Understanding Methodological and Disciplinary Differences in the Data Practices of Academic Researchers

Travis Weller and Amalia Monroe-Gulick

"This article is (c) Emerald Group Publishing and permission has been granted for this version to appear here. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited."

Introduction

In response to the rise of data driven research, academic libraries have expanded their research data services. Rather than adopt a blanket, “one-size fits” all model, these research data services should be provided with a detailed and nuanced understanding of their users. In this article, the authors examine how research data practices and future research needs vary by research methodology and academic discipline based on results of a locally conducted survey. With this information, academic libraries can offer targeted services to different user populations, which would be more efficient for the libraries and more useful to the researchers.

Background

The prevalence of electronic data has changed the process of research. While a researcher’s underlying research questions may remain consistent, the materials available to investigate those questions, the tools available to analyze materials, and the methods for disseminating results of that analysis have all shifted in recent years.

As part of the campus research infrastructure, academic libraries have been evaluating their role in this era of data driven research (Hey and Hey, 2006; Jones, 2008; Luce, 2008; Soehner, et. al., 2010). A 2010 survey of Association of Research Libraries (ARL) member libraries found that 73% were involved in supporting data driven science, in some way, at their institutions (Soehner et. al., 2010). However, the types of data services offered varied by institution from helping to locate relevant data to data curation and preservation. A report released two years later, in 2012, found that only half of libraries at research institutions offered reference support for data, and approximately a quarter of libraries at these institutions offered data management consulting, data standard consulting and more comprehensive research data services (Tenopir et. al., 2012). The same study found though that libraries planned to dramatically increase the services they offer in the near future. A recent study of four midwestern land-grant universities showed significant variation in the data management services offered by four schools that are incredibly similar in terms of the student body, research profile, geography and institutional mission (Deards, 2013).

The role of individual librarians is similarly unresolved. Most libraries reassign existing staff to provide assistance with research data (Soehner et. al., 2012). Despite this, relatively few academic librarians directly interact with researchers on data topics such as the data lifecycle, data management planning, data analysis, and preservation (Bresnahan and Johnson, 2013). A recent survey targeted specifically to academic librarians most likely to provide data support services found that less than one-third of those librarians provide research data services as an “integral” part of their job responsibilities. (Tenopir et. al., 2013). At the same time, there seems to be an emerging number of academic librarians who serve as research data managers (Cox and Corrall, 2013).

To help guide the services they offer their researchers, a wave of institutions have conducted campus-wide studies to better understand the research needs on their own campus (Marcus et. al., 2007 (Univ. of Minnesota); Peters and Dryden, 2011 (Univ. of Houston); Akers and Doty, 2012 (Emory
Inspired by efforts at other institutions, and in order to better understand the needs of their own researchers, the University of Kansas (KU) Libraries decided to conduct a similar local study of the campus. KU is a major, public research institution. With over $286 million dollars in research expenditures in fiscal year 2012, it ranked 75th among all institutions in research dollars. It is a doctoral granting institution with 5,691 graduate students and 1,608 faculty members. The University of Kansas campus hosts a wide range of disciplines, from engineering to humanities, however the medical school is separate and was not included in the study. The results of the survey at KU will help guide service provision at this campus, but also contribute to the broader understanding of research data practices and needs.

**Literature Review**

Significant progress has been made in recent years in understanding the data practices of academic researchers. At the broad, disciplinary level, the United Kingdom (UK) based Research Information Network (RIN) produced studies that examined how life science and humanities researchers interact with their data (RIN, 2009; RIN, 2011).

Life science researchers in particular worked with large amounts of data, some of which was standardized and machine generated, while others worked with more diverse data (RIN, 2009). These researchers felt highly protective of their data but were willing to freely share the software, tools and codes they used to analyze the data. Life science researchers saw a need for better standardization of data, but were also concerned about managing and having the tools to analyze the large amounts of data they generate.

Humanities researchers engaged with a wide range of resources, from paper materials and microfilm to advanced digital resources (RIN, 2011). At the time of the RIN study, few humanists were engaged in highly collaborative, highly computationally demanding research. Although their work was highly complex and varied significantly from project to project, humanists were not as likely to use the state-of-the-art available technology. They were concerned about linking data stored in various archives so that it was available for their own research. There was also a need to help humanists become aware of, adopt and use new tools and digital resources for their research.

For social scientists, the Council on Library and Information Resources issued a report on the data practices of social scientists in 2012 based on interviews they conducted with faculty, postdoctoral fellows, graduate students and other researchers representing multiple universities in the United States (Jahnke and Asher, 2012). Social scientists expressed little interest in modifying their data for long-term storage and preservation; they focused instead on the usefulness of the data in preparation for publication. Social scientists constantly adjusted their research practices, from collection to analysis, to meet their immediate needs.

These broad, disciplinary focused reports are helpful but there is also a need for a more individualized understanding of researcher data practices. Purdue University and University of Illinois faculty developed the *Data Curation Profile* to help librarians construct detailed understandings of a single researcher’s data needs (Witt et. al., 2009). The resulting compilation of the completed data profiles, launched in 2009, provides in-depth, qualitative information about the data practices of individual researchers as varied as demographers to physicists.

Librarians have also completed a number of institution-specific studies to better understand the research data needs of graduate students and faculty members on their own campuses.

In a 2007 report, librarians at the University of Minnesota described the results of interviews and focus groups they conducted with faculty and graduate students to better understand their campus’s research practices and needs (Marcus et. al., 2007). Their research data curation was
“idiosyncratic, haphazard and in great need of attention.” (Marcus et. al., 2007 at 10). Researchers organized data in a practical, utilitarian manner. It needed to be organized in a way it could found easily and so they could do something with it.

A survey of faculty at Georgia Tech, completed in 2010, found that a majority of faculty members at their campus were interested in data storage and preservation, data sharing tools, and information on best practices for data management (Wells Parham et. al., 2012). Fewer than half of respondents wanted assistance in drafting data management plans or meeting the data management requirements imposed by funding agencies.

Librarians at the University of Houston focused their interviews, conducted in 2011, on science faculty that had received external funding for their research (Peters and Dryden, 2011). These principal investigators (PIs) saw project data management as their responsibility, but the study concluded that there was often confusion among researchers about what constituted data management.

In 2012, the University of Iowa conducted a survey of faculty and staff to understand the research data management needs of their campus’s researchers (Gu and Averkamp, 2012). Respondents primarily stored their data on their personal computer or on a shared university server and the vast majority felt their data was confidential or sensitive. Although over one-third of the Iowa researchers did not, at the time, receive assistance with managing their data, most of those were interested in data management assistance in the future.

While some of these studies focused on particular user groups others were conducted for the entire institution. Nevertheless, none of the studies discussed above reported comparative differences among user groups on their campus. Comparative, quantitative, information about researchers in different disciplines is still relatively rare. Studies at Oxford (Wilson, 2013) University of Nottingham (Parsons, et. al., 2013), and Emory University (Akers and Doty, 2013) are a few notable examples of studies that reported differences in researcher practices by discipline in a quantitative, comparative way.

In particular, the recent survey and analysis by Emory University librarians Akers and Doty demands attention for its investigation of how faculty members’ research practices varied by academic discipline (2013). They found that basic science researchers at their campus stored higher amounts of data, were most familiar with funding agency requirements for data management, and most likely to share their data with colleagues outside of their research group. Emory social scientists were unlikely to share their data or archive it. Akers and Doty speculated that this was due to concerns over the private or sensitive nature of their data. Arts and humanities researchers in their study were unlikely to use central data storage options, like shared university servers, were relatively unfamiliar with funding agency requirements for data management, and least likely to share their data with others. Compared to others, humanists at Emory were most interested in digitizing the physical research materials that they already have.

This paper adds to the literature by reporting on data practices, influences and future data-related challenges of academic researchers at a major public research institution. In addition to analyzing the results by academic discipline, the survey responses are also examined for differences between researchers who employ different research methodologies.

