Embedded Data Librarianship: A Case Study of Providing Data Management Support for a Science Department

Rutgers University has made this article freely available. Please share how this access benefits you. Your story matters. [https://rucore.libraries.rutgers.edu/rutgers-lib/47849/story/]

This work is an ACCEPTED MANUSCRIPT (AM)
This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.


Terms of Use: Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

Article begins on next page
TITLE
Embedded Data Librarianship: A Case Study of Providing Data Management Support for a Science Department

ABSTRACT
This case study details how a data services librarian and a science librarian collaborate to provide embedded data management support for the research-oriented Department of Earth and Environmental Sciences at Rutgers University–Newark. Combining their familiarity with emerging professional practices and resources, their efforts to gain a deeper understanding of the specific data management needs of researchers in the department, and their research into the evolving research data infrastructure in that particular discipline, the two are able to successfully connect researchers with the best practices in data management, suitable data repositories, and experts in the campus’ Computing Services unit.

KEYWORDS
Data management support, data services, embedded librarianship, science librarians, academic libraries
INTRODUCTION

New roles for academic librarianship have been an intriguing topic recently – especially given the change in eScience research paradigm and as those we serve in academia march deeply into the digital age. If the coming of the digital age draws panic from librarians that we will easily be forgotten by researchers who will see us simply as gatekeepers of print collections, then the eScience research paradigm shift presents us with a promising opportunity to be perceived as even more useful to researchers as information experts who excel at managing and preserving digital assets, including research data. Academic librarians might envision a new function serving as a bridge between the ever-changing information and communication technologies and the researchers who are mostly traditionally trained, but need to conduct research in a data-intensive and increasingly collaborative scholarly environment (Jones 2009). This new library service concerning research data provides an additional terrain for embedded librarianship (Carlson and Kneale 2011), a model that has been recognized in the library profession as the focus of the future (Kesselman and Watstein 2009).

This case study describes how a data services librarian is collaborating with a science librarian to provide data management support to a research-intensive science department. The needs of the science department were analyzed and useful resources from the information and technology landscapes that best meet the department’s needs were identified. This article discusses how the librarians connect these worlds in order to successfully deliver data management services through multiple embedded ways: data management training for graduate students, data
management plan (DMP) consultation, and building connections between the department and Computing Services unit on campus.

LITERATURE REVIEW

Traditionally, in the print information world, librarians are book and journal curators, controlling users’ access to information and at the same time, serving as the gateway and guide to a well-organized information world (Hawkins 2012). Now, in the digital information age, library collections, especially academic journals, are almost all in digital format and users can access them easily by themselves without needing to set foot in the library. At the same time, research data is becoming an important component of research materials, but has not been curated or managed very well (Jahnke, Asher, and Keralis 2012; Marcum and George 2010). This is where academic libraries see an opportunity for revival – by transforming their information management and preservation expertise into new fields of applications (Tenopir, Birch, and Allard 2012).

There are many uncertainties as to what academic librarians should do and how they might prepare to be competent in their envisioned new roles. Alvaro, Brooks, Ham, and Poegel’s (2011) recent analysis of job advertisements for eScience librarians showed that although there seems to be an emerging need for eScience librarians, it is unclear how exactly these librarians will be involved with eScience research – as data librarians focusing merely on data support or as subject librarians with data support responsibilities. Neither is it clear what specific skill gaps need to be filled since no standard or consistent job responsibilities were identified. A survey
conducted by the Association of Research Libraries (ARL) showed that many science librarians are unclear about their roles in providing research support to other campus units, and that there is no agreement among science librarians about what skills will be required of them for this type of support – whether it is new, research data-related skills or traditional, subject specialist/liaison librarian skills (Antell et al. 2014). A qualitative interview study of UK library staffs revealed great variation in libraries’ research data management initiatives at the institutional level due to the complicated local environments where governance, resources and skills are all important influencing factors (Pinfield, Cox, and Smith 2014).

Even given the above-mentioned uncertainties, there are some theoretical and empirical cases in the academic librarianship literature that optimistically discuss the future role of academic librarians, especially in terms of research support. One vision for new roles in academic librarianship expands on the traditional information-literacy educator’s role to one that teaches data information literacy to researchers. Data literacy content and competencies have been continuously explored, with attempts to integrate it into the existing information literacy standards in the higher education environment (Carlson et al. 2011; Prado and Marzal 2013). Additionally, a series of data information literacy education practices were carried out in a variety of disciplines, departments, and institutions in a diverse number of formats, such as online courses, workshops, embedded librarianship, mini-courses, and assigned readings followed by discussion-based instruction (Carlson et al. 2013).

Another vision for new roles is based on initial practices that see academic librarians as collaborators and partners who directly contribute to and even embedded in research projects
Although academic librarians may be leveraging their traditional technical skills in metadata, information/data organization, and curation for these projects, they are viewed by the researchers as trusted partners and are approached for collaboration beginning with the early stages of research projects, sometimes even invited to apply for grants together with the researchers (Carpenter et al. 2011). When examining the librarians’ potential roles at each stage of the research life cycle – from generating ideas, writing proposals, finding interdisciplinary partners, and performing research, to publishing and preserving research results – a group of health science librarians easily identified unique contributions that librarians are able to bring to the table, such as helping researchers identify existing data resources, demonstrating impact of their research, performing effective data management, and using metadata to facilitate research results discovery (Hamasu, Jones, and Kelly 2012).

Successful research data services can broaden the range of interactions between librarians and researchers and thus create additional niches for practicing embedded librarianship. Based on a comprehensive survey and interview study of members of Special Library Association, Shumaker and Talley (2009) found that typical models of embedded librarianship include librarians participating in an academic course teaching information literacy skills, librarians holding office hours in service departments or units, and the integration of informationists into clinical care teams. New areas of embedded librarianship were recognized by other studies. For example, Kesselman and Watstein (2009) noted embedded librarianship cases where librarians collaborate with faculty in scholarly communication activities and play a role in multidisciplinary research. Carlson and Kneale (2011) observed that embedded librarianship

(Brandt 2007; Carlson and Kneale 2011; Fortin and Mueller 2013; Garritano and Carlson 2009).
could also be applied in the research context in two ways: one is project-based, where faculty-librarian collaboration depends on the needs of a project; the other is program-based, where a librarian is hired by an organization and supports multiple projects on an ongoing basis.

Although there has been an increase in the recent literature about research data services, the idea of providing data support as a form of embedded librarianship has not been further and empirically illustrated to the point of helping academic librarians gain a deeper understanding of the rich possibilities this concept implies. The reason could be that academic institutions are just beginning to transform themselves and are still in the early stages of developing research support services, so many libraries and librarians have not yet implemented any such programs or services (Soehner, Steeves, and Ward 2010; Tenopir et al. 2013). Therefore, at this stage, more qualitative case studies from librarian-level experiences are especially valuable. This article describes such a case study. It presents a case of librarians who connect researchers with various data management infrastructure and resources of which the researchers may not be aware. The librarians are also embedded in multiple settings and research stages to provide support to a research-oriented science department.

DESCRIPTION OF SERVICE

Background

The Data Services Librarian at the John Cotton Dana Library has been providing data services to Rutgers University–Newark campus faculty members and graduate students since 2009. She has
assisted faculty members and students with data analysis and visualizations on the basis of individual request and taught general workshops to graduate students about how to use statistical software packages. The Data Services Librarian has also collaborated with the Physical Sciences Librarian and the Life Sciences Librarian on campus to offer data management workshops to faculty members, attempting to introduce some basic data management strategies and information about data management plans.

The Department of Earth and Environmental Sciences (DEES) on the Rutgers University – Newark campus is a research-oriented department with nine faculty members and twenty-seven graduate students who specialize in atmospheric chemistry, environmental geochemistry, fault mechanics and processes, near-surface geophysics, and phytoplankton physiology and biogeochemistry. The Physical Sciences Librarian at the John Cotton Dana Library serves as the subject specialist/liaison librarian for DEES and has been an embedded librarian for the department through participation in their weekly Seminar Series aimed at graduate students, provision of “office hours” in the department, and library research instruction sessions for undergraduate courses.

When the graduate program director of DEES learned that the John Cotton Dana Library provides data services, including data management-related assistance, he contacted the Physical Sciences Librarian for more information. The Physical Sciences Librarian suggested that a workshop presented along with the Data Services Librarian during the DEES’ Seminar Series might be helpful, and the director agreed. Since then, a series of data management-related services have gradually been developed by the two, including surveying the department about
their current data management practices and concerns, providing a workshop during the department’s annual graduate student retreat, and advising on DMPs for grant applications in the department. These services are discussed in greater detail below, except the required data management training for the graduate students in this department, which is described only briefly with an emphasis on the collaboration of the authors. Detailed description of that program can be found in another article by the authors (Fong and Wang 2015).

Preparing to be Embedded Librarians

Be aware of emerging professional practices and resources in data services. To successfully serve users in the present day data-intensive research environment, it is essential that librarians learn as much as possible about the research data lifecycle and stay current with emerging practices. This requires familiarity with both recently updated resources as well as new resources available as published literature or online training modules.

It’s not uncommon for academic librarians to share experiences and learn from each other during conferences, through scholarly writing, and by participating in professional training. Although social science data services, which are not very different from traditional reference services (with the exception of focusing on data access, collection and related services) have existed in academic libraries for a long time, and the International Association of Social Science Information Technology Services (IASSIST) has been organizing annual meetings and supporting this profession for decades, a new type of data service, which focuses more on data management and data curation, began to bloom after 2010 (Xia and Wang 2013). However,
research data services are new and few library school programs offer formal training in this area. Some librarians migrated from the more traditional librarian role into one that focuses on data services. Thus, it’s especially important that professional organizations such as IASSIST and the Association of College and Research Libraries (ACRL) formed interest groups and coordinated conference sessions to facilitate discussions and communications among this new group of academic librarians. In the case of Rutgers University–Newark, the newly hired Data Services Librarian has been actively participating in these professional groups and conferences, which helps to familiarize her with both the current state and future directions of data services. Through interactions and observations of her peers at other institutions, she has developed an awareness of issues, challenges and opportunities for this new librarian role.

With the support of grant funding, a few leading institutions are working on excellent projects that establish some very important groundwork which enables data services librarians to learn about these new academic library services at their own pace. These same training materials can be used to teach data literacy or the best practices in data management. For example, The University of Edinburgh’s MANTRA Research Data Management Training Module (http://datalib.edina.ac.uk/mantra) provides customized training modules for graduate students and researchers as well as DIY (do it yourself) training kits that librarians can further customize for their local environments. In the United States, the Data Information Literacy project (http://www.datainfolit.org), a collaborative of four academic institutions, began with a data management needs survey and then developed into different types of training programs that fit the needs of each institution’s departments. The project Web site gathers all related project outputs (http://www.datainfolit.org/publications) and these publications can be very useful for
librarians developing their own services. Another noteworthy project is the New England Collaborative Data Management Curriculum (NECDNC) (http://library.umassmed.edu/necdmc/index). This customizable instructional tool for data management training includes rich cases and examples from various disciplines. At Rutgers University–Newark, the Data Services Librarian shared these general best practices introductory materials with her science librarian colleagues, and they worked together to customize training for particular departments and other audiences.

**Gain a deeper understanding of your users’ evolving data needs.** As useful as general knowledge of data management strategies and best practices may be, this by itself is not enough. Faculty and students in a department are best served by embedded librarians who understand their discipline-specific needs and know of any special data management requirements, concerns, and resources. At Rutgers University–Newark, the combined knowledge and efforts of both the Physical Sciences Librarian and Data Services Librarian were crucial in order to meet the needs of DEES.

After some investigation, the librarians found that the Earth sciences discipline is actively developing a national data infrastructure to solve many of their data interoperability issues due to the interdisciplinary nature of the subject area. This issue had resulted in rich data assets being in danger of not being usable or reusable even within their own disciplinary community. Earth science scholars from all over the nation are collaborating to solve challenges in research data preservation caused by the huge volume of data and the great diversity in data types. The large-scale National Science Foundation funded project titled “EarthCube” (http://earthcube.org) is a
multi-year project being carried out in an effort to build the necessary data infrastructure. The Data Services Librarian joined the EarthCube virtual community and participated in a group focused on data curation. Through the platform, she learned about a summer institute training opportunity on data infrastructure for geoscientists and applied to participate (http://www.geongrid.org/index.php/education/summer_institute). The librarian’s application was accepted and she travelled to San Diego to attend the weeklong institute. Among the 15 participants were two other librarians and one library school student. The training programs, networking opportunities with trainers, conversations with geoscientists, and collegiate communications with other librarians all helped the Data Services Librarian to become more familiar with the discipline’s needs and issues with regard to data and the ongoing efforts to address them. This helped her feel more confident about working with DEES to discuss data management strategies. In addition, she developed valuable contacts and later was able to ask one of the trainers, who is a leading scholar in the EarthCube project, about suggestions for suitable disciplinary data repositories and better options for data storage and discovery systems for the DEES department.

Apart from the above-mentioned national data infrastructure project, the librarians identified other rich resources that can guide researchers toward better data management practices. One such example is the Data Observation Network for Earth (DataOne) (https://www.dataone.org). This project has built a cyber-infrastructure for multi-scale and multi-discipline data preservation and access and also developed a comprehensive database of data management best practices, including a data management education module that instructors can customize for their own use. Another resource, the Federation of Earth Science Information Partners (ESIP), offers both in-
person and online video data management training for scientists
(http://commons.esipfed.org/datamanagementshortcourse), with the latter being freely available. These discipline-specific training materials allowed the Data Services Librarian and Physical Sciences Librarian to gain a deeper understanding of Earth Science data, infrastructure, and relevant metadata standards and data repositories.

**Embedded Librarianship in Action**

**Data management training: connecting graduate students with data management best practices.** There was much interest from DEES faculty in providing data management training to graduate students in the department. Following a basic introduction to data management during the department’s Seminar Series, the Physical Sciences Librarian and Data Services Librarian had a few informal conversations with the faculty (including the department chair) to learn more about their data management concerns. The biggest issues seemed to be how graduate students are handling research data in the short-term and how to handle long-term data preservation and discovery both within the department and within the broader research community. Familiar with the Data Curation Profile Toolkit (http://datacurationprofiles.org), the authors suggested adapting it to create a questionnaire to survey graduate students in DEES about their current research project – from type and size of data collected to how data is managed (including organization and documentation strategies) to comments about data management challenges they are facing or expect to face in the future (Fong and Wang 2015). All graduate students doing research in DEES were asked to complete the questionnaire. The authors summarized the results and presented it to the department chair and another interested faculty member. The report included
basic facts about the data characteristics, information about back-up strategies, and student concerns. Faculty had some of these same concerns, such as insufficient data storage, inadequate security strategies, and the lack of a clear data organization/management plan. Students wanted more specific guidance and expressed interest in documentation and metadata standards. Many were conscious of the increasing volume of data they are dealing with and the need to better organize and manage that data.

The authors were invited to present a workshop during the department’s annual graduate student retreat addressing some of these concerns – specifically, data organization and documentation (Fong and Wang 2015). Prior to the retreat, DEES faculty asked students to complete a homework assignment that was designed by the authors. This assignment consisted of preparing a short presentation about how they have been handling and documenting their data. After the student presentations, the librarians gave a presentation about the best practices in data handling and documentation, which included sharing some highly useful resources and tools. Students were then given the opportunity to work in small groups to review current data files and discuss with their peers how to apply some of the organization and documentation strategies they just learned. The contents from all DEES data management training sessions have been compiled into a library guide that also includes links to many data management resources relevant to DEES (http://libguides.rutgers.edu/data_EES).

**DMP assistance: connecting researchers with DMP examples & data repositories.** The Data Services Librarian and Physical Sciences Librarian have assisted both faculty and a graduate student in DEES with writing DMPs. While most DMPs have been for NSF grant proposals, one
was for a private grant. When approached for assistance, the librarians seek out the particular directorate’s requirements, a generic DMP template, and examples of DMPs to share with the faculty or graduate student researcher. The researcher is asked to draft a DMP accordingly, and then the librarians provide feedback. Typically, clarification is needed for the sections pertaining to data types and plans for preserving the final research data. Identifying an appropriate research data repository is a common question researchers have. At Rutgers University–Newark, the institutional repository is suitable for most types of data and many of the university’s grant proposals could include it.

In one instance, the authors were asked for assistance in developing a general departmental DMP as part of grant application for purchasing research instruments for DEES. They advised the researcher to focus on data that would be generated from the instrument when writing the DMP – e.g., how these particular data will be generated, documented, and preserved. They provided the researcher with a document on best practices for preparing environmental data sets to share and archive (Hook et al. 2010), which included very detailed guidance on how to create a data dictionary, what fields should be included in data documentation, and how files should be organized and structured. The researcher integrated this information into the DMP and also added into the plan that graduate students in the department would continue to receive training on data management best practices.

In one case concerning a private grant, the librarians were surprised by how well-developed the DMP needed to be at the application stage and the additional requirements expected of successful grantees. The DMP necessitated the identification of a dedicated data manager who
would be responsible for working closely with the funding agency’s data archive staff to ensure the research data would be continuously well-managed during the entire research project. The grantor was developing a data archive and sharing system. It has an overarching DMP in place and a DMP template for each research consortium to use, and makes all existing supported research consortia’s DMPs available to the public. The authors shared these discoveries with the researcher and helped him develop a comprehensive DMP for his grant proposal. The researcher appreciated how much effort librarians put into assisting him with his DMP and expressed interest in collaborating on DMPs again in the future. He was quite pleased with the results and indicated he would be willing to share his plan with other Rutgers University—Newark researchers.

**Data storage & backup: connecting researchers with Computing Services.** Although librarians may theoretically be aware of best practices as far as research data storage and backup are concerned, the library cannot always provide the equipment or technical expertise in these areas. Smaller academic departments may also lack an IT expert on staff to handle such questions. In these instances, librarians might consider connecting researchers with the campus Computing Services unit. Computing Services could advise researchers who have funding as to what kind of equipment they should purchase to meet their short-term and longer-term research data needs. If researchers lack funding, however, Computing Services might advise them on their next best option. In the case of DEES, the department had already purchased and set up a server for storing the local research data. Unfortunately, the server was underutilized and did not provide an option for external collaborators to access the local data stored on it. Faculty members
in the department were interested in suggestions for improving their system of data storage, backup, and access.

The authors contacted Computing Services with their questions. In response, Computing Services asked for more details about the kind of research data collected by DEES so that they might make a more informed recommendation. Results from the modified Data Curation Profile completed by graduate students in the department were summarized and provided to Computing Services. A meeting between DEES, Computing Services, and librarians was then set up. It was suggested that DEES consider purchasing another server to back up the first one. For short-term storage, faculty and student researchers can utilize the campus server, which offers a more secure back up strategy. At Rutgers University–Newark, cloud storage space is available to individual faculty members, graduate students, and/or entire departments. This information was useful for DEES to learn, while increasing their understanding about the research support needs of the faculty was helpful for the Computing Services unit – a win-win for all involved.

NEXT STEPS

Thus far, the authors have worked with DEES on items of most urgent concern to faculty and student researchers in the department – the everyday aspects of data management such as data storage, data organization, and data documentation. However, DEES also has interest in more in-depth training about metadata standards and about preparing data for long-term preservation, including identifying appropriate data repositories. These topics may be discussed at a future department retreat.
While DEES researchers are currently most concerned with access to research data generated by their own department, they recognize that providing access to external researchers is also important. As funding agencies look beyond DMPs and begin focusing on data sharing, so too will researchers and the librarians who support them. After all, preserving and providing access to research materials is at the heart of what academic libraries do – why not do it with research data?

Encouraged by an invitation from a DEES faculty member to be the data manager and a member of his research data team, the authors see working with research groups in DEES as the logical next step. They could help the groups develop and implement concrete data management standards and strategies, and continue to improve these as needed. Deeper involvement with the researchers will enable the authors to even better identify the specific data management issues facing the faculty and students and find the most suitable resources and tools to address those issues. The authors envision a role for embedded librarians that goes beyond just a few educational sessions and DMP consulting assistance, turning towards both broader and deeper connections with the research community.

CONCLUSION

Just as many of the traditional skills librarians possess for finding, organizing, storing, and sharing information are applicable when providing data support, many components of traditional embedded librarianship can be found in the embedded data librarianship model the authors
developed at their institution. At Rutgers University-Newark, however, the Data Service Librarian and Physical Sciences Librarian work together as embedded librarians providing data support for DEES. The combination of data knowledge and skills with subject expertise best serves the needs of faculty and student researchers. Although the librarians may not always be able to immediately address data concerns of researchers, it is understood that both parties are learning together and may even at times seek assistance from other campus units or explore external resources provided by disciplinary organizations or grant agencies. One of the most valuable aspects of embedded data librarianship is the increased engagement between librarians and researchers. As librarians gain a better understanding of researchers’ data needs, they know better where to focus their professional development and training efforts, thereby continuing the lifelong learning cycle and improving how they serve their users.

REFERENCES


http://crln.acrl.org/content/68/6/365.full.pdf


http://crln.acrl.org/content/72/3/167.full


doi:10.7191/jeslib.2015.1067

doi:10.1080/15420353.2013.767765

http://www.istl.org/09-spring/refereed2.html


http://publications.arl.org/rli262/13


doi:10.1371/journal.pone.0114734


doi:10.1177/0340035212473089


doi:10.5860/crl13-435