

2016

Embedded Option: A Common Framework

Cynthia Hudson-Vitale

Washington University in St Louis, chudson@wustl.edu

Follow this and additional works at: https://openscholarship.wustl.edu/lib_papers



Part of the [Library and Information Science Commons](#)

Recommended Citation

Hudson-Vitale, Cynthia, "Embedded Option: A Common Framework" (2016). *University Libraries Publications*. 17.
https://openscholarship.wustl.edu/lib_papers/17

This Book Chapter is brought to you for free and open access by the University Libraries at Washington University Open Scholarship. It has been accepted for inclusion in University Libraries Publications by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

Embedded Options: A Common Framework

Cynthia Hudson-Vitale

EXPERIENCE AND RESEARCH has shown that given the complex nature of research data services, various university units and departments must work together to provide appropriate services to create, manage, store, educate, archive, and preserve research data. Organizationally, this can prove to be a challenge. One viable option for meeting these challenges is the embedded librarian model. In the research data services sense, librarians may be embedded into a faculty-led research group, assisting in creating metadata and managing active data; into the university research office, helping with federal requirements for open data compliance; and into a campus information technology unit providing assistance with big data transfer and data storage issues, to name just a few examples. This chapter provides a common framework that describes the responsibilities and skills of an embedded research data services librarian and then presents various case studies as examples of implementation.

Embedded Research Data Librarian Framework

In general, frameworks serve to provide structure and outline a concept. They provide the scaffolding and common components that describe a system or conceptual schema. The framework for embedded research data services librarians articulates the common skills and responsibilities of the role within the context of location or level of personalized research data services.

Research Data Librarian Responsibilities

The responsibilities and skills expected of a research data services librarian make up a major component of the common framework. A 2003 report titled *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the Na-*

tional Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure was one of the first documents to articulate how new research methods required new technical and social infrastructure to support them.¹ The report highlights the role that digital libraries may play in supporting the development of community data holdings, including the creation of tutorials and documents on data format, quality control, and interchange formatting, as well as tools for data preparation, data fusion, data mining, knowledge discovery, and visualization.

In addition, data information literacy is a useful term to describe a set of skills that data librarians need to provide support and data related services to researchers.² These proficiencies include understanding a researcher's culture of practice; data conversion and interoperability; data curation and reuse; data management and organization; data preservation; data processing and analysis; data quality and documentation; data visualization and representation; databases and data formats; discovery and acquisition of data; metadata and data description; and ethics and attribution. As exemplified in the case studies below, many research data services librarians collaborate with faculty in these areas.

Taking the technical skills and requirements of the research data services librarian a step further, another report published by the National Science Board in 2005 expands the role of the data scientist, or data librarian, as having the responsibility to assist researchers with many of their data-related needs throughout the research lifecycle while directly contributing to innovations in data technology and scholarship. This service and research would include helping researchers conduct research using digital data collections; implementing and developing innovative methods and technologies for data storage, visualization, and discovery; and serving as a mentor to those interested in pursuing data related fields and careers.³

In sum, these reports and scholarly publications provide research data librarians with a set of services and academic pursuits that are most needed by the domain or practice and the faculty with whom they work. A research data librarian may expect to provide outreach, training, and education around data-related skills (such as file management, analysis, cleaning, visualization, and curation, to name a few), while also applying those skills to faculty, researcher, and library projects and ongoing collaborations.

What it Means to Be Embedded

The other core component of the embedded research data services librarian framework is a common understanding of what is meant by embedded librarianship. In general, to be embedded can take many different forms; it does not necessitate being physically located outside of the library, but does focus on providing a high level of personalized services for a researcher. In this role, the embedded librarian should be located where the users are (either virtually or physically) while actively assessing user needs and creating services to fulfill those needs.⁴

The embedded librarian is not a new role in the library profession. In 1993 various authors first began to discuss getting outside of the library to provide additional services of value to users in the research setting.⁵ More recently, Carlson and Kneale define embedded librarianship as a role that enables librarians to create partnerships with their clientele and apply their information expertise in ways that will have a direct and deep impact on research and teaching.⁶ Noticeably absent from this definition is the necessity of being located where the users are; instead, Carlson and Kneale focus on the level and quality of service that an embedded librarian provides. This definition also positions librarians less as service providers and more as collaborators within research groups and to individual faculty, an important reframing of the librarian's role in the research ecosystem.

Carlson and Kneale further outline two modes of embedded librarianship in the research setting that establish these partnerships. The first is a project-based model that involves librarians as collaborators with faculty in a particular research project with defined responsibilities. The second model is programmatic, with a librarian working in an ongoing manner for a research organization to handle various information management tasks. Funding for these types of partnerships can also take many different forms. The embedded librarian may be supported through grant add-ons, co-funded by both the library and the embedded point of service, or even fully funded by the library, with time donated to the embedded service.

Embedded Case Studies

Though an understanding of common research data services and embedded services is helpful, little comparison has been conducted on the common skills and responsibilities of an embedded research data services librarian. To address this gap, embedded research data librarian case studies and examples were mined from the scholarly literature and personal interviews.

Researchers and Research Groups

The embedded data librarian can provide programmatic or project-based services that span the entire research data lifecycle while working with researchers and the research group. These collaborations can develop in a number of ways, but arise predominately through cultivated relationships, grants, and via funded grant add-ons.⁷

The first example provides a programmatic approach to providing embedded research data services at the University of California, Los Angeles. Two librarians worked directly with two distinct research groups to provide a number of data-related services. The first project was funded through the NIH and involved

the librarians providing services to improve data management practices within a research group setting. In this role, the librarians provided presentations on best practices for file management and recommended a metadata schema and file/data versioning software (such as Apache Subversion[†]). The second project was initiated after a researcher attended a university-wide presentation by a librarian about data management practices. For this project, several librarians worked with the research team and led investigations into useful software for data aggregation (such as REDCap[‡]) and best practices for organizing electronic notebook entries.⁸ While not physically located at the research group offices, the librarians regularly attended lab meetings and provided customized solutions to research group data-related challenges.

Customized solutions may also be found in a project-based approach at Washington University in St. Louis. In this instance, a data librarian collaborated with a faculty member in Energy, Environmental and Chemical Engineering to develop and integrate data types into an air quality data catalog. This collaboration was initiated by the librarian and sustained through grant funds for a duration of eight months. The data librarian's activities included grant writing, the registration of data types, the linking of data types to the existing catalog metadata, and the final report writing.⁹ Communication was conducted through weekly meetings, Skype chats, and other mechanisms. The librarian did not have dedicated space for working with the faculty in the research lab, but rather offered a high level of service and regular communication to successfully conclude the project.

Clinical Care Group

In the clinical setting, “informationists”[§] often subsume research data related services. In this role, they provide clinical care group support for patient health and the training of the next generation of medical practitioners. In 2011 Greyson, Surette, Dennett and Chatterley surveyed 191 Canadian health librarians to determine the most common responsibilities for informationists.¹⁰ They found the top responsibilities included searching literature, attending research team meetings, analyzing, scoping or summarizing literature, providing general reference services and maintaining current awareness for the clinical group. Recently, some clinical informationists have begun formally including research data related services in their practice.

† Apache Subversion is an open source versioning control software. It manages files and directories, and tracks changes to them over time.

‡ REDCap (Research Electronic Data Capture) is a secure web application for building and managing online surveys and databases.

§ A term coined by Davidoff and Florence in 2000 to describe the embedded medical librarian.

Beginning in 2012, Informationists at the New York University Health Sciences Libraries provided data-related services to a surgeon-scientist at the NYU College of Dentistry and a protein chemist in a project-based manner. They created an automated literature searching system, introduced tools and workflows for sharing citations, and introduced tools and workflows for sharing data and specimens.¹¹ To develop the automated literature searching system, the librarians wrote a Matlab program to access the NCBI programming utilities[¶] and then processed the results based upon the clinicians' needs.

