May 2013

The Role of the Library in the Research Enterprise

Christopher J. Shaffer
Oregon Health & Science University, shafferc@ohsu.edu

Corresponding Author(s)

Follow this and additional works at: http://escholarship.umassmed.edu/jeslib

Part of the Library and Information Science Commons, and the Medicine and Health Sciences Commons

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

Recommended Citation

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
The Role of the Library in the Research Enterprise

Christopher J. Shaffer

Oregon Health & Science University, Portland, OR, USA

Abstract

Libraries have provided services to researchers for many years. Changes in technology and new publishing models provide opportunities for libraries to be more involved in the research enterprise. Within this article, the author reviews traditional library services, briefly describes the eScience and publishing landscape as it relates to libraries, and explores possible library programs in support of research. Many of the new opportunities require new partnerships, both within the institution and externally.

Introduction

Data collection, management, and analysis technologies are changing the landscape of research. Digital technologies, from sensors to analytical instrumentation, are increasingly a core component of observational and experimental research. Meanwhile, changes in scholarly publishing offer new opportunities for researchers to share the products of their work in ways that weren’t previously possible.

There has been an increasing interest in the library field to better connect with the research needs of faculty and students, and to explore how the skills, knowledge, and practices of librarianship could be applied towards supporting evolving eScience paradigms, particularly in the area of data curation.

Traditional Library Services

In the naïve view, all researchers want from the library are journals, journals, and more journals – free and online. They see little need to visit the library or communicate with librarians. Information professionals can certainly do many things to improve the usability of the journal literature. More and better tools are available to study researcher use of articles, from reading to citation patterns, which can be leveraged to better target purchasing and licensing decisions. Search engines, from Google Scholar to PubMed, are continually being improved to enhance retrieval of relevant information. Libraries have long provided training and research consultation services to improve the efficiency with which end-users search literature databases. Librarians help researchers manage citations and articles, providing training and support for products such as EndNote and RefWorks.

Proponents of open access are trying to introduce new financial models in support of transparent sharing of research results (Butter et al. 2012). The scholarly publishing crisis (aka the library funding crisis) is forcing research institutions to rethink services to
support researcher access to the literature, especially in times of shrinking budgets in higher education. Increasingly, when researchers can’t access the journal articles they need, they bypass traditional library services such as document delivery and interlibrary loan, which may be perceived as expensive and cumbersome, and instead email authors and colleagues. Commercial organizations, from publishers to aggregators, are marketing individual articles via pay-per-view, in partnership or competition with libraries. Breaking down the traditional unit of the journal volume or issue into commercially marketable units challenges the old models of collecting and acquiring journal literature for researchers.

Changes in Research and Researchers

As Jim Gray described it, eScience is a “transformed scientific method” or “the fourth paradigm.” Research was originally empirical. In the last few hundred years, theoretical models emerged. More recently, researchers have been able to use computational tools to explore simulations of complex environments. Now we have access to vast quantities of data from experiments and instruments, massive simulations, meta-analysis of research results, and more. Gray argues that this is a new way of doing research and requires a new model for conducting scientific inquiry (Gray 2009). However, not all research that falls under the rubric of eScience is conducted at the grand scale of particle physics or genomic experiments. There are many challenges facing researchers working at a variety of scales of data.

The explosion of publishing, driven by an increasingly competitive tenure and promotion environment and the growing specialization of science, has made a vast amount of journal literature available to researchers. Researchers are reading more and more articles every year, yet spend less and less time reading each individual article (Tenopir 2009). It is clear that technology is also presenting new data management challenges for researchers. Resource Navigators working with The eagle-i Consortium discovered that the vast majority of academic biomedical research laboratories do not have an effective inventory system for managing physical or digital resources (Shaffer 2012). The proliferation of computer files can transform the traditional lab notebook into a complex mess of spreadsheets and documents that can only be interpreted by the producer, if they can be found and interpreted at all.

Beyond the simple, yet massive, increase in the volume of research data collection, the complexity and diversity of data is increasing. Data manipulation technologies and algorithms can be so intricate that some researchers have posed fundamental questions about the reproducibility of computational research (Stodden 2010). Technology is also allowing the integration of quantitative and qualitative data in ways not previously possible, raising new data management issues (Estabrooks 2009). Funding agencies are beginning to mandate data sharing plans in grant applications to facilitate data reuse and eliminate redundancy. Technology is facilitating the sharing of research information prior to traditional publishing patterns, as seen in the emergence of “Science 2.0” or open science (Waldrop 2008).

