1-1-2013

Strategies for Data Management Engagement

Deborah H. Charbonneau

Wayne State University, dcharbon@wayne.edu

Recommended Citation

Charbonneau, D.H. (2013). Strategies for data management engagement. *Medical Reference Services Quarterly*, 32(3): 365-374.
Available at: <http://digitalcommons.wayne.edu/slisfrp/112>

This Article is brought to you for free and open access by the School of Library and Information Science at DigitalCommons@WayneState. It has been accepted for inclusion in School of Library and Information Science Faculty Research Publications by an authorized administrator of DigitalCommons@WayneState.

NOTICE IN COMPLIANCE WITH PUBLISHER POLICY: This is an Author's Accepted Manuscript, formatted for archiving (pagination added), of an article subsequently published in *Medical Reference Services Quarterly* 32(3):365-374, Copyright 2013 The Author, and available online at:
<http://www.tandfonline.com/10.1080/02763869.2013.807089>

Strategies for Data Management Engagement

INFORMATICS EDUCATION, Toni Hoberecht and Kimberly Pullen, Column Editors

Deborah H. Charbonneau

Abstract The research landscape is growing dramatically, and librarians are examining new roles, services, and types of collaborations to support data-intensive research. This column describes curricular enhancements at one School of Library and Information Science in the United States. Several key areas of data management in which health sciences librarians may wish to build or enhance their skills are outlined. Possible roles and opportunities for health sciences librarians to strategically engage in data management initiatives are also presented.

Keywords Curriculum, data management, health sciences librarianship, research, research data management

Author Deborah H. Charbonneau, Ph.D. (dcharbon@wayne.edu) is Assistant Professor, School of Library and Information Science, Wayne State University, Detroit, MI, USA 48202.

Comments and Suggestions should be sent to the Column Editors: Toni Hoberecht (tonihoberecht@ouhsc.edu) and Kimberly Pullen (kpulle@lsuhsc.edu).

Introduction

The research landscape is changing dramatically, and the volume of data being generated is growing immensely. As a result, health, science, and research in general have witnessed an explosion of data. This flood of data has been referred to as the “data deluge.”¹ Clearly, the nature of research is rapidly changing and so must the role of libraries in supporting data-intensive research.² Librarians are examining new roles, services, and types of collaborations needed to support data-centric research. Graduate

Strategies for Data Management Enhancement

schools of library and information science (LIS) are also enhancing their curricula to better prepare an information workforce skilled at addressing these e-science, e-research, and data management challenges.

The term *e-science* has been described as a new research methodology, fueled by networked capabilities and vast amounts of data requiring new strategies for research support,¹ whereas *e-research* is a broader term encompassing all domains of research which are also challenged by conducting “research in computationally enhanced and networked environments.”³ In this growing era of data-intensive research, opportunities exist for health sciences librarians to develop unique and valuable services.

Data management is a comprehensive approach to effectively sharing, managing, curating, preserving, and reusing the rapidly growing volumes of data generated by research.⁴ Data are “integral to the knowledge base that underpins scholarship, provide insight into our complex world, and inform decisions about our present and our future.”⁵ Research data are very diverse and may include observational, computational, and experimental data.⁶ Data types can range from “spatial, temporal, instrument-generated, models, and simulations” to images, video, and many others.⁷

Data management presents tremendous opportunities for librarians and graduate students of LIS alike to build their knowledge and skills to support data-intensive research. The purpose of this column is two-fold. First, it describes curricular enhancements at one school of library and information science in the United States. Throughout this discussion, several key areas of data management in which health sciences librarians may wish to build or enhance their skills are outlined. Examples of tools useful for supporting data management are also described. Second, this column presents possible roles and opportunities for health sciences librarians to strategically engage in data management initiatives at their institutions or organizations.

Curriculum Enhancements

According to Auckland, LIS schools have an opportunity to develop courses “to ensure new professionals are equipped to support the current and emerging research environment.”² The Wayne State University School

of Library and Information Science in Detroit, Michigan is one such school that embarked on an initiative to enhance the Master of Library and Information Science curriculum. Ogburn states that it is imperative for professional graduate programs to incorporate data management into their curricula in order “to grow effective future librarians.”⁵

Piorun et al. suggest topics suitable for an overview of research data management, including managing and sharing research data, understanding relevant public policies, and gaining knowledge of the data life cycle.⁸ Selected data management skills and knowledge relevant to supporting data-intensive research were integrated into the LIS curriculum at Wayne State University. While not a comprehensive set of data management skills and topics, librarians may find this sample of research data management course topics to be beneficial for beginning to build knowledge, expand expertise, or keep up-to-date:

