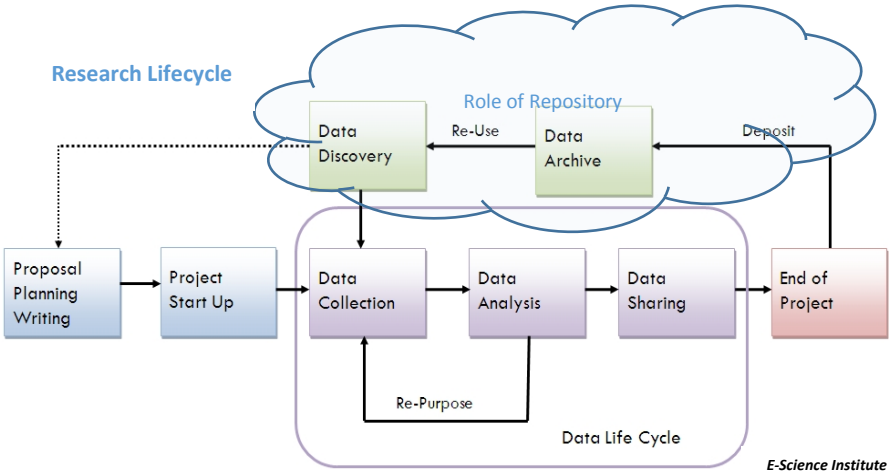
Acquiring, organizing, making accessible, and preserving information and data, including faculty archives, university records, theses/dissertations, and other output created by the university, is the *library's reason-for-being*. This experience positions them to now address "born-digital" gray literature (that is, much of the content in the sidebar below) and data sets produced by faculty and students. Without a repository and data management program such materials usually lie unknown and inaccessible on personal computers or departmental servers and without curation will eventually disappear.

Over the past decade and more, academic research libraries have made the leap to curate and provide access to digital resources in online repositories, both the born-digital and digitized special collections. Implementing research support programs also plays well to *skills and experience librarians* have developed, including: digital library collections and repositories, software development skills, project management and leadership, information management expertise, advanced knowledge of metadata, and ontology development expertise.

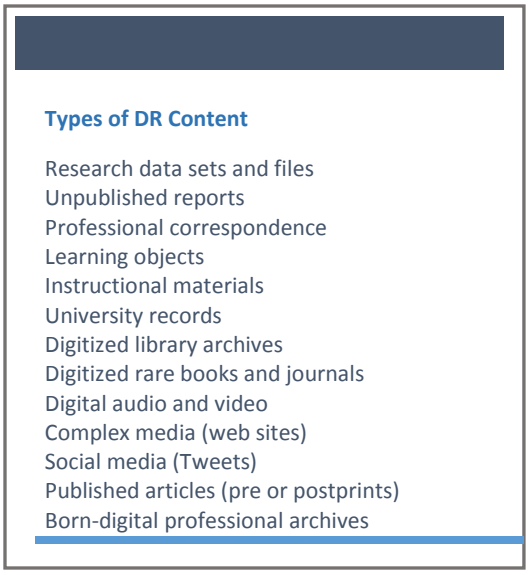
The world of 'big-data' and the computation science it requires is an excellent and natural fit with both the mission of libraries, and the skills and roles librarians have acquired with contemporary digital initiatives. Such roles include: developing and operating digital repositories; managing data and other research output throughout the research lifecycle with emphases on accessibility and preservation for productive reuse; consulting with faculty on data analysis and visualization; metadata/ontology development; open-access publishing; and other scholarly communication focused initiatives.

Digital Repositories

A digital repository (DR) *captures, organizes, makes accessible, enables reuse* and reformatting (ex. data visualization), and preserves the *intellectual output of an institution's research enterprise* and research-significant library collections of unique materials.



This output may be in *any digital format* from unpublished text documents, digital media, Web pages, social media, and data sets, to scholarly publications. A DR provides workflow functions that enable *content management and curation during the research lifecycle*, including embargoing, retention scheduling, and restricting access. A DR may also provide *workflow frameworks* for open-access journal publishing, archiving course materials from a learning management system, and managing retention of university records. Strategies for federating DRs within geographic and discipline-based consortia are critical.