Data practices based on research methodologies are important for librarians and data archivists to understand because the methods determine how the researchers interact with their data. A particular method of research cuts across traditional departmental, disciplinary and institutional lines. Someone who does archival research is more likely to be similar to someone else who does archival research in terms of how they work with their data even if one archival researcher studies sociology and the other studies geography. The theoretical foundation and the subject of study may vary widely, but the data practices themselves are likely to be similar. To date, this is the first study to report on differences in data practices among researchers who use different research methodologies.
Methodology

During the fall semester of 2013, a group of librarians and a campus grant-specialist, conducted a needs assessment of faculty, researchers, and graduate students at the University of Kansas. A survey instrument was selected as the tool for conducting the assessment. The four objectives of the survey were to:

1. Identify the current and future research needs of KU faculty, researchers, and graduate students.
2. Assist the Libraries with the development of an implementable plan for services that address the current and future research support needs of KU faculty, researchers, and graduate students.
3. Assist with the development of new services, programs and supports for researchers, as appropriate, based on the results of the survey and additional assessment activities.
4. Establish benchmarks based on initial data collection that will provide the baseline for monitoring future services and programs.

One component of the survey was research data practices and needs of KU researchers. The survey was pre-tested among representatives from all groups in the identified population. The survey was then distributed via two email distribution links that would reach the largest number of potential respondents. The final response rate was 7% overall, with 14% of faculty and 5% of graduate students responding. There was overrepresentation of faculty. However, when the results were analyzed by broad disciplinary area, the results more closely aligned with the make-up of the university (see Table 1). The authors intend for the data to be posted publicly through KU ScholarWorks within a year after internal decisions are made.

Table 1: KU Population Compared to Survey Respondents by Disciplinary Area

<table>
<thead>
<tr>
<th>KU Disciplines</th>
<th>KU Overall</th>
<th>Survey Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humanities</td>
<td>27%</td>
<td>26%</td>
</tr>
<tr>
<td>Social Sciences</td>
<td>51%</td>
<td>51%</td>
</tr>
<tr>
<td>Sciences</td>
<td>22%</td>
<td>23%</td>
</tr>
</tbody>
</table>

Limitations:

As with any study, the KU Libraries faculty/staff and graduate student research needs survey has limitations. First, the response rate is not high enough to test for significance, and may decrease the generalizability of the results. In addition, since the survey was distributed to the entire population, the responses may be skewed towards those researchers who have stronger opinions about the library. With the inclusion of graduate students in the survey population, results may not be entirely reflective of the local environment because of potential differences in research at other locations.

---

1 The entire survey instrument is available in the KU intuitional repository: KU Scholarworks
2 Only 76% of respondents provided an answer for this question. Out of 415 responses, 402 provided a department or center that could be classified into a broad disciplinary area.
Results

Multiple Methodologies

Respondents were asked to identify which methodologies they used in their research from a list of 21 options, including an “other” open-ended response. The authors developed the list of methodologies through past experience working with researchers and it was refined during the pre-test. Rather than force a respondent to choose a principal research methodology, they were allowed to select as many methodologies as applied (see Chart 1).

On average, each researcher responded that they used five different research methodologies. Less than 10% responded that they only used a single methodology. An equal percentage responded that they used *nine or more* of the different methodologies listed. The number of methodologies selected by individual respondents is displayed in Chart 2.
Some of the overlap between different methodologies was anticipated. For example, of the 46 oral historians who responded, 61% also responded that they use qualitative methods, which is understandable given the qualitative nature of oral histories. Similarly, 85% of the statisticians also responded that they used quantitative methods.

However, individual researchers also reported using methodologies that normally would not be associated at all. Of the 94 respondents who do data mining, 24% also do field work. Of the 42 respondents who work with geospatial data, 43% are also historians. In fact, the two broadest, and generally opposing, categories – quantitative and qualitative – also had some of the greatest overlap. 55% of quantitative researchers also responded that they were qualitative researchers and 63% of qualitative researchers responded that they were quantitative researchers. For people who selected "qualitative" as one of their methods, the most common other method selected is "quantitative." Similarly, "qualitative" is the second most commonly selected method by those who also marked "quantitative."

