Supporting the research process through expanded library data services

Supporting the Research Process through Expanded Library Data Services[1]

(This is a post-print article published in
Program: electronic library and information systems, 47(3), 282-303)

Minglu Wang, Data Services Librarian, John Cotton Dana Library, Rutgers-Newark

Acknowledgements

This paper is developed with contributions from my colleague, Ka-Neng Au, Head of Media & Digital Library Services, John Cotton Dana Library of Rutgers, and it is based on a conference paper, "Data needs and data services: a case study," co-presented with Ka-Neng Au in the 2011 Asia-Pacific Conference on Library & Information Education and Practice: Issues, Challenges and Opportunities, Putrajaