

The DataRes Research Project on Data Management

By: Martin Halbert, William Moen, and Spencer Keralis

“The DataRes Research Project on Data Management,” (with William Moen and Spencer Keralis) in *Proceedings of the 2012 iConference*. Edited by Jens-Erik Mai. ACM, New York, 2012, pp. 589-591. <http://dx.doi.org/10.1145/2132176.2132300>

© The authors | ACM 2012. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in ACM Digital Library, <http://dx.doi.org/10.1145/2132176.2132300>.

Abstract:

The University of North Texas together with the Council on Library and Information Resources, have received \$226,786 from the Institute of Museum and Library Services for a two year research project to investigate how the library and information science profession can best respond to emerging needs of research data management in universities. This project will address broad new issues concerning the emerging roles, expectations, and practices arising from requirements announced by NIH, NSF, IMLS and other funding agencies for data management plans as part of proposals.

Keywords: Management | Documentation | Security | Legal Aspects | Verification | Research data management | Policy

Article:

1. Introduction

Federal agencies including NSF, NIH, and IMLS have now mandated data management plans as requirements for submitting research grant applications. As a result, research universities across the country are now struggling to develop consistent policies and programmatic implementations for institutional data management functions. Research libraries and library and information science (LIS) programs in particular are scrambling to respond to these new requirements and to understand emerging requirements for curricula and training for both students and working information professionals. Recent surveys of the field and major white papers provide evidence that there is an acute need for research that will inform this process of curriculum and training development; research that documents the emerging patterns in data management policies, and the expectations of major stakeholders in the research cycle regarding data management roles, responsibilities, and professional training and preparation for those taking on data management responsibilities.

There have been a number of recent ARL and other interinstitutional studies which highlight the need for systematic research efforts in institutional data management issues. [1] There are only a handful of universities that have implemented or begun serious planning for the implementation of comprehensive new data management functions at this point. [2] The emerging landscape of

research data management is both complex and very new to universities. [3] When taken together, the surveys and guides reviewed in the course of preparing for the DataRes project make a strong case that:

A. The research library community together with other research stakeholders is investing in new information profession positions to take on research data management responsibilities.

B. To be effective, information professionals assuming data management responsibilities must be informed by descriptive research concerning the perspectives and concerns of all stakeholders in the research cycle and trends in emerging institutional research data management policies.

C. This urgently needed descriptive research has not been conducted to date. Education and training efforts across the nation for the current and next generations of information professionals requires basic descriptive research in these areas.

2. Project Aims

Targeted descriptive research into stakeholder expectations and emerging data management policies is being undertaken in this project, which will inform the educational needs of the professionals who maintain and provide public access to data generated by the research community. This research project will include the following activities.

Creating a broad descriptive baseline study that documents emerging institutional research data management policies, and that empowers transformation of LIS curriculum and training: The implementation of data management plan requirements by federal agencies that provide most of the public research funding in the United States is a significant sea change in expectations about the preservation of research outputs. This change is rapidly leading to dramatic new thinking about the nature of information professionals that support the research cycle. There is no roadmap for this transition. Rather, we are seeing a great collective experiment that must be documented if we as a profession are to respond effectively. The information profession is justified in seeing this as an opportunity to take on new responsibilities, but this will only be possible if we understand the transition that is taking place.

Building a greater understanding of the skills needed for research data management functions through surveys of key stakeholder groups in the research cycle: Without a broadly based understanding of stakeholder perceptions and concerns, curricula and training will be uninformed of the actual requirements of work responsibilities in emerging data management programs. There is no more effective recipe for failure of service providers than not understanding the desires of their clientele. Unless the desires of all stakeholders in the research cycle are understood, data management duties will be misdirected.

Providing urgently needed research findings that will enable information science programs and academic libraries to work together to provide certification and training opportunities in research data management in order to better recruit and educate the next generation of information professionals: This is a remarkable moment for LIS educators and practitioners to come together to update educational offerings at a critical juncture. But this joint effort will be

stillborn if there is no guidance as to the direction that it must take. A research project documenting and informing this new direction is critically important to undertake now.

3. Research questions

Project personnel will investigate the following descriptive research questions:

Data Management Plan Policies and Practices: What trends are observable in the data management plans and associated institutional policies now being implemented at research institutions in response to federal requirements? What patterns will be demonstrated in a collection of institutional data management policies assembled as a key product of this project?

Stakeholder Expectations: What are the expectations of key stakeholders in the research community (including researchers, administrative officials, librarians, funding agency officials, research equipment vendors, and others) concerning the long-term management of research data generated in universities and the role of information professionals in such efforts?

Preparation Needed: What skills, infrastructure, training, and other preparation are needed for professionals charged with data management responsibilities, based on both expectations of stakeholders and observed trends from data management policies now being implemented?

4. Research stakeholder populations to be studied

There are a range of key stakeholders in the research community whose perspectives on the question of research data management are not well-known or documented. A central claim of this project is that until the perspectives of all of these key stakeholders concerning expectations, best practices, and required preparation are understood, progress cannot be made in establishing effective research data management functions by the majority of institutions. The following are descriptions of these key research community stakeholder groups, and rationale for including each of them in this study:

Researchers: An obvious and central group to assess, researchers produce the datasets and other forms of information through research that are the focus of the new federal data management plan requirements.

Provosts/CAO Officials: Provosts (or other Chief Academic Officers) of institutions receiving significant federal funds are another important group to study, inasmuch as these officials are charged with leading the academic affairs of their institutions.

Vice-Presidents for Research/Campus Research Officers: In the university hierarchy, the vice-president (or other senior leader) for research efforts is a key stakeholder, who has great influence in setting policies for research activities on the campus, especially grant applications and campus policies concerning such applications.

LIS Faculty: Faculty of Library and Information Science programs are seeking to respond in various ways to new data management plan requirements, for example through curriculum enhancements and new certification programs.

Librarians: Academic librarians are charged with collecting and managing the scholarly information assets of their institutions.

Campus IT Managers: As demonstrated by the 2010 ARL survey [4], in the relatively few instances in which a unit has been designated for data management to date, these units are often some variant of campus IT center, whether the center is organizationally part of the library or some other campus unit.

Funding Agency Officials: As federal agencies have implemented data management plan requirements, agency officials have had to make decisions about how to assess proposal statements concerning such plans.

Research Publishers: Publishers of research also have a vested interest in the long-term survivability of research data, yet their perspectives are infrequently taken into account in statements concerning data management plans.

Research Equipment Vendors: A group that has a significant insight into the needs for data management in research settings is that of research equipment vendors. In support operations for university research groups, such vendors are often able to observe first-hand the results of poor data management practices.

5. Project activities and products

Through this research project, educators and practitioners in the LIS profession will be provided with basic quantitative and qualitative information about stakeholder expectations, current policies, and needed preparation for information professionals taking on emerging responsibilities in data management. This information will establish a baseline for institutions to plan toward in developing new curricula and training. The following are some of the project activities and products that will assess the research questions through descriptive study:

Primary Survey: A survey of research community stakeholders will be undertaken to investigate research questions through multiple sub-questions. This survey will be conducted online, with phone follow-up to ensure acceptable response rates.

Focus Groups: Focus groups will be convened, timed in conjunction with professional meetings. The aim of these focus groups will be to engage groups of stakeholders in facilitated conversations concerning the research questions.

Interviews: Personal interviews will be scheduled with selected individuals, some of whom may be difficult to reach through other means, notably senior university administrators.

Data Management Policies Database: University data management policies will be collated and encoded to identify patterns in the implementation of data management plans and implications for training and education of information professionals.

Secondary Survey: a second survey will be undertaken to investigate unexpected questions that arise following the first survey, focus groups, and interviews.

Published Papers and Presentations: Individual papers on research findings will be published in professional publications or presented at professional conferences.

Symposium: An open symposium will be held to provide a forum for peer-reviewed papers and discussion concerning the future of research data management and implications for the LIS field.

White Paper: A comprehensive synthesis of all findings will be published by CLIR to report on overall conclusions of the project.

6. Acknowledgments

We wish to thank the Institute of Museum and Library Services for funding this research.

7. References

[1] Hswe, P. and Holt, A. 2010. Guide for Research Libraries: The NSF Data Sharing Policy. Association of Research Libraries. URL: <http://www.arl.org/rtl/eresearch/escien/nsf/index.shtml>

[2] Soehner, C., Steeves, C, and Ward, J. 2010. E-Science and Data Support Services: A Study of ARL Member Institutions. Association of Research Libraries: Washington, DC. URL: http://www.arl.org/bm~doc/escience_report2010.pdf

[3] Association of Research Libraries. 2007. Agenda for Developing E-Science in Research Libraries: ARL Joint Task Force on Library Support for E-Science Final Report & Recommendations. Association of Research Libraries: Washington, DC. URL: http://www.arl.org/bm~doc/ARL_EScience_final.pdf

[4] Soehner, *ibid.*