Research Data Management/Curation

"The activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and reuse. For dynamic datasets this may mean continuous enrichment or updating to keep it fit for purpose. Higher levels of curation will also involve maintaining links with annotation and other published materials." (Lord, Macdonald, Lyon, and Giaretta)

The activities required for effectively managing data include: appraisal, accession, arrangement, description, storage, preservation, and access. By embedding librarians with research teams throughout the research lifecycle, they bring their skills and expertise to each phase of a research project. This may include:

- The preparation of data management plans (DMPs), now mandated by the National Science Foundation and the National Institutes of Health. Most government granting agencies are expected to require such plans, as well as many foundations and other grant-making institutions. Librarians at some institutions have developed DMP templates for various agencies.
- Consultation on data structures to ensure greatest compatibility with digital repository requirements.
- Implementing digital preservation strategies that identify and fix bit degradation and monitor format obsolescence, by engaging in such emerging programs as the Academic Preservation Trust and the Digital Preservation Network.
- Facilitation of data deposit into local and regional repositories, as well as national and international discipline-based repositories.
- Enabling access of data sets and other research documents by creating DOIs (Digital Object Identifiers) for data sets and by creating metadata recognized and indexed by search engines.
- Developing ontologies for specific sub-disciplines for which no metadata structures exist and expanding current metadata standards to adequately describe the provenance, context, and correct use of data in deposited data sets.
- Normalizing researchers' names so that an individual's complete output is accessible both in a repository and across other repositories, literature indexes, and library catalogs. This is a traditional "authority control" function performed by libraries. Research identity management is currently being addressed on several fronts, including ORCID.
- Providing a level of subject description that places data sets in a broader research context, as librarians now do for all other forms of digital and analog research resources.

Scholarly Communication/Digital Scholarship

University libraries are also pursuing collaborative ventures in supporting many aspects of scholarly communication and digital scholarship. These efforts are related – directly or indirectly – to issues of big data science and support provided by digital repositories and research data management.

- *Intellectual Property and Copyright Consultation.* Often in collaboration with offices of legal counsel or entrepreneurial affairs, libraries provide consultation to faculty and students on such questions as: Can I use this video in my class? In a distance learning course? A MOOC? Do I have to assign my copyright to a journal that has accepted my paper for publication? If I put my data sets in the repository, what rights to I retain? Can I restrict access?
- *Open-Access Publishing.* The inflation of commercially published journals has caused universities to consider other scholarly publishing models, such as establishing open-access publishing, in which the institution provides a publishing infrastructure and authors retain all rights to their publications. Libraries are often the locus for open-access publishing operations.
- *Database Design and Strategies.* Faculty often have needs for collecting and manipulating modest-size data sets. Scholarly communications offices are providing consultation and database design services to meet this need.
- *Data Analysis and Visualization Tools and Consultation.* Libraries are assisting faculty in analyzing, visualizing, and extracting meaning from data. PhDs in relevant fields, who often become data experts during graduate school, are seeking alternative careers as data experts in library scholarly communication offices. Custom facilities, often called research commons or digital scholarship centers, replete with high-end computers, projection capabilities, and advanced software for individual consultation and group sessions, are becoming prevalent in libraries.

Libraries and Advanced Regional and State Networks: Leverage

Efforts by libraries and regional/state networks to support university research can and should be complementary and supportive. Areas of overlap, commonality, mutual interest, and complementarity need wide discussion and hold tremendous promise. Just a few such areas might include:

- Federating multiple institutional digital repositories.
- Providing common data stores when such an approach is optimum.
- Providing remote collaboration tools among researchers and libraries in multi-institutional projects.

Other Resources

- Academic Preservation Trust (http://academicpreservationtrust.org/)
- University of California Curation Center (UC3) (http://www.cdlib.org/services/uc3/)
- Digital Preservation Network (http://www.dpn.org/)
- ORCID (http://orcid.org/)
- Purdue Distributed Data Curation Center (D2C2) (http://d2c2.lib.purdue.edu/)