So, it is common for an individual researcher to use both qualitative and quantitative methods to understand their data. Even though qualitative and quantitative data, tools and analysis — and the skills needed to manage them — vary greatly, a single researcher likely uses both. This methodological overlap also influences the results of this survey. If a respondent answered that they used historical methods and quantitative methods, then their responses about how they manage their data is included in both the discussion of historical methodologies and quantitative methodologies.

Data Storage Practices

To develop a more detailed understanding of how researchers store their digital files, respondents were asked to identify the storage method for different file types including text documents, image files, sound or video recordings, downloaded online material, and data sets. Regardless of the file type, the most common method of digital file storage is saving it to a hard drive or a CD.iii Cloud-based commercial storage like Dropbox or Google Drive was the next most common storage method for text (16%), images (13%), recordings (12%) and downloaded web content (16%). For data sets though, a central University server (16%) was more common than cloud-based storage (13%). For text files, printing and saving the physical versions of the files is still utilized relatively regularly (14%).

---

iii Text files, 58% of respondents; image files, 71%; sound or video recordings; 73%, web content, 67%; data sets, 66%. 
The five most common research methods among survey respondents, from the 21 methods offered, were quantitative, qualitative, statistical, experimental, and historical. These methods represent a vast array of different data types. As a result, the authors hypothesized that data storage practices and influences would vary significantly between these disciplines as well. See the chart below for detail on how the data storage for text files varied by research methodology.

Researchers who do historical analysis are more likely to print and save physical copies of text. Historians are also the least likely to utilize data sets. More than half of historian respondents reported that they did not use data sets.

Similarly, there was little variation in data storage when the results were analyzed by academic discipline, although social scientists and humanists were more likely than natural scientists to use cloud-based commercial storage.
Influences on Data Storage

To better understand why researchers decided to store data in certain places, the authors asked respondents to identify factors that influence their data storage methods. They were allowed to select as many factors as applied. While the data storage practices are relatively consistent across researchers of different methodologies, the reasons they cite for their data storage practices varies by methodology.

Ease of the storage method (88%) and long-term sustainability (62%) are more influential for historians than researchers using other methodologies, however historians are less likely to be influenced by grant requirements (8%). Privacy and security concerns are motivators for statistical (51%), quantitative (50%), experimental (49%) and qualitative researchers (47%), but are less influential on historians (34%). Concerns over file size and back up needs are more pressing for quantitative and statistical researchers, while cost concerns seem to motivate quantitative and experimental researchers at higher levels than the average researcher.

Regardless of the research methodology, ease of storage is the primary influencing factor. Other interesting trends emerge, when the responses to this question are examined by researcher’s discipline.
Humanists were less likely than the average researcher to be influenced by file size or concerns over cost or privacy and security. Social scientists were more likely than others to be influenced by physical space requirements and privacy and security concerns. Natural scientists were more influenced by back up needs, file size, and cost than their counterparts, but found physical space requirements much less influential.

Although in general, grant requirements were relatively not important compared to other concerns, grant requirements are more influential than the norm for natural scientists (19%) and social scientists (15%) and least influential for humanities researchers (4%). Akers and Doty found that basic scientists were most likely to be familiar with funding agency requirements, followed by social scientists, with humanists least likely to be familiar with them (2013). So, the variation in the influence of funding agency requirements may be based on their familiarity with them or may reflect the availability of external research funding in the humanities.