- Research and data life cycles
- Principles and practices in data management and use
- Tools for managing, using, and sharing data
- Metadata standards for describing data (i.e., datasets)
- Workflows in scientific research processes
- Roles of librarians in providing research data management services

Opportunities for Health Sciences Librarians

The following section explores the potential roles and opportunities for health sciences librarians in data-intensive research environments. As librarians grow their skills and expertise in data management, they can become uniquely situated to support data management issues throughout the research life cycle. Within this context, several promising areas exist where librarians can play active roles in providing researcher data awareness, training, and services at their institutions. These areas of potential data management engagement, as will be discussed, are:

1. Data Interviews
2. Data Management Planning
3. Funder Compliance

Strategies for Data Management Enhancement

4. Data Literacy Instruction
5. Data Ethics
6. Data Citation and Attribution
7. Data Sharing
8. Data Publishing
9. Emerging Data Metrics
10. Data Curation and Preservation

Data Interviews

Information about the needs, practices, and attitudes of scholars and their data will be essential to the creation and delivery of relevant library support.⁹ In order to better understand the data needs and outputs of researchers, librarians can conduct data interviews. Data interviews can illuminate the types of data researchers generate and are willing to share, potential concerns, and desired services.¹⁰ Although librarians have been trained in collection management, much of this training was related to traditional formats such as monographs and journals rather than datasets and digital research data.¹¹ Conducting data interviews can assist librarians in better understanding researcher data needs. Questions during the data interview should seek to provide a context for how research data will be produced during a research project, what format the data are in, the expected lifespan of the data, and how the data could be used or re-used. Additional data interview questions might seek to clarify ownership of the data, potential audiences for the data, and what publications have resulted from the data.¹¹

Following the data interviews, librarians can assess potential areas in which they are able to provide researcher data advice. The ability to consult on research outputs and the discovery, access, dissemination, and preservation of research data are key skills needed for librarians to support a rapidly changing research environment.² Essentially, conducting a data interview is an important approach to collaboratively working and engaging with researchers for data management. Conducting data interviews can be instrumental in eliciting helpful information about researcher data needs while also affording librarians an opportunity to demonstrate their ability to advise on various data

management planning issues. In turn, this collaborative data interview process may help to increase researcher awareness of data management issues and the range of services or expertise librarians can offer.

Data Management Planning

Data management planning entails the development of policies and procedures to manage data throughout its life cycle.⁷ An understanding of various research life cycles, workflows, and communities of practice is central so that librarians can advise on data management planning issues. One tool in particular assists with developing an understanding of various scientific workflows. The resource called myExperiment <<http://www.myexperiment.org>> is a collaborative research environment where scientists can share their scientific workflows, share expertise, and exchange digital items related to their research. In fact, “myExperiment is now the largest public repository of scientific workflows.”¹²

Several tools support data management planning efforts and may be useful for librarians and researchers. A data management plan (DMP) describes the data that will be produced during research and “how the data will be managed and made accessible throughout its lifetime.”⁷ Existing tools include the DMPTool <<https://dmp.cdlib.org/>>, DMP Online <<http://www.dcc.ac.uk/dmponline>> developed by the Digital Curation Centre, and the Data Curation Profiles Toolkit <<http://datacurationprofiles.org/>>. While these tools have some similarities and differences, each one provides guidance related to the management of research data throughout its life cycle. These tools furnish guidance on a range of topics including crafting a data management plan, documentation and metadata, file formats, data security and backup, ethical and legal issues, and sharing and archiving data. With this in mind, health sciences librarians can provide critical training and access to such tools for assisting with data management planning efforts.

Data management training efforts may range from the development of guides, web pages, online tutorials, guidelines, and checklists to presentations and data consultations. Other methods for engaging with researchers about data management issues can be more formal and at a strategic level. These may include participation in faculty meetings and related research committee work to advocate for institutional policies.² A useful strategy is to learn where data-intensive research activities are

Strategies for Data Management Enhancement

occurring within an organization, as they may be taking place in clinical research, basic sciences, or clinical and translational science programs.¹³ This is one important way health sciences librarians can “continue to span boundaries” and understand emerging information and data management needs.¹³ A survey conducted by the Association of Research Libraries of member institutions emphasized the importance of collaborations for supporting data-intensive research.¹⁴ As such, other collaborative approaches may include working with key stakeholder groups such as academic senates, computing or information technology divisions, or other research divisions to advance data management awareness.