Similarly, at Johns Hopkins University, informationists in Welch Library work with faculty in the Department of Radiology and Radiation Oncology as part of a grant-funded project.¹² In this capacity, the librarians provide bibliographic instruction to team members, conduct iterative literature searches, and develop guidelines on data sharing for the larger community. As the project has involved large amounts of literature and text-mined data, the informationists are also implementing a number of data management best practices, software, and tools.

Curriculum and Teaching

The embedded data librarian has also been found providing data-related support in the curriculum and through other training opportunities. According to Shulte, librarians embedded in course curricula have overall positive effects on student learning, specifically writing assignments.¹³ While librarians embedding in the curriculum is not a new concept, the skills they are teaching students are new. In addition to, or even instead of, the traditional bibliographic instruction, embedded research data librarians are teaching data information literacy, data science, and data visualization skills.

Purdue University provides a good example of data information literacy skills embedded into an existing course curriculum.¹⁴ Here, librarians partnered with teaching assistants in the School of Engineering to provide data information literacy activities throughout the engineering design cycle. This project was particularly concerned with data management of the code that students produced as part of the course's final project. The librarians were embedded in the course by regularly attending class sessions and holding customized workshops on data management skills.

At Weil Cornell Medical College, a librarian co-taught a course that introduced students to data mining and computational analysis in health informatics.¹⁵ In addition to creating a resource guide, the librarian taught various R skills and computational methods, including such topics as support vector machines,

¶ NCBI programming utilities are a group of eight programs that allow stable access and querying of the databases that make up Entrez.

Monte Carlo method, Naïve Bayes and Adaptive Boosting.[†] Rather than stand-alone workshops on skills related to data information literacy, the librarian integrated those skills within the data science sessions she was teaching. This allowed the students to implement best practices while conducting course related assignments.

Examples of librarians co-teaching or leading data related courses may also be found at Washington University in St. Louis, where a Geographic Information Systems (GIS) Data Librarian has been added to the program faculty¹⁶ for the International and Area Studies (IAS) department. In this role, the librarian teaches GIS courses and workshops and collaborates with other IAS faculty to conduct GIS research. Within the GIS courses the librarian includes data information literacy skills and practices. The librarian also actively attends faculty meetings and oversees the GIS curriculum for the department.

University Research Office

The role of the data librarian in the university research office can take many forms including providing and analyzing institutional author data to assist Research Offices with determining federal funding compliance and providing training and education in data management best practices. At the University of Oregon, the data management librarians partner with the Research Compliance Services to provide data management instruction as a component of the Responsible Conduct in Research Education program for graduate students in Biology, Psychology, Human Physiology, Geology, and Planning, Public Policy and Management.¹⁷ Data librarians are also alerted when a faculty member applies for a federal grant requiring a data management plan. This alert allows the librarians to reach out and offer assistance in reviewing the data management plan and other data related services.¹⁸

Additional Examples

In addition to the organizational units described above, which may be found at many institutions, embedded opportunities exist in other Centers and multi-university projects.

[†] These topics are computational algorithms and machine learning models. Support vector machines are models that conduct supervised learning to analyze data, recognize patterns, classify objects, and conduct regression analysis. The Monte Carlo evaluation is a computational algorithm that conducts repeated random sampling to obtain numerical results in three main areas: optimization, numerical integration, and generating draws from probability distribution. Naïve Bayes is a family of machine learning probabilistic classifiers. Adaptive Boosting is an optimization algorithm.

Biomedical Informatics Service Providers

The Center for Biomedical Informatics (CBMI) at Washington University in St. Louis is currently collaborating with a data librarian to provide reproducible research services for their clinical trial and bio-specimen databases and search queries. Using the outcomes and recommendations of the Research Data Alliance working group on Data Citation: Making Dynamic Data Citable,¹⁹ the data librarian is responsible for the phased implementation and management of a data citation initiative that will serve to make the CBMI databases, queries, and code reproducible and persistent. This collaboration began as a time donation, but the work has proven useful, and thus turned into a grant-funded project in little more than a year's time. With the grant funding, the aims and outcomes have expanded into improving the CBMI research reproducibility. In this example, the data librarian physically locates herself at the CBMI, where she has dedicated space for conducting her work.

Research/Data Communities

Data-related needs are not limited to just an individual team or unit on campus; often, they are a domain and discipline need. It therefore makes sense for data librarians to also heavily involve themselves in the research and data communities that are developing data-related recommendations and procedures that many of the faculty may adopt.

For example, the Sediment Experimentalist Network (SEN), an Earth-Cube Research Coordination Network (RCN), has included a data librarian from the University of Minnesota on its steering committee.²⁰ RCN grants serve to bring together domain researchers to collaboratively address discipline-specific issues in a virtual manner. Many data librarians have played various roles in these grants. For the SEN, the librarian provides expertise in data management planning, community education, and metadata development.²¹ To be successful in this capacity it is important for the embedded research data librarian to attend all available meetings and conference calls. Additionally, the scope and extent of the expertise the data librarian will provide should be made explicit.

Conclusion

Various opportunities exist for data librarians to embed themselves into the academic scholarship within their home institutions, research domains, and data process communities. The case studies above offer examples of research data librarian embedded options that range significantly along a spectrum of ease of implementation (see Figure 2.1).

Figure 2.1. Spectrum of Initiating and Implementing Embedded Services

Embedded Options	Low Barrier	Resource Intensive
	→	
<i>Securing Funding</i>		
Grant Supported		
Time Donated		
<i>Cultivating University Relationships</i>		
E-mail Faculty/Research Office		
Attend University/Department Receptions		
Chat/E-mail about Research Data Services Regularly		
<i>Embedded Services</i>		
Provide DIL Instruction in the Classroom/Lab/Office		
Provide Data Science Instruction in the Class/Lab		
Provide Data Science Services		
Recommend Practices for Data Management		
Act as Data Manager for Lab		
Set up RO DMP Alerting Service		
Review Faculty DMP's with Research Office		
<i>Assessing Services</i>		
Develop a Model to Assess Services		

Providing embedded research data services is not without challenges, though, especially when it comes to scalability. The services that each of the librarians in the case studies provide are incredibly specialized, tailored to the needs of the embedded location, and incredibly time consuming. Rather than supporting the research and teaching of an entire department, they are only reaching a handful of faculty, at most.

Additionally, for many in the university research community (faculty, graduate students, and post-docs), metadata and data information literacy skills are not understood to be essential to conducting research. Thus, the time and effort that are needed for proper data management and outreach to be successful is sometimes lacking.

Some university libraries are also calling on subject liaisons to develop data-related skills to scale up the embedded data model. However, these librarians

will need additional training and free time from other responsibilities to support research data. For many subject liaisons, the question of how to balance traditional librarian activities with data-related activities requires further exploration.

Future research in this area should focus on the impact that the embedded research data librarian model has had upon research and teaching. Data and information that determines if the level of tailored support provided through the embedded research data services model is of enough value to warrant a continued involvement is significantly lacking. To make an informed decision about the viability of this role, a mechanism of evaluation needs to be developed and implemented. Additional metrics of how the embedded research data services model impact compares to the traditional model impact, also do not yet exist. If embedded librarianship is to be understood as an advantage over traditional librarianship, benchmarks need to be established and points of comparison further explored.

Ultimately, to have significant impact and show value to library users, McCluskey finds the embedded librarian model needs to shift from one of support and collaboration, to one of knowledge creation itself.²² She asserts that the work embedded librarians conduct should reflect or build upon the practices of the librarian community. It is through the drive to improve the data librarian ‘practice’ that true embedding takes place.[†]