Data Related Research Challenges

To better understand barriers to research, respondents were asked to identify which stage they found most challenging or time consuming. The authors divided the research process into stages – identification of data, acquiring access to the identified materials, managing data once they have been acquired, and disseminating the research results and respondents were instructed to select one option – the piece of the research data process they found most challenging. Chart 6 displays stage selected as most challenging by research method, compared to the overall results.
Among all respondents, results were relatively consistent across the first three phases of the research process – identification, acquiring and managing materials. However, significant differences emerged once the sample was divided by research methodologies. Quantitative and experimental researchers were less likely to struggle with acquiring access to data (21% for quantitative, 22% for experimentalists compared to 26% overall). In contrast, historians find acquiring access to materials the most challenging, but research dissemination not as significant a barrier.

Additional interesting results emerge by examining the results by disciplines. Social scientists
were most likely to struggle with identifying relevant materials/data. Of the three disciplines, humanists were far less likely to struggle with research dissemination, but more likely than others to find managing materials challenging. For natural scientists, acquiring access to materials (30%) was more challenging than those in other disciplines (28% humanists, 19% social scientists), or the total population (26%).

Data Related Future Needs

Finally, in an intentional effort to gain information about how the library can best position itself to support researchers moving forward, researchers were asked to anticipate their future needs related to their research data. Respondents were allowed to select multiple options.

Overall, researchers primarily anticipate needing assistance with data analysis (51%), data storage (50%) and data dissemination (52%). Although it was selected least often, more than one-quarter of respondents expect that they will need assistance with drafting data management plans in the future. These responses demonstrate that researchers acknowledge they are going to need help with a wide range of data related tasks in coming years.

Significant differences emerge though when results are analyzed by different research methods. Chart 8 details the future needs of researchers by different methods, and compares their responses to the total.

Chart 9: Future Needs (by methodology)

Assistance with data analysis for quantitative (65%), statistical (64%) and experimental (64%) researchers is a much more significant concern than the total population (51%).

Historians anticipated needing less assistance writing data management plans (21%) and conducting analysis (34%) than the population as a whole (28% and 51%) respectively, but anticipate a much greater need with digitization (53% compared to 37% of the total).
An unexpected number of social scientists – 72% - reported need assistance with data analysis. Only 21% of humanists anticipated needing assistance with data analysis. Instead, humanists are more likely to need assistance with disseminating and publishing their findings, storing and preserving their data and digitizing resources. It is curious that only 8% of humanists named “disseminating research” as the most challenging aspect of their research but 58% of them acknowledged a future need for assistance with data dissemination and publication. This may reflect an increasing pressure on humanists to connect their work to others given concerns with standardization, quality, reliability and functionality of different resources (RIN, 2011).

Discussion/Conclusions

These survey results have significant implications for KU and, potentially, other research universities and academic libraries.

The overlapping use of different research methodologies by single researchers forces the reconsideration of the typical view of academic researchers as a specialist in a specific type of research method. Instead, researchers are approaching their primary subject of study using a range of research methods. This diversity of methodologies can be extraordinarily challenging for libraries and archivists. Data management strategies for a single project may have to include a host of different software packages and file types. Skill development in analytical tools has to be offered broadly; more than just the geographers are working with geospatial data.

This challenge is further compounded when multiple researchers, each employing multiple research methods, collaborate on a single research project. A library liaison assigned to support a specific department or discipline within the institution will need to be able to provide support to a wide range of researchers employing a host of methods. Alternatively, libraries could develop and train individual experts in a specific methodology and then make that person available to support researchers across campus without regard to department or discipline. For example, academic libraries could have a single librarian who is an expert in survey research and that person supports all the researchers who conduct surveys on campus. This is an area that would benefit from further investigation to determine
the more specific needs of those in different disciplinary areas but who utilize similar methodologies from those in other disciplinary areas.

Given researchers’ continued reliance on hard drives and CDs for data storage, and that researchers cite “ease of storage” as the most common influencing factor in their data management decisions, institutions must make data storage easy. Institutions offer their university servers as the secure alternative to hard drives and commercial services. If institutions want to encourage the use of central research data storage, their primary concern should be to make it as simple to use as saving materials to a hard drive. Connaway and Dickey recommend making data storage responsive to disciplinary differences but still simple enough that it integrates into a researcher’s everyday workflow (2009). Beyond ease of storage, universities could persuade researchers to use institutional repositories by emphasizing different features of the repositories. This is an important area for academic libraries, since the library could be in charge of at least one of the institutional repositories, as is the case with KU. If this approach is taken to assist researchers with data management and preservation, the library will need to decided what it can and cannot do based on their existing infrastructure and potential growth in the future to support growing needs.

Due to the lack of influence of grant requirements on data storage decisions, citing the storage requirements of external funders is not likely to encourage the use of centralized data storage. The lack of influence of grant requirements may be due to researchers’ lack of familiarity with these requirements (Akers and Doty, 2013). So, institutions could try raising awareness of grant requirements, and then leveraging that new knowledge to increase use of university servers. However, that strategy is unlikely to prove successful in fields where grant funding is less available.

Additionally, these results can help libraries tailor the services they provide as they go through the process of expanding their research data services.

Academic libraries have devoted a significant amount of attention in recent years to the final stage in the research process (the advent of institutional repositories and participation in the open access movement are two notable examples of this). These efforts have produced significant positive changes in research dissemination in recent years. However, these results indicate that researchers are more likely to find the earlier steps of the research process most challenging – identifying, acquiring and managing their research data, with some variation by research method and academic discipline. The 2012 ACRL survey indicates that libraries at research institutions plan to offer services in each of these areas, though the nature of these services will inevitably vary from institution to institution (Tenopir, et. al., 2012). To help advance the research missions of their institutions, academic libraries will need to ensure that they offer services, or can connect their campus’s researchers to appropriate services elsewhere, for all stages of their research.

Given that institutions are organized by academic department and discipline, it may be helpful for libraries to use the disciplinary results of this survey to target their outreach efforts. Then, the methodological results can be used by libraries to ensure that they have staff with sufficient expertise to help researchers using various methodologies with method-specific issues.

Digitization services are most in demand by historians and humanists. As a result, libraries could target their outreach for the digitization services they provide to these communities by visiting department meetings or holding demonstration sessions in collaboration with a humanities research center.

Assistance with data analysis is pressing for quantitative, qualitative statistical and experimental researchers as well as social scientists. Libraries could consider expanding the data analysis services that they offer, or identify other places where assistance with data analysis is available, such as campus research centers or even off-campus sources.

That natural scientists acknowledge that data storage and preservation is an area of future need is a significant opportunity for libraries, since libraries often manage their institutional repositories.
However, libraries may have to develop expertise in the types of data generated by natural scientists, such as machine generated equipment outputs and ensure that repositories are capable of storing the large datasets produced by natural scientists.

Compared to others, researchers do not anticipate a pressing need for assistance with writing data management plans. Nevertheless, over 40% of institutions surveyed by ACRL either already offer or plan to offer consultation services on data management plans (Tenopir et. al., 2012). This is not to imply that libraries should not offer assistance with data management plans. Instead, data management planning should be just one facet of a library’s suite of research data services.

The results of the survey conducted at the University of Kansas of faculty, staff and graduate student research practices resulted in some unanticipated results related to data services needs that have implications for academic libraries. In addition, the results also indicate that further investigation into methodological approaches, and not only disciplinary areas, will assist libraries with designing the most useful and adaptable data services for researchers at all levels.

Acknowledgements

The authors would like to acknowledge the contributions of Brian Rosenblum and Julie Waters for their contributions to the development and dissemination of the survey instrument.

Endnotes


References


**About the authors**

Travis Weller is a Grant Specialist at the Institute for Policy & Social Research located at the University of Kansas. He holds a BA in Sociology from the University of Kansas, a JD from the University of Wisconsin Law School, and an MLS from the University of Illinois at Urbana-Champaign. Travis Weller is the corresponding author and can be contacted at: weller@ku.edu

Amalia Monroe-Gulick is a Strategy and Assessment Librarian at the University of Kansas Libraries. She holds a BS and MS in Political Science from the Illinois State University, and an MLS from the Indiana University.