

Embedded Librarians to Support Data-management Needs of a Multidisciplinary Research Program

Ms. Qianjin Zhang, University of Iowa

Qianjin (Marina) Zhang is Engineering & Informatics Librarian at the Lichtenberger Engineering Library, The University of Iowa. Her work focuses on providing data management support and developing library research data services. As a subject librarian, she also provides instruction, reference and consultation services for the engineering faculty and students. Marina holds a MA in Information Resources & Library Science from The University of Arizona (Tucson, AZ), and a BS in Biotechnology from Jiangsu University of Science and Technology (Zhenjiang, China).

Mr. Brian Westra, University of Iowa

Brian Westra is the Data Services Librarian for the University of Iowa Libraries. He assists researchers with the management and sharing of research data. Brian teaches workshops and a course on data management and consults with researchers on data management plans and implementing good data practices across the data lifecycle. He curates data published in the institutional repository, and works with stakeholders across the university to improve data sharing and broaden the impact of research through FAIR-aligned practices and infrastructure.

Embedded Librarians to Support Data Management Needs of a Multidisciplinary Research Program

Abstract

This paper describes the establishment of a partnership between the Libraries and a multidisciplinary research program, and some of the products and outcomes from immersive and embedded roles within that program. Several factors contributed to the development of this partnership: outreach efforts by the Engineering Library and the Data Services Librarian to faculty, staff, students, and research administrators; a research program director who has a history of engagement with the Libraries; and the funder's data management and sharing mandates in the funding opportunity announcement for the research program.

The result has been the inclusion of two librarians in the Data Management and Analysis Core of this program, which is expected to continue for the duration of the 5-year funding cycle. This approach yielded the launch of a data management course, and ongoing development of program- and team-specific guidance and strategies to improve data sharing and integration. This paper indicates that these kinds of partnerships can increase the awareness of librarian skills in research data management, support compliance with funder requirements, and enhance the impact and value of research outputs.

Introduction

The University of Iowa Libraries has been providing research data services to faculty, students and researchers for several years. A Research Data Interest Group (RDIG) was initiated in 2015 by an Engineering and Informatics Librarian and a Sciences Instruction Librarian to support the development of the Libraries' research data services. The RDIG included librarians and campus partners from Information Technology Services, the Iowa Informatics Initiative, the Division of Sponsored Programs, and other stakeholders. The RDIG facilitated communication about research data management topics, and attempted to raise institutional awareness of data management needs and opportunities by hosting several seminars and workshops, and advocating for the creation of a Data Services Librarian position in the Libraries in 2017.

The Data Services Librarian has led efforts to build research data services through outreach to research faculty, staff and students, and deeper connections with institutional stakeholders. These actions have resulted in a set of services that includes data management plan consultations, workshops and course-based instruction, data curation for deposits in the institutional repository, and involvement in institutional data policies discussions.

The Iowa Superfund Research Program (ISRP) is a multi-project center grant, funded by the National Institute of Environmental Health Sciences (NIEHS) since 2006, to conduct collaborative research on sources, exposures, toxicities, and remediation of polychlorinated biphenyls (PCBs). This multidisciplinary program is currently composed of 22 faculty, 10 staff, and 25 students (also called trainees), from civil and environmental engineering, biomedical engineering, microbiology and immunology, medicinal and natural products chemistry, urban planning, human toxicology, and occupational and environmental health. The program is funded through 5-year grants, and funding for the 2015-2020 fiscal years including supplemental administrative funds for short-term projects was a little over \$12 million.

In the 2021-2025 grant, there are five project teams conducting research related to airborne PCBs. These project teams are supported by four program-wide support cores, and two research support cores. The Program Support Cores provide infrastructure and build and sustain research capacity through training trainees (Research Experience and Training Coordination Core), engaging with affected communities (Community Engagement Core), administering the grant and disseminating discoveries (Administrative Core), and the Data Management and Analysis Core [1].

In addition, there are two Research Support Cores. The Synthesis Core synthesizes and authenticates chemical compounds and analytical standards, and the Analytical Core provides chemical analyses services and support.