-
1. Daniel Atkins (chair) *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, National Science Foundation (2003), 42.
 2. Jacob Carlson, Lisa R. Johnston, and Brian Westra, “Developing the data information literacy project” in *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, eds. Jake Carlson and Lisa R. Johnston (West Lafayette: Purdue University Press, 2015).
 3. National Science Board. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. (National Science Foundation, 2005), 27, accessed April 30, 2015 <http://www.nsf.gov/pubs/2005/nsb0540/>.
 4. David Shumaker and Laura Ann Tyler, “Embedded Library Services: An Initial Inquiry into Practices for Their Development, Management and Delivery,” paper presented at the Special Libraries Association Annual Conference, Denver, Colorado, 2007.
 5. Tom Davenport and Larry Prusak. “Blow up the Corporate Library,” *International Journal of Information Management* 13, no. 6 (1993): 405–412.
 6. Jake Carlson and Ruth Kneale, “Embedded librarianship in the research context: Navigating new waters,” *College and Research Libraries News* 72, no. 3 (2011):167–170, accessed March 1, 2015 <http://crln.acrl.org/content/72/3/167.full>.
 7. Lisa Federer, “Embedded with the Scientists: The UCLA Experience,” *Journal of eScience Librarianship* 2, no. 1 (2013): 6–7, doi: 10.791/jeslib.2013.1039.
 8. Vessela Ensberg, Personal Interview, April 24, 2015.

† I would like to thank all of the librarians who allowed me to interview them for this article, especially Vessela Ensberg, Lisa Johnston and Jennifer Moore. Additionally, Lauren Todd and Jennifer Moore were invaluable in providing their feedback and comments on this chapter.

9. “RDA grant” *DataFed.net*, February 25, 2015, http://datafedwiki.wustl.edu/index.php/2015-02-25:_RDU_Project_-_DataFed_Metadata_Stack.
10. Devon Greyson, Soleil Surette, Liz Dennett and Trish Chatterley, “You’re just one of the group when you’re embedded’: report from a mixed-method investigation of the research-embedded health librarian experience,” *Journal of the Medical Library Association*, 101, no. 4 (2013): 287–297.
11. Alisa Surkis, Aileen McCrillis, Richard McGowan, Jeffrey Williams, Brian L. Schmidt, Markus Hardt, and Neil Rambo, “Informationist Support for a Study of the Role of Proteases and Peptides in Cancer Pain,” *Journal of eScience Librarianship*, 2, no. 11 (2013): 35–40.
12. Victoria H. Goode and Blair Anton, “Welch Informationists Collaboration with the Johns Hopkins Medicine Department of Radiology,” *Journal of eScience Librarianship*, 2, no. 1 (2013): 16–19.
13. Stephanie Schulte, “Embedded Academic Librarianship: A Review of the Literature,” *Evidence Based Library and Information Practice*, 7, no. 4 (2012): 122–138.
14. Jake Carlson and Megan Sapp Nelson, “Addressing Software Code as Data: An Embedded Data Librarian Approach,” in *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*, eds. Jake Carlson and Lisa R. Johnston (West Lafayette: Purdue University Press, 2015).
15. Daina Bouquin, “New Experiences with Teaching—Computational Methods in Health Informatics” (blog), June 17, 2014, <http://dainabouquin.com/teachingcompmethods/>.
16. Jennifer Moore, Personal Interview, April 30, 2015.
17. University of Oregon, “Responsible Conduct of Research,” accessed May 15, 2015 <http://orcr.uoregon.edu/content/responsible-conduct-research>.
18. Brian Westra, “Developing Data Management Services for Researchers at the University of Oregon,” in *Research Data Management: Practical strategies for Information Professionals*, ed. Joyce M. Ray, (West Lafayette: Purdue University Press, 2014).
19. Research Data Alliance, “Data Citation WG,” accessed May 15, 2015, <https://www.rd-alliance.org/group/data-citation-wg.html>.
20. EarthCube, “SEN,” accessed May 15, 2015, <http://earthcube.org/group/sen>.
21. Lisa Johnston, Personal Interview, July 11, 2015.
22. Clare McCluskey, “Being an embedded research librarian: supporting research by being a researcher,” *Journal of Information Literacy*, 72 no. 2 (2013): 4–14.