

Project Title: Library Capacity Assessment and Development for Big Data Curation

National Leadership Grants for Libraries: Planning Grant; National Digital Platform

Summary: The research team from Indiana University Indianapolis (IUPUI) and Indiana University Bloomington (IUB) will develop a conceptual framework for assessing libraries' capacity for big data curation. The purpose of this study is to provide a research-based foundation for a full project, which will lead to the development of a big data curation capacity assessment toolkit that will be freely usable by any library.

Statement of Need

Data are increasingly being recognized as “first-class intellectual objects” that support scientific inquiries and improve the quality of human life, and libraries become responsible for curating data in order to satisfy their mission by protecting and disseminating data.¹ While more libraries—mostly academic but a few public libraries as well—have been trying to establish data services, recent examination of libraries' data services has demonstrated that there is a significant level of variation in their services and programs, perhaps due to their differing organizational capacities.² This is more true with so-called big data, “high-volume, velocity and/or variety information assets.”³ Because big data are fundamentally different from “spreadsheet science,” curating them requires “innovative forms of processing that enable enhanced insight, decision-making and process automation.”⁴ Many have asserted the necessity of building sustainable and scaleable data curation programs in libraries that will be effective in continuing to support communities and scholarship that may be overwhelmed by big data,⁵ but a holistic approach to understanding libraries' capacity for big data curation is needed.

The goal of this project is to develop a conceptual framework for assessing libraries' capacity for big data curation, which will be essential in implementing sustainable and scaleable big data curation programs. Big data curation programs require sufficient capacity in various dimensions, such as technology infrastructure, policies, values, skills, and culture as well as leadership. Defined as “the ability to perform tasks and produce outputs, to understand and solve problems, and make informed choices,” capacity is a useful concept to assess libraries' necessary conditions to perform curation activities as well as the potential for success. Previous literature suggests that assessing capacity is a prerequisite for interlinked decisions in any organization, such as strategic and operational choices, ongoing policy dialogue, and further capacity development processes.⁶ Data curation programs must be tailored to existing capacity to be implementable in the long term, so they can ensure effectiveness and avoid unintended consequences.

In order to develop a framework to assess capacity, the overarching aim of this project is to define capacity in the context of big data curation and to identify its various dimensions from existing literature on the topic found in various fields (e.g., organizational studies, development studies, economics, education, etc.) within the context of libraries and big data curation based on the examination of current library practices.

Intended Results and National Impact

As an outcome of this planning grant, our project will produce a conceptual framework for defining big data curation capacity with details of the dimensions that comprise a library's capacity and specify necessary conditions to increase capacity. This initial investigation will lead to a larger project aimed at developing an assessment toolkit that will be freely usable by any library.

Our project will have a major impact on libraries, both academic and public. Despite agreement on the importance of having big data curation programs in libraries and recent efforts to build such programs, there has been little previous effort to systematically understand the library's capacity to perform curation work and build

¹ Heidorn, P. B. (2011). The emerging role of libraries in data curation and e-science. *Journal of Library Administration*, 51(7–8), pp. 662–672.

² Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: A content analysis of libraries' websites. *College & Research Libraries*. Preprint retrieved from: <http://crl.acrl.org/content/early/2016/11/16/crl16-948.full.pdf>

³ LeHong, H., & Laney, D. (2013). Toolkit: Board-ready slides on big data trends and opportunities. *Gartner*.

⁴ Ibid.

⁵ NISO (2013). *Webinar: Research Data Curation, Part 2: Libraries and Big Data*.

http://www.niso.org/news/events/2013/webinars/data_curation/

⁶ European Commission. (2005). *Institutional assessment and capacity development: Why, what and how?*

a sustainable program. A conceptual framework will assist library staff in understanding their current environments as well as potential impediments to building a successful curation program. Additionally, the framework will be used in a subsequent project to develop a standardized capacity assessment instrument, which will be tested, evaluated, and freely available to any library upon its full development.

Project Design

Major project activities include: (1) conducting a systematic review of the literature on organizational capacity, data curation, and big data practices; (2) conducting a large scale landscape survey with libraries to learn about their current practices and/or future efforts planned; (3) developing a framework for library capacity based on the findings of (1) and (2); and (4) vetting the framework with a panel of leading experts who handle various aspects of big data curation in academic and public libraries. Our literature review will first focus on the concept of institutional/organizational capacity from organizational studies, development studies, information and library science, information technology, and other relevant disciplines, to understand the meaning of capacity and identify capacity dimensions. We will also conduct a review of literature on big data and curation practices to contextualize our capacity review. Next, we will develop an online survey to understand current library data curation practices. This will complement our understanding of the literature on library practices. Survey questionnaires will be reviewed by our advisory board and distributed online through library association listservs (e.g., RDAP, PLA). Outcomes from (1) and (2) will enable us to define capacity as it pertains to big data curation by mapping current big data curation practices with organizational capacity concepts and dimensions, and will be used to produce a framework for institutional capacity for big data curation.

We will use several strategies to evaluate this planning project and outcomes. After creating the conceptual framework, we will invite 7 advisory board members (from both academic and public libraries) to come to Indianapolis, IN to review our framework and provide feedback. The advisory board members will participate in a Delphi study aimed at generating a consensus on what constitutes capacity for big curation while considering our framework. The advisory board will include representative experts from the digital curation discipline, who possess the expertise to give us feedback on our project progress and outcomes: **Eben English**, a librarian at the Boston Public Library, **Robert McDonald**, the Associate Dean for Research and Technology Strategies at IUB, **Nancy McGovern**, Head of Curation and Preservation Services at the Massachusetts Institute of Technology, **Elaine Westbrooks**, the Associate University Librarian for Research at the University of Michigan, **Michael Witt**, the Head of the Distributed Data Curation Center (D2C2), Purdue University Libraries, and **Tim Rogers**, Executive Director at Metropolitan Library System. (One to be determined).

The project information and outcomes will be circulated through multiple venues, such as the project's website, publications, conference presentations (e.g., International Digital Curation Conference), and social media, for any comment and feedback. Presentations and peer-reviewed publications will help us to evaluate and improve our work.

Project Personnel

Ayoung Yoon, Ph.D. is the **PI** and an Assistant Professor in the Department of Library and Information Science at Indiana University Indianapolis (IUPUI). She is also a RDA/US data share fellow. Her dissertation, *Data Reuse and Users' Trust Judgments: Toward Trusted Data Curation*, received the Eugene Garfield Doctoral Dissertation Award in 2015. Her research focuses on data curation, data reuse, and open data.

Devan Ray Donaldson, Ph.D. is the **co-PI** and an Assistant Professor in the Department of Information and Library Science at Indiana University, Bloomington (IUB). He is an internationally known digital curation researcher who studies digital repositories and data sharing practices. His research has been funded by the Alfred P. Sloan Foundation and the United States Department of Energy.

Budget Requested: \$49,997 (Total direct costs: \$31,744, Indirect costs (F&A @57.5%): \$18,253)

- Project personnel: \$20,044 (PI and Co-PI summer salaries, Graduate Assistant (\$13/hx200) with summer FICA)
- Conference travels: \$4,000
- Advisory board: \$7,700 (Travels to Indianapolis, hospitality)