The Data Management and Analysis Core (DMAC) is a new core required for all proposals to the most recent funding opportunity announcement for this program. The NIEHS added this core to formalize and enhance each program's capacity to support data management and sharing, and data analysis. The DMAC has two sets of goals: facilitate data management, integration, sharing and interoperability of data assets in alignment with FAIR data principles (Findable, Accessible, Interoperable, and Reusable); and coordinate and manage data analysis and data quality throughout the lifecycle of data products [2].

Literature Review

Research data management services offer an opportunity for librarians to apply their skills toward organizing, describing, preserving and sharing information, meeting a growing demand for services that can assist researchers with sharing data [3]. These services intersect with the data lifecycle at multiple points, and can be developed and sustained through partnerships with researchers and institutional stakeholders to provide the training, infrastructure, and services necessary to facilitate good data management and sharing practices. It may be useful to consider the challenges of data management as a "wicked problem" [4] which requires innovative thinking, flexibility, and collaboration. Put another way, successful research data services are frequently associated with approaches that may fall under the rubric of embedded librarianship.

For the purposes of this paper, the authors consider embedded librarianship to encompass a broad spectrum of practices and implementation pathways [5-7]. Some of the attributes of this approach to research data management support are: access to and collaboration with the research team to understand their research methods and data practices [8], coupled with librarian experience and interest in acquiring knowledge, skills and abilities in data management [7]; and a willingness to innovate, take risks and experiment [8, 9]. On an interpersonal level, it requires developing and sustaining trusted relationships between the data or liaison/subject librarian(s) and researchers [9]. At the organizational level, the support of library administration and institutional prioritization of data services are critical factors [8].

The depth and duration of these interactions with researchers may range from shorter-term “project-based” interactions [9], to “deep dives” into the data practices of an individual or team [10, 11], to long-term, “program-based” relationship-building and collaborations with a program or department [7].

Some of the outcomes of these service and support frameworks in an academic setting may include: Responsible Conduct of Research training [12]; data information literacy training [13]; data management plan consultations [14]; active data management infrastructure and training [15, 16]; infrastructure and services for data curation [17] and alignment with FAIR data principles [18], and repository services and development [19, 20].

Establishment of Partnership between the Program and the University Libraries

The ISRP Principal Investigator (PI), an environmental engineering professor, has a long-standing working relationship with the Engineering Library which is part of the University Libraries. The PI also has a strong interest in data management and sharing and has attended the Libraries’ data seminars and workshops prior to taking over the ISRP.

As noted above, the NIEHS included a requirement for a DMAC in the 2019 call for proposals. This led the PI to invite the Data Services Librarian to join the program’s team in the proposal writing process. The Data Services Librarian also recommended the Engineering and Informatics Librarian to the PI. With approval and support by library administration, the librarians participated in the ISRP grant-proposal process, contributing to the DMAC section of the proposal, and providing a data management plan template for the data sharing statements required for several components of the proposal. The Engineering and Informatics Librarian also helped format bibliographies for the entire proposal.

This participation in the proposal-writing process set the stage for further engagement with the program. The proposal review and funding decision took about a year, and while the librarians were not yet formally part of the grant, they were deeply engaged with the program. In addition, a supplemental administrative grant included the librarians as key personnel immediately prior to the anticipated startup date of the new grant. During this time, the librarians were invited to participate in the PI’s weekly research group meetings to learn about and respond to data

management and sharing needs. Data management considerations across the research lifecycle were incorporated into the meeting agenda so the librarians could present on these topics and facilitate discussions about best practices and resources. These presentations offered an informal setting in which instructional materials and examples could be developed, and became the basis for a one-credit graduate-level Research Data Management course for ISRP trainees.

The proposal was accepted for funding by NIEHS for 2020-2025, starting in April of 2020, with both librarians funded at 10% of their time. There are a range of aims for the DMAC related to data management and sharing that span the five years of this grant. Outlined below are some of our activities during the first year.

Outcomes

Research Data Management Course

A one-credit graduate-level course called Research Data Management Seminar was offered through the College of Engineering's Civil and Environmental Engineering department (CEE:5110). Because the Libraries are not an academic department, the program director served as the course supervisor, and facilitated the process of creating this new course listing. The course was initially offered in-person and moved online during the COVID-19 pandemic. The syllabus covered foundational data management and sharing topics and used data management planning, and data curation as frameworks for applying these concepts. The instructors took advantage of the wide variety of openly accessible data information literacy instructional materials and resources that have been developed by data librarians, curators, and others. These include lessons, examples, activities and curricula from Data Carpentry [21], DataONE [22], FAIR stakeholders [23, 24], Data Information Literacy outputs [25] and others.

In particular, the course covered essential practices in research data management including file organization, data documentation, lab notebook best practices, tabular data structure, and cleanup tools. The course also introduced factors to consider when choosing how and where to share and publish data, including metadata and identifiers, FAIR data principles, licensing, and citation. The course style was a mix of lectures and small group discussion. The assignments included creating a data management guide and preparing a dataset, with metadata (readme file, data dictionary, DataCite metadata) for publication. Most of the students in the course were ISRP trainees because the course was recommended for them, but one non-ISRP student and two ISRP staff members also enrolled in the course. The course evaluation indicated student satisfaction and excitement. A few comments are shown below:

“It is helpful in understanding what data management is and its importance, can be especially useful if one is taking the course at early stage of research.”

“I feel like I have a better grasp of what should go into a data management plan and how data should be prepared for sharing. So that's great!”

“[The instructor] is very knowledgeable and was able to point us toward good resources.”

In an effort to expand awareness of and outreach for the course, the spring 2021 course was cross-listed with the College of Public Health’s Occupational and Environmental Health department (OEH:5110).

DMAC Activities

In addition to participating in research group meetings and teaching the course, the librarians are members of the DMAC led by a Professor of Biostatistics and a Professor of Engineering, Planning, and Environmental Policy within the ISRP. In this role, the librarians are collaborating with PIs and project teams to establish and implement strategies to improve data sharing and integration across projects. These include collaborations with project team to develop guidance and best practices for data structure to facilitate data analysis, and plans to explore controlled vocabularies and infrastructure to improve data sharing and alignment with FAIR data principles. So far, the librarians have created an electronic request form for data analysis advice, developed the tabular data structure guidance and assisted ISRP researchers with data publishing.

In her role within the DMAC of promoting consistent data practices, the Engineering and Informatics Librarian facilitated a discussion on data preparation with the DMAC and Analytical Core staff members and created a tabular data structure guide with several examples to address issues occasionally found in data submitted by trainees seeking data analysis support. The guide outlined a set of best practices for tabular data organization, key components in a ReadMe file and data dictionary, and examples which were collected from the ISRP researchers and improved based on data management best practices and other guidance. The tabular data structure guide has been circulated among the ISRP research project team leads.

The Data Services Librarian has assisted ISRP researchers with complying with NIEHS data sharing requirements by curating datasets for deposit in the University Libraries’ institutional repository, and consulting with researchers on the use of domain-specific repositories. The data curation workflow and examples of deposits have been useful for instruction and curation within the program, and as exemplars for working with other researchers and campus partners.

These data curation activities have also informed library decisions about repository systems, metadata support, and data deposit and publishing processes. Researchers view local curation support and the institutional repository positively as they gain experience with and see the value added by robust metadata, responsive staff, and a metadata schema and repository system that is aligned with their needs and FAIR data principles. Although there are relatively few deposits in the institutional repository so far, it is anticipated that these numbers will increase as the benefits of curation are recognized. Based on his data curation experience, the Data Services Librarian has developed data deposit guidance and will incorporate it into the Libraries’ research data services website in upcoming revisions and updates.

Conclusions

By embedding two librarians in a large multi-year grant to provide services and infrastructure across the data lifecycle, the Libraries and this program have embarked on a fairly unique partnership.

Building relationships with researchers through embedding in a grant enables the conversations and collaborations that are at the core of a responsive and engaged data services program. The libraries provide a range of one-shot workshops on research data topics, which can also facilitate subsequent connections with researchers for data management assistance. In both cases, these interactions serve to increase awareness of services, infrastructure, and resources, sometimes leading to follow-ups initiated by researchers. Opportunities to work directly with the research community are trust-based, and reliant on solutions and partnerships that can contribute to desirable outcomes. Demonstrating impactful outcomes, such as more efficient data workflows, broader discovery and reuse of data, reproducibility, compliance with funder policies, and the possibilities that are enabled by alignment with FAIR data principles, does result in increased uptake of these services.

The return on investment both for the program and for application to other research groups has more than justified the work this has required. Several factors would impinge on scaling up this embedded approach to multiple simultaneous projects. Broader expansion of data services as a program with deeper integration across the institution would likely require additional Libraries staff, or expanding expectations for involvement of more liaison/subject librarians. This is a challenge that is not unfamiliar among higher academic institutions, particularly in the face of shrinking budgets. Librarians as self-learners could attend a variety of trainings or conferences to obtain data management skills and experience, and a workshop was previously held for those interested in building an understanding of essential data management practices and the data lifecycle.

As with most research projects, time requirements for the librarians have been dependent on research milestones and outcomes, and the academic calendar. The COVID-19 pandemic negatively impacted in-person research capacity, but seemed to also shift work toward data analysis, curation and publishing. In the lead-up to the course, instructional design and course development might require an investment of over 50% of available time per librarian at the beginning of the semester in the initial year of the course, and less time as the course was fully developed. Library administrative support to provide this flexibility is critical.

Probably the other most significant challenge is also relatively common, and that is the need for tools and systems that can accommodate and enable more streamlined approaches to data integration, without relying on local application development. The librarians are hopeful that the DMAC community will continue to share insights into infrastructure and practices that can improve all of their capacities for knowledge management.

The two librarians have enjoyed being engaged in the opportunities that this collaboration with ISRP has opened up to them and are thankful for library administrative support. They plan to continue participating throughout the duration of the grant (2020-2025). The strategic plan for DMAC activities, with input from the program's external advisory board and the NIEHS will continue to shape data services within the grant, and the librarians' awareness and responses to the needs of ISRP faculty, staff and students. This experience has already provided use cases that exemplify librarians' potential when librarians discuss data management and sharing services with other campus partners in research, technology, and administration.

Future Activity Plans

There are several activities planned for the coming year and beyond. In the current, second year of the research data management course, "flipped classroom" approaches have been used more frequently. Based on student feedback, the creation of a data management plan through cumulative assignments have been replaced with more use cases to directly apply concepts. The Data Services Librarian is collaborating with other data librarians to share instructional curricula and materials among Superfund programs. The course will continue to be revised to strengthen an evidence-based learning philosophy, and to reflect feedback from students about the content and alignment with application to their work.

To improve curation of data produced by the ISRP, the Data Services Librarian proposed the creation of a data curation primer for mass spectrometry data, which was accepted by the Data Curation Network [26]. This project is a collaboration with chemistry librarians from several institutions and is nearing completion.

Last but not the least, the DMAC team plans to meet with each project team to review data sharing plans and synchronize them with program-wide guidance and practices. The librarians are also interested in exploring infrastructure and practices that might leverage existing ontologies and controlled vocabularies toward linked data. The NIEHS-hosted quarterly meetings of DMAC personnel from all programs provides opportunities to learn about and explore infrastructure and resources from other programs and the librarians plan to collaborate with other DMAC teams and data scientists in these areas.

References

- [1] The University of Iowa. (2021, April 17). *Iowa Superfund Research Program* [Online]. Available: <https://iowasuperfund.uiowa.edu/>.
- [2] National Institute of Environmental Health Sciences. (2018). *Superfund Hazardous Substance Research and Training Program* [Online]. Available: <https://grants.nih.gov/grants/guide/rfa-files/RFA-ES-18-002.html>.
- [3] S. Pinfield, A. M. Cox, and J. Smith, "Research data management and libraries: Relationships, activities, drivers and influences," *PLoS ONE*, vol. 9, no. 12, 2014.

- [4] A. M. Cox, S. Pinfield, and J. Smith, "Moving a brick building: UK libraries coping with research data management as a 'wicked' problem," *Journal of Librarianship and Information Science*, vol. 48, no. 1, 2016.
- [5] S. Schulte, "Embedded Academic Librarianship: A review of the literature," *Evidence Based Library and Information Practice*, vol. 7, no. 4, pp. 122-138, 2012.
- [6] O. Olivares, "The sufficiently embedded librarian: Defining and establishing productive librarian-faculty partnerships in academic libraries," *Public Services Quarterly*, vol. 6, no. 2, pp. 140-149, 2010.
- [7] M. Wang and B. L. Fong, "Embedded Data Librarianship: A Case Study of Providing Data Management Support for a Science Department," *Science and Technology Libraries*, vol. 34, no. 3, pp. 228-240, 2015.
- [8] H. Coates, "Building Data Services From the Ground Up: Strategies and Resources," *Journal of eScience Librarianship*, vol. 3, no. 1, pp. 52-59, 2014.
- [9] J. Carlson and R. Kneale, "Embedded librarianship in the research context," *College and Research Libraries News*, vol. 72, no. 3, pp. 167-170, 2011.
- [10] M. S. Bracke, "Emerging Data Curation Roles for Librarians: A Case Study of Agricultural Data," *Journal of Agricultural & Food Information*, vol. 12, no. 1, pp. 65-74, 2011.
- [11] J. Carlson and M. Stowell-Bracke, "Data Management and Sharing from the Perspective of Graduate Students: An Examination of the Culture and Practice at the Water Quality Field Station," *portal: Libraries and the Academy*, vol. 13, no. 4, pp. 343-361, 2013.
- [12] L. Johnston, M. Lafferty, and B. Petsan, "Training Researchers on Data Management: A Scalable, Cross-Disciplinary Approach," *Journal of eScience Librarianship*, vol. 1, no. 2, 2012.
- [13] J. Carlson and L. Johnston, *Data information literacy : librarians, data, and the education of a new generation of researchers*. 2015, pp. 271-271.
- [14] G. Steinhart, E. Chen, F. Arguillas, D. Dietrich, and S. Kramer, "Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements," *Journal of eScience Librarianship*, vol. 1, no. 2, 2012.
- [15] A. L. Whitmire, "Implementing a Graduate-Level Research Data Management Course: Approach, Outcomes, and Lessons Learned," *Journal of Librarianship and Scholarly Communication*, vol. 3, no. 2, pp. eP1246-eP1246, 2015.
- [16] J. N. Jeffryes and L. Johnston, "An e-learning approach to data information literacy education," 2013, pp. 12-12.
- [17] L. Johnston *et al.*, "Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services," *Journal of eScience Librarianship*, vol. 6, no. 1, pp. e1102-e1102, 2017.
- [18] K. M. Hettne, P. Verhaar, E. Schultes, and L. Sesink, "From FAIR leading practices to FAIR implementation and back: An inclusive approach to FAIR at leiden university libraries," *Data Science Journal*, vol. 19, no. 1, pp. 1-7, 2020.
- [19] R. Rice, "DISC-UK DataShare Project: Final Report," 2009, [Online]. Available: <http://ie-repository.jisc.ac.uk/336/>.
- [20] M. P. Newton, C. C. Miller, and M. S. Bracke, "Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force," *Collection Management*, vol. 36, no. 1, pp. 53-67, 2010.
- [21] D. Paul and C. Macdonell. (2017). *Data Carpentry: Data Cleaning with OpenRefine Ecology lesson (2017.04.0 ed.)* [Online]. Available: <https://datacarpentry.org/OpenRefine-ecology-lesson/>.
- [22] DataOne. (2012). *DataONE Education Module: Data Entry and Manipulation* [Online]. Available: <https://old.dataone.org/education-modules>.
- [23] Australian National Data Service. *The FAIR data principles* [Online]. Available: <https://www.ands.org.au/working-with-data/the-fair-data-principles>.
- [24] S. Jones and M. Grooteveld. (2017). *How FAIR are your data?* [Online]. Available: <http://doi.org/10.5281/zenodo.1065991>.
- [25] J. Carlson and M. S. Nelson. Data Information Literacy Case Study Directory [Online] Available: <https://docs.lib.purdue.edu/dilcs/>
- [26] *Data Curation Network* [Online]. Available: <https://datacurationnetwork.org/>.