Technology is also changing the culture of research. The emergence of team science challenges investigators to work together in new ways. In the example of health sciences, the dominance of the R01 grant is slowly giving way to the rise of Program Project and Center Grants. Wuchty, et al. (2007) showed that teams are growing larger, and their articles are more highly cited than solo authored articles. The National Institutes of Health’s emphasis on translational science is bringing basic science investigators together with clinicians to speed the transfer of knowledge from the bench to the bedside. Schools are revising tenure policies to recognize that not every researcher will have the opportunity to be first author, and articles
with 20 or more authors are not uncommon. The institutional organization needed to manage multidisciplinary and team research and is promoting the development of new skill sets and support structures (Boardman 2013). Superstar researchers are managing teams of hundreds, rather than individual labs staffed with a small group of students, research associates, and postdoctoral scholars.

The Research Enterprise

In order to identify new roles for libraries in the research enterprise, librarians must first gain a deep and multi-faceted understanding of the research environment at their own institutions. In the DuraSpace/ARL/DLF E-Science Institute, teams from dozens of research libraries examined their local environments through interviews with stakeholders, surveys, identification of primary areas of research emphasis, and analysis of institutional culture. The landscape analysis conducted by the teams took place with an understanding that an exploration of the research environment must include perspectives that are outside of the normal context of library research. Participation in planning by researchers, research administrators, and other service providers is essential. Outside voices provide important contextual information and opinions that help to inform the broader discussions taking place around eScience and data management.

The E-Science Institute teams, which included at least one person external to the library, created an inventory of the services and resources currently available to research teams. Some teams found that there was significant centralization of research administration, information technology, financial services, and other units providing services, while other teams found silos and fragmentation. Understanding the often complex array over overlapping services and providers helped libraries begin to identify gaps, which

Figure 1: Knowledge Transfer Data Life Cycle (Humphrey 2012).
the library might help fill. Libraries also dis-
covered unexpected pockets of research in
institutes, departments, and units that were
not traditionally considered research centers
on their campuses (The DuraSpace/ARL/
DLF 2012 E-science Institute 2012).

At the University of North Carolina at Chapel
Hill (UNC), the Provost’s Task Force on the
Stewardship of Digital Research Data, which
included representatives from the Library
and across the campus, conducted a re-
search data stewardship survey (Provost’s
Task Force on the Stewardship of Digital Re-
search Data 2012). In partnership with the
Oregon Health & Science University (OHSU)
Research Council, the OHSU Library imple-
mented the UNC survey (revised to better fit
the local setting in Oregon). Getting direct
feedback from researchers at the local insti-
tution is crucial to identifying their pain points
in management of data and other research
products.

The Role of the Library

There are many roles that libraries have as-
sumed in supporting eScience. In 2009, the
Association of American Universities, Asso-
ciation of Research Libraries, Coalition for
Networked Information, and the National As-
sociation of State Universities and Land
Grant Colleges issued a call to action urging
libraries to become involved in the dissemi-
nation of the full range of products of faculty
research and scholarship throughout the re-
search lifecycle (Hahn 2009). The E-
Science Institute was one of many respons-
es to that call. However, there is little con-
sensus on which, if any, objectives research
libraries should pursue in this arena.

New NSF and NIH regulations requiring re-
searchers to include data management
plans in grant applications appear to offer
libraries a new entrée into the research pro-
cess. The data lifecycle model of describing
the products of research provides a way for
librarians to examine issues related to the
curation of information from the inception of
an experiment or project through the publica-
tion of results (Humphrey 2012).

Research data services can be seen as a
natural extension of the research library’s
mission to collect, preserve, and make avail-
able to scholars a documented record of re-
search. Libraries have traditionally fulfilled
this charge at the end of the research pro-
cess: making articles available via journal
subscriptions, assisting with citation man-
agement, assessing research impact
through bibliometrics and citation analysis,
and assisting researchers with finding rele-
vant published literature. In recent years,
libraries have begun assisting with regulato-
cy compliance, most notably in assisting re-
searchers with required deposit of article
manuscripts in PubMed Central to comply
with the NIH Public Access Policy. In re-
response to changes in scholarly communica-
tion, librarians have promoted open access
and formed organizations like the Scholarly
Publishing and Academic Resources Coalition
(SPARC).

Librarian expertise with metadata design,
selection, and application could be applied to
the data curation and sharing process. Les-
sons learned in preservation and archiving
seem to be applicable to the challenges re-
searchers face in storing data and making it
available for analysis and reuse. The sug-
gestion that data citations could be used in
tenure and promotion has a clear analogue
to article citation and bibliometric analysis.

However, there are many potential barriers
to data sharing. Some potential partners,
such as technology transfer and business
development offices, might want to restrict
data sharing in ways that seem antithetical
to many librarians’ philosophy of free and
open sharing of information. Libraries must
find ways to work with these partners in
the service of researchers, rather than treating
them as competitors and wasting limited re-
sources on conflict. There are many rea-
sons that privacy of information may be
more important to the institution or the re-
searcher than data sharing. The severe penalties associated with release of individually identifiable health information under the Health Information Portability and Accountability Act, protection of the safety of researchers working in controversial fields like primate research, and the need to respect the cultural and privacy rights of study populations are just three examples. In some disciplines, researchers have a natural inclination to keep data secret to prevent being 'scooped,' or because they fear data misuse. In any case, curating data — publishing and archiving for preservation — is difficult and time consuming. This should not be discounted as perhaps the largest impediment to data sharing.

