

**Bridging the Gap between Scientists, Institutional Repositories and Data Management Practices****Devan Ray Donaldson, PhD/Indiana University/IMLS Early Career Development****Project Description**

In this Early Career Development project, Dr. Devan Ray Donaldson from the Department of Information and Library Science at Indiana University Bloomington requests \$295,695 from the Laura Bush 21<sup>st</sup> Century Librarian program for a three-year empirical investigation into the use of Institutional Repositories (IRs) by new users in order to address the following research questions:

- How do scientists use IRs during Data Management Plan (DMP) implementation?
- What barriers arise as scientists use IRs during DMP implementation?
- How do IR staff interpret and respond to information about scientists' use of IRs?

IRs are a set of services that universities offer to members of their communities for the management and dissemination of digital materials created by the institutions and their community members (Lynch, 2003). IRs allow faculty and other researchers to self-deposit a broad range of scholarly outputs including: journal articles and preprints, presentations, theses and dissertations, project reports, video, lab publications, and more recently, research data. Increasingly, funding bodies are requiring grant-holders to develop and implement DMPs that “state what data will be created and how, and outline the plans for sharing and preservation, noting what is appropriate given the nature of the data and any restrictions that may need to be applied” (Digital Curation Centre, 2018, n.p.). IR staff and other information professionals, in turn, have focused on supporting researchers with the creation and implementation of DMPs (Novak & Day, 2018; Witt, 2008). In practice, IRs vary in the services they provide, and scientists vary widely in their creation and implementation of DMPs as well as their utilization of IRs during this process. The aim of this project is to investigate how scientists actually use IRs during DMP implementation, identify any barriers, and to understand the perceptions of IR staff about these barriers in order to continue bridging the gap between scientists, their data, and IRs as they co-evolve.

This project takes seriously the National Research Council's (2015) recommendation that “educators in institutions offering professional education in digital curation should create cross-domain partnerships with educators, scholars, and practitioners in data-intensive disciplines and established data centers” (p. 79). In a series of case studies, LIS graduate students enrolled in the PI's digital curation courses will partner with scientists to study how they: implement their existing data management plans; identify which data should be deposited into IRs; prepare the data for deposit; deposit the data into IRs; and publicize access to the deposited data. The students will observe and support the scientists as they undertake these activities, recording their experiences in field notes and diaries, noting any challenges that may arise. To increase the generalizability of the study, the PI has already established contacts with scientists from a diverse range of disciplines (e.g., atmospheric sciences, chemistry, computer science, ecology, and neuroscience) that vary in how disposed they are to data sharing in practice. Additionally, the PI has established contacts with staff at three IRs that vary in the software platforms they use, but match the institutions where the scientists' lab groups are located; the IR staff have expressed an interest in allowing the PI to collect empirical data about their perceptions of the scientists' use of the IRs as it relates to the management of their data.

The PI has already conducted research on data management, including data sharing and digital repositories from the users' (Donaldson et al., 2017; Fear & Donaldson, 2012) and staff's (Donaldson & Bell, 2018; Donaldson & Conway, 2010; Donaldson et al., 2016) points of view. Building on his existing research agenda and using case study methods, Donaldson will investigate a diverse range of scientists and IRs where the commitment to supporting scientists with data management is real, but engagement from scientists could be improved upon and better understood.

**Project Design and Methods**

Scientists representing fifteen lab groups from five disciplines have committed to this project, enabling the PI to study their data management practices and to study the incorporation of use of IRs into those practices. Additionally, three IRs have agreed to partner in this research, providing the PI with access to personnel (e.g., IR staff). Donaldson will employ a multidimensional approach to inquiry and methods, from ethnographic

**Bridging the Gap between Scientists, Institutional Repositories and Data Management Practices**

Devan Ray Donaldson – IMLS, LB21 Early Career

observation, interviews, and case study. A graduate student researcher (GSR) will be recruited to assist the PI during the project; the PI will mentor the GSR on multiple aspects of the research enterprise, including data collection, data analysis, and writing for publication. Four experts have agreed to serve as advisors: John Chodacki (California Digital Library), Margaret Hedstrom (University of Michigan), Clifford Lynch (Coalition for Networked Information), and Carol Tenopir (University of Tennessee); they all have expertise in digital curation, IRs, and research data management.

**Broad Impact**

This project will ultimately result in a better understanding of researcher data management practices and needs, and thus a better ability to design and position services to meet these needs. To measure the impact of this project on the scientists, the IRs staff, and the LIS students who participate, the PI will deploy a series of pre- and post-tests (e.g., surveys) regarding: the scientists' perceptions of the management of their data for the long term, including data access and preservation; the IR staffs' understanding and awareness of the new users' data management issues and challenges; and LIS graduate students' understanding of how to support the management of scientists' data. This study is intended to provide a new model for LIS graduate education related to digital curation by actively providing students with opportunities to build relationships with scientists across domains to understand and support their data management practices. Research that creates these types of opportunities for LIS students is critical to positioning LIS graduate programs for 21<sup>st</sup> century practice (Sands et al., 2018).

**Performance Goals and Outcomes**

In years 1 and 2, the PI will secure IRB approval; recruit a GSR for the project; develop interview protocols and data collection instruments; deploy pre-test surveys; observe and support the scientists with use of IRs for data management; recruit interview participants (e.g., IR staff); meet annually with advisory board and present preliminary findings at professional conferences. During years 2 and 3, the PI and GSR will deploy post-test surveys; code data collected from years 1-2 for themes using NVivo; compute inter-rater reliability, and synthesize findings for publication. Outputs of the project will include the publication of results in scholarly journal articles and peer-reviewed conference proceedings (e.g., *Library Quarterly*, *College & Research Libraries*, *Open Repositories*; *IDCC*); an open educational report (OER) intended for use by practitioners and educators in library programs and iSchools; SAA Research Forum presentations; CNI breakout sessions; LITA conference presentations; ACRL virtual conference webcasts; and a final report to IMLS.

**Diversity Plan**

The PI will collaborate with underrepresented minority students in the IU-Undergraduate Research Opportunities in Computing (UROC) program and summer research students from Historically Black Colleges and Universities (HBCUs) and other Minority Serving Institutions (MSIs) through the IU-MSI STEM Initiative.

**Budget Summary**

Indiana University overhead rates of 57.5% suggest an overall budget of \$295,695 for the project, which includes summer salary support for the project director for three years, a Doctoral Graduate Student Researcher for three full years, travel to field sites, research supplies, and dissemination of findings in professional and academic venues.

**Project Director**

Devan Ray Donaldson, PhD, is an Assistant Professor in the Department of Information and Library Science at Indiana University Bloomington where he directs specializations in Digital Curation and Archives and Records Management. He has published fifteen articles on a broad range of issues regarding research data management, digital repositories, and their users. He is a recipient of the 2017-2018 Indiana University Trustees Teaching Award in recognition of excellence in teaching.