Funder Compliance

While data management planning is good practice for all researchers, funders may require data sharing or data management plans. Funding agency data policies in the United States range in their scope and require grant recipients to make research outputs accessible, including publications and research data. Numerous funding agencies now require data-related management, sharing, and preservation plans as part of grant applications. Another way that librarians are engaging with researchers to increase awareness of data issues is by offering valuable guidance for complying with funder policy requirements. In fact, several studies show that researchers rarely have the skills required to appropriately manage their data.¹⁵ Researchers need advice concerning data management plans, technical standards, metadata standards, and preservation management.¹⁶ Here, librarians can provide critical data management training to improve researchers’ skills. Libraries can offer, and may already offer in some cases, assistance to their research communities by creating, writing, and implementing data management plans in response to funding agency requirements.¹⁷

Understanding policies and data requirements researchers encounter with respect to funder compliance represents a key area for engagement. Dietrich et al. provide a nice overview of data policies from major funding agencies in the United States, including the Public Access Policy from the National Institutes of Health (NIH) and the National Science Foundation’s (NSF) data management plan requirement.¹⁷ In addition, investigators submitting a grant application to NIH “are expected to include a plan for sharing data or explain why sharing data is not possible.”¹⁸ As funding

agency requirements continue to expand, it will become necessary for researchers to address their plans for effectively managing their research data, including the long-term storage, preservation, and access to their research data.¹⁷ To this end, librarians can play an important role in the education, awareness, and support of funder compliance.

Data Instruction, Ethics, and Citation

Data provide opportunities for growth in the area of *data instruction*.¹⁸ One recommendation is to use data more frequently in instructional efforts. As Gore states, data literacy instruction includes understanding how to retrieve, select, assess, manipulate, and cite data.⁶ Librarians can leverage instructional opportunities and incorporate data into training, provide guidance on where to locate data, and demonstrate how to properly cite data. *Data citation* refers to providing a reference to data or datasets in the same way bibliographic references are routinely attributed to resources such as journal articles or books.¹⁹ Data are beginning to be recognized as a primary research output. In the past, data were seldom cited in the same way as a journal article or another publication. Yet, as researchers look for alternative ways to assess their research impact, ideas about data as an important research output are starting to change.

Although data citation standards vary across disciplines, DataCite <<http://datacite.org>> is one initiative attempting to “make it easier for research datasets to be handled as independent, citable, unique scientific objects.” This particular tool supports researchers and others by helping people to find, identify, and cite research data. This is largely done by using digital object identifiers (DOI) as permanent identifiers for datasets.

In one study, faculty members reported that *data ethics* was another area where students need assistance.¹⁵ Data ethics “includes intellectual property rights, ownership of data, issues of confidentiality/privacy, human subjects, implications for sharing (or not sharing) data with others, and assigning attribution in order to gain recognition of one’s work.”¹⁵ Hence, providing data-related instruction so that research data can be properly cited would be helpful. The area of data ethics offers a promising topic for engagement with researchers, including students.

Data Sharing and Publishing

Data sharing and ensuring the accessibility of data are crucial for future scientific purposes. Since research data are a valuable resource, the benefits of sharing data include facilitating new scientific inquiry, promoting potential new uses of data, and encouraging the validation of research. Educating faculty and researchers about the importance and benefits of data sharing provides one avenue for librarian engagement.

Tools, services, and infrastructure to support the sharing and dissemination of research data and outputs vary. DataBib <<http://databib.org/>> is an annotated and searchable list of data repositories. It is an excellent resource for both data users (i.e., those looking for data to use) and data authors (i.e., researchers submitting their data). This resource can be used as a starting point in consultations with researchers to aid in the identification and selection of appropriate data repositories. In addition, an institutional repository may also be a viable option for data sharing and data preservation purposes. The worldwide online directory of academic open access repositories, known as OpenDOAR <<http://opendoar.org/index.html>>, is another useful resource that can be searched for institutions with open access repositories.

Data repositories and institutional repositories are one way to comply with journal data sharing policies. The Joint Data Archiving Policy (JDAP) mandates supporting data to be publicly archived and made available as a condition of publication.²⁰ JDAP has been adopted by journals across various disciplines, including major scientific journals such as the *British Medical Journal* (BMJ), *Nature*, *Nucleic Acids Research*, and *Public Library of Science* (PLOS). Datasets and other supplemental materials are being embedded in electronic journal articles and “adding immense value.”⁶ Other journal policies require data to be accessible to colleagues or to be placed in public repositories such as Dryad <<http://datadryad.org/>> to facilitate the location of data. Once again, librarians can offer valuable guidance on how to comply with journal data policies by assisting researchers with the identification of possible data repositories appropriate for their research data sharing, archiving, and publishing needs.

As more researchers are willing to contribute research data to repositories, data tracking metrics are another way to help assess the impact of research outputs. Many repository platforms, including Digital

Commons <<http://digitalcommons.bepress.com/>>, DSpace <<http://www.dspace.org/>>, and Fedora <<http://fedora-commons.org/>>, already provide usage statistics for content contributors. In addition, tools such as figshare <<http://figshare.com/>> and the abovementioned Dryad data repository are just a few more tools offering researchers such data tracking metrics.

Data Curation and Preservation

Librarians must consider their role in data management, data curation, and data stewardship in the scientific areas.²¹ To clarify, according to an ARL task force, *data preservation* “consists of the management practices based on standards that guide and build metadata (and data) throughout the research life cycle” and “the subsequent long-term care for these products.”²² *Data curation* is used to refer to “ways of organizing, displaying, and re-purposing preserved data collections.”²² In fact, existing subject librarian liaison functions can be extended to incorporate data curation and preservation efforts.²³

Gold further notes that some reference and subject librarians have already incorporated data services into their work, particularly in the social sciences and geospatial data arenas; due to this, she argues that what is needed at this point is an expanded scope of librarian involvement with research data.²⁴ Health sciences librarians can support the life cycle of research data by providing and facilitating services for data storage and permanent access. Depending on the available infrastructure, this may entail developing or utilizing an institutional data repository. Librarians can also play a role in exploring or negotiating storage solutions for research data in cooperation with computing and information technology units. Alternatively, librarians might assist researchers in locating an existing data repository for their data such as a discipline-specific or cloud-based data repository. Nevertheless, librarians can provide important consultation services for research data curation and preservation thereby engaging researchers in the long-term preservation of their research data. Such engagement helps to ensure access and usability of research data over the long term.

Moving Forward

Librarians have often been “pioneers in the development of systems, processes, and innovative approaches” to the delivery of information.³ As libraries move forward in the data era, health sciences librarians can be proactive and responsive to research data management issues. This column focused primarily on librarians serving as advisors, trainers, and supporters of researchers and their data management needs. While beyond the scope of this column, librarians are also active in other data roles such as data creators, collectors, custodians, and consumers.

Promising areas for engagement discussed here are relevant to the dynamic research landscape. It should be noted, however, that these topics and skills are not meant to be an exhaustive list. In fact, data management issues are often interconnected and overlapping areas of potential engagement. For these reasons, the highlighted data management topics and sample of tools summarized in this column offer a foundation from which librarians can continue to build. With the increasing data-intensive nature of research, the ability to advise on potential data manipulation tools and metadata are additional areas that librarians might want to grow to better meet the challenges of data-driven research.^{2, 25} Advanced skills or specialized coursework in the areas of data analytics, visualization, relational databases, and data mining may also be beneficial.

Conclusion

Data management presents tremendous opportunities for health sciences librarians to contribute to and support endeavors in health, science, and research in general. Given the proliferation of research and scientific data, an information workforce prepared to address data management issues in partnership with the research community is warranted. This column highlighted curricular developments at one graduate school of library and information science in the United States as an effort to prepare an information workforce responsive to data-intensive research environments. With this in mind, librarians can enhance or expand their knowledge of data principles, practices, and tools. Numerous opportunities for librarian engagement in the area of data management were also presented. As library and information professionals continue to

enhance their repertoire of data management skills, they will be well-positioned to offer distinctive services and expertise for increasingly complex and exciting data challenges.

References

1. Hey, T., and J. Hey. "E-Science and Its Implications for the Library Community." *Library Hi Tech* 24, no. 4 (2006): 515-528.
2. Auckland, M. "Re-Skilling for Research: An Investigation into the Role and Skills of Subject and Liaison Librarians Required to Effectively Support the Evolving Information Needs of Researchers." *Research Libraries UK*. January 2012. [http://www.rluk.ac.uk/files/RLUK Re-skilling.pdf](http://www.rluk.ac.uk/files/RLUK_Re-skilling.pdf).
3. Brandt, D.S. "Librarians as Partners in E-Research: Purdue University Libraries Promote Collaboration." *College & Research Libraries News* 68, no. 6 (2007): 365-396.
4. Lyon, L. "The Informatics Transform: Re-engineering Libraries for the Data Decade." *International Journal of Digital Curation* 7, no. 1 (2012): 126-138.
5. Ogburn, J.L. "The Imperative for Data Curation." *Libraries and the Academy* 10, no. 2 (April 2010): 241-246.
6. Gore, S.A. "E-Science and Data Management Resources on the Web." *Medical Reference Services Quarterly* 30, no. 2 (April-June 2011): 167-177.
7. DataONE. "Best Practices." Accessed April 21, 2013. <http://www.dataone.org/all-best-practices>.
8. Piorun, M.E., D. Kafel, T. Leger-Hornby et al. "Teaching Research Data Management: An Undergraduate/Graduate Curriculum." *Journal of eScience Librarianship* 1, no. 1 (February 14, 2012). <http://dx.doi.org/10.7191/jeslib.2012.1003>.
9. Scaramozzino, J.M., M.L. Ramírez, and K.J. McGaughey. "A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University." *College & Research Libraries* 73, no. 4 (2012): 349-365.
10. Carlson, J.R. "Demystifying the Data Interview: Developing a Foundation for Reference Librarians to Talk with Researchers about their Data." *Purdue University e-Pubs. Libraries Research Publications* no. 153 (2011).

Strategies for Data Management Enhancement

- http://docs.lib.purdue.edu/lib_research/153.
11. Witt, M., and J. R. Carlson. "Conducting a Data Interview." *Purdue University e-Pubs. Libraries Research Publications* Paper no. 81 (2007).
http://docs.lib.purdue.edu/lib_research/81.
 12. myExperiment Wiki. "Main Page." Last modified January 30, 2013.
<http://wiki.www.myexperiment.org/index/php/Main-Page>.
 13. Rambo, N. "E-Science and Biomedical Libraries." *Journal of the Medical Library Association* 97, no. 3 (July 2009): 159-161.
 14. Association of Research Libraries. "E-Science and Data Support Services. A Study of ARL Member Institutions." Accessed April 1, 2013.
<http://www.arl.org/storage/documents/publications/escience-report-2010.pdf>.
 15. Carlson, J., M. Fosmire, C. Miller, and M. Nelson. "Determining Data Information Literacy Needs: A Study of Students and Research Faculty." *Libraries and the Academy* 11, no. 2 (April 2011): 629-657.
 16. Shearer, K., and D. Argáez. "Addressing the Research Data Gap: A Review of Novel Services for Libraries." *Canadian Association of Research Libraries* (2010). Accessed March 28, 2013.
http://www.carl-abrc.ca/uploads/pdfs/library_roles-final.pdf.
 17. Dietrich, D., T. Adamus, A. Miner, and G. Steinhaart. "De-Mystifying the Data Management Requirements of Research Funders." *Issues in Science and Technology Librarianship* (Summer 2012).
<http://www.istl.org/12-summer/refereed1.html>. DOI: 10.5062/F44M92G2.
 18. National Institutes of Health. "National Institutes of Health Data Sharing Guidance." Updated March 5, 2003.
http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm.
 19. Australian National Data Service. "Data Citation." Accessed March 28, 2013. <http://www.ands.org.au/cite-data>.
 20. Dryad. "Joint Data Archiving Policy (JDAP)." Accessed March 28, 2013.
<http://datadryad.org/pages/jdap>.
 21. Lynch, C., and D. Carleton. "The Impact of Digital Scholarship on Research Libraries." *The Journal of Library Administration* 49, no. 3 (April 2009): 227-244.
 22. Joint Task Force on Library Support for E-Science. "Agenda for Developing E-Science in Research Libraries." Association of

Research Libraries. November 2007.
<http://www.arl.org/storage/documents/publications/escience-report-final-2007.pdf>.

23. Gabridge, T. "The Last Mile: Liaison Roles in Curating Science and Engineering Research Data." *Research Library Issues: A Bimonthly Report from ARL, CNI, and SPARC* 265. August 2009.
<http://publications.arl.org/rli265/16>.
24. Gold, A. "Cyberinfrastructure, Data and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries." *D-Lib Magazine* 13, no. 9/10. September/October 2007.
<http://www.dlib.org/dlib/september07/gold/09gold-pt2.html>.
25. Qin, J., and J. D'Ignazio. "Lessons Learned from a Two-Year Experience in Science Data Literacy Education." In *Proceedings of the 31st Annual International Association of Scientific and Technology University Libraries Conference, West Lafayette, Indiana, June 20 - 24, 2010*.
<http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1009&context=iatul2010>.