Some libraries are already promoting best practices in data management. Laboratory information management systems (LIMS), once limited to the largest and best-funded labs, are now available as web services and have been promoted as tools to better organize and describe research resources and data. Librarians are assisting lab managers with the development and implementation of metadata schema, and libraries have drafted templates for data management plans to use in grant applications. Information scientists and domain specialists are developing ontologies and implementing linked open data (LOD) to facilitate data harvesting and reuse.

But data isn't everything. At the E-Science Institute, teams were encouraged to consider potential services in:

- Scholarly communication (connecting data to articles, “data papers”);
- Virtual organizations (distributed research teams, shared compute);
- Physical space (visualization labs, group study);
- Policy development (IP or sharing, integrating, citing, preservation);
- Collection development (policies, areas, methods);
- Technical infrastructure (cyberinfrastructure, HPC, grid, storage).

There are also potential roles for libraries in promoting the scholarly outputs of their institutions. At OHSU, the Library participates in Research Week, an annual celebration of campus research that brings together people from across disciplines. Librarians and bio-informaticists at the Bernard Becker Medical Library of Washington University have developed a model for assessment of research impact (Sarli 2010). Expertise systems are being used by libraries on some campuses to highlight interests and accomplishments of researchers at an institution.

Expertise systems, such as VIVO, SciVal Experts, and Harvard Profiles, have been implemented at many research institutions. These tools can be used to promote an institution, to help build multidisciplinary or translational research teams, and to help research administrators make investment and recruiting decisions. The eagle-i Consortium is indexing research resources, such as core facilities, model organisms, antibodies, and plasmids, to facilitate sharing and re-use (Vasilevsky 2012). At many campuses, libraries are key players in building and publicizing expertise systems. However, other campuses have implemented systems without library involvement, so this is an area where the role of the library is not universally accepted.

The OHSU Library Ontology Development Group is developing an Integrated Semantic Framework (ISF), which will merge expertise and resource ontologies and allow for the integration of linked open data between disparate systems (Torniai 2011). This project is one of many multi-institution ontology collaborations with partners from industry and academia, bringing together domain specialists, computer scientists, librarians, and other information scientists. This work is a nat-
ural extension of libraries’ traditional role to index and catalog information for retrieval.

The E-Science Institute demonstrated that there is wide interest in providing eScience services by libraries. It also demonstrated that libraries are all over the map in developing and implementing services. There is as yet no consensus on what services should or will be considered essential for libraries to provide. This is not the first time that libraries have developed new services and programs, which are then seen as essential for research libraries to adopt in order to stay current. However, the institutional repository serves as a cautionary warning. Even where there is a defined need and service to respond to the need, it’s not clear that libraries will be the ones to take on new roles - competing service providers may be more successful (Carlson 2013). Perhaps the library can play a role as a connecting unit that helps unify the institution it serves, as the “Switzerland” of the research enterprise (Wirz 2012).

The five organizational stages for data curation developed by Kenney and McGovern (2003) could provide libraries that are considering development of eScience services a tool with which measure their current status and future potential. Institutions may acknowledge that there is an issue to address (stage 1), determine that e-research is of interest locally; act (stage 2), initiating relevant projects; consolidate (stage 3), shifting from projects to programs; institutionalize (stage 4), incorporate the broader environment and rationalizing programs; or externalize (stage 5) embracing inter-institutional collaboration and dependencies.

There are also questions about the workforce required for e-science initiatives. In addition to traditional library roles, such as cataloger, systems specialist, and reference librarian, libraries will also need new types of expertise, such as domain specialists, ontologists, and data curators. Will the existing workforce need retooling? Will non-librarians need to be hired? Or will new professional roles need to be developed? In the remainder of this issue, librarians and information scientists will try to answer these questions, as they describe a new kind of collaboration: information professionals embedded in research teams as informationists.

References


Stodden, Victoria C. “Reproducible research: Addressing the need for data and code sharing in computational science.” Computing in Science & Engineering 12, no. 5 (2010): 8-12, http://dx.doi.org/10.1109/MCSE.2010.113


Wuchty, Stefan, Benjamin F. Jones, and Bri-

**Disclosure:** The author reports no conflicts of interest.

All content in Journal of eScience Librarianship, unless otherwise noted, is licensed under a Creative Commons Attribution-Noncommercial-Share Alike License

[http://creativecommons.org/licenses/by-nc-sa/3.0/](http://creativecommons.org/licenses/by-nc-sa/3.0/)

ISSN 2161-3974