Analyzing the Data Management Environment in a Master's-level Institution

Anthony Stamatoplos *, Tina Neville [1], Deborah Henry [2]

Nelson Poynter Memorial Library, University of South Florida St. Petersburg, 140 Seventh Avenue South, St. Petersburg, FL 33701, USA

ARTICLE INFO

ABSTRACT

The data management environments at research-intensive institutions have been studied extensively. Few studies, however, have assessed the environments at institutions that are not classified as research-intensive, where scholarship and obtaining external funding is still highly encouraged. Using results from semi-structured interviews with faculty from an array of disciplines, the authors describe the research processes and data concerns at a Master's-level institution. A comparison of the results illustrate that, at least at this institution, faculty face very similar issues as those identified at research-intensive organizations and many of the same practices and services could be implemented on a smaller scale.

© 2015 Elsevier Inc. All rights reserved.

INTRODUCTION

Data are an essential part of research yet, for many years, they were not formally acknowledged as a part of the research process that should be managed, preserved, and shared. The National Institutes of Health (NIH) was one of the first federal agencies to make formal requirements for data management (U.S. National Institutes of Health, 2003). The release of the February 2013 Office of Science and Technology Policy (OSTP) memorandum on increasing access to federally funded research spurred additional activity as both funding agencies and researchers from around the United States reconsidered their data management practices to ensure compliance (U.S. White House. Office of Science and Technology Policy, 2013).

A great deal has been written about the effects of the federal data management mandates on academic researchers, participation of libraries and librarians, and the policies and processes that have been implemented at large research-intensive institutions to respond to the mandates. The focus of the current research study is to examine the environment, concerns, and challenges on a small, Master's-level campus where research and grant procurement is a prominent expectation of the faculty for promotion and tenure and for growth of the institution.

The university environment studied in the project is a separately accredited campus of a larger system of institutions, with approximately 200 faculty, more than 4100 undergraduates, and nearly 600 graduate students. The university is classified as Master's M by the Carnegie Classification system (Carnegie Commission on Higher Education, 2015). Undergraduate degrees are offered in 24 majors within the Colleges of Arts & Sciences, Business, and Education. Master's level degrees and graduate certificates are also offered in education, psychology, digital journalism, liberal arts, environmental science, and business administration. The faculty have variable teaching loads and graduate student assistance is limited. The campus strategic plan values and encourages faculty and student research and all faculty are required to conduct research for promotion and tenure. Not all faculty, however, need outside funding to conduct their research. Faculty conduct research in a wide variety of disciplines and participate in national and international collaborations.

The authors test the use of semi-structured interviews as a mechanism for gaining in-depth information about faculty research processes. From this knowledge, the authors seek to determine if the processes relating to data management that are already underway at research-intensive universities might translate effectively to the research practiced at Master's-level institutions.

LITERATURE REVIEW

Numerous articles and reports detail the assessments and studies conducted by librarians at large research institutions on the data management environment and data practices on their campuses (Antell, Bales Foote, Turner, & Shults, 2014; Delserone, 2008; McLure, Level, Cranston, Oehlerts, & Culbertson, 2014; Peters & Dryden, 2011; Rolando, Doty, Hagenmaier, Valk, & Parham, 2013; Shen & Varvel, 2013). In 2011, Tenopir, Birch, and Allard (2012) conducted an extensive survey of Association of College and Research Libraries (ACRL) member

* Corresponding author. Tel.: +1 727 873 4124 (voice); fax: +1 727 873 4196.
E-mail addresses: stamatoplos@mail.usf.edu (A. Stamatoplos), neville@mail.usf.edu (T. Neville), henry@mail.usf.edu (D. Henry).
[1] Tel.: +1 727 873 4091 (voice); fax: +1 727 873 4196.
[2] Tel.: +1 727 873 4977 (voice); fax: +1 727 873 4196.

libraries including institutions awarding Associate, Baccalaureate, and Doctoral degrees, to establish a baseline of the current data services offered and those planned for the future. Though Doctoral institutions were most likely to provide data services, the authors recommended that non-Doctoral institutions should consider offering assistance in these areas. To meet the staffing demands of new services, many academic libraries were reassigning existing staff and also partnering with other offices on campus.

Additional surveys discovered disciplinary differences in data management practices. A 2012 survey of Emory University faculty found that basic scientists were the most aware of the federal data mandates and were more likely to deposit data into repositories than other disciplines. Social scientists and medical researchers expressed concern over data sharing, related at least in part, to the confidential nature of much of their research, while faculty in the humanities were the most willing to share data with the public. The authors discussed the role of data in the humanities, noting that these faculty may not define their research results or artifacts as data (Akers & Doty, 2013). An earlier survey of more than 1300 international researchers found that social scientists were less likely to share data or have a specific practice for data management. The majority of social science researchers, however, agreed that the lack of access to others' data is an obstacle to scientific progress (Tenopir et al., 2011).

A number of studies examined the role of the library in the institution's data management environment and the advantages of the library reaching out to faculty. In a 2012 survey of Association of Research Libraries (ARL) science librarians, respondents perceived a broad range of roles for librarians, from traditional metadata organization and research instruction to new skills specific to data management and grant-writing (Antell et al., 2014). After interviewing STEM researchers, University of Houston librarians proposed training library liaisons on data management practices, providing them with the skills needed to discuss data management practices with faculty in any discipline (Peters & Dryden, 2011). A Purdue subject librarian worked closely with an agriculture professor to create a prototype repository for data collection, including the development of appropriate metadata schemas and copyright protection. In-depth conversations with researchers across the institution highlighted new roles that librarians could play, and provided a better understanding of the bigger research picture on campus, suggesting partnerships where the faculty themselves may not have been aware of potential collaborators (Bracke, 2011).

While several investigations explored the situations at institutions where research is not necessarily described as "big science/big data," there are only a few reports directly related to non-research-intensive institutions where research is also an important part of the culture. Scaramozzino, Ramírez, and McGaughey (2012) surveyed primarily teaching-focused faculty who are expected to incorporate the results of their individual research into their teaching. While their institution is not small in student population (~19,000), the authors described the scholarship conducted on their campus as small science. Their survey found that half of the faculty respondents lacked confidence in their data management skills and were open to additional education in that area, but most did not consider the library when looking for assistance with data management education or storage of their data. Cox and Pinfield (2014, p. 308) studied the research data management activities of a cross-section of academic and research organizations in the United Kingdom including, "teaching-led" organizations, but they reported their results in aggregate and so did not present a clear picture of the environment at those non-research-intensive institutions. They noted "distinct differences" in the number of services provided by the group described as large research-intensive organizations compared to all the other groups participating in their study.

The enormous amount of small data being produced is a topic that was raised by Salo (2010), who discussed the issues facing libraries that were taking up the "data challenge." Salo contended that widespread and disparate research practices make increased communication necessary to produce an "acceptable standard" of data management. Shorish (2012) added to the small data conversation with a discussion of the importance of Master's and Baccalaureate institutions' participation in data curation. These types of academic institutions far outnumber research-intensive ones according to Carnegie Foundation data, and many produce quality research data that require the same attention as big data sets. While describing some of the first steps taken at the author's institution to raise awareness of data management requirements, Shorish suggested that an assessment of faculty needs to be conducted. Goldstein and Oelker (2011) presented a game plan for data curation at liberal arts colleges where significant research is performed. The authors employed an informal survey of faculty in the natural and physical sciences to gather information about their institution. They found that conversations with a nearby research university was a useful learning experience that gave them confidence in their own newly created data management decisions. Toups and Hughes (2013) published the process and outcomes of their small liberal arts university's foray into data curation. To assess the data needs of researchers at Trinity University, the librarians conducted focus groups composed of 10 faculty from diverse disciplines. In addition to learning more about their researchers' data processes, the conversations also revealed that their faculty did not think intuitively of the library as a partner in data management.

Many of the studies cited here offer benchmark data or make a case for the importance of assessing individual institutional environments. Each presents unique considerations in addition to the common themes and challenges that many authors have documented. There is limited literature relating to the needs of colleges and universities that are not designated as research-intensive or ARL institutions. Only a few studies examine the data management environment at smaller or non-research-intensive institutions. Issues examined in prior studies are quite varied and center on areas outside the main goals of this project: 1) to obtain in-depth information about faculty research processes in a variety of disciplines at a non-research-intensive institution; 2) to see if the data management needs of these faculty are similar to those at research-intensive institutions, and; 3) to test the use of semi-structured interviews to gain this knowledge.

METHODS

This investigation seeks an understanding of the thinking and practices of a small, but diverse population of faculty researchers regarding data management. The concept of data management itself is ambiguous, with multiple and varying meanings and perspectives. As noted in the literature review, disciplinary orientations, methods of research and analysis, and individual experience may affect how researchers see and talk about data management. Additionally, in the context of plans and mandates, some issues may be unfamiliar or new to many researchers.

The authors searched the system's database of grant activity to identify local faculty who had either applied for or received an external grant since 2009. As the focus in this exploratory study is related to data mandates, the authors identified researchers likely to be most directly and immediately affected by current or future data requirements. The search identified 36 faculty who met these criteria, thus forming the population for this study. Collaboration may mask the real number of funded research projects occurring at this institution. The system-wide database used to identify the initial population reported names of researchers who are principal investigators. There could be others involved in externally-funded projects in supporting roles but who would not appear in a search. A much broader survey of all faculty would be needed to verify that all researchers with external funding have been identified.

Various techniques for collecting relevant data from faculty members were examined. Several earlier studies employed surveys and focus groups to gain a broad sense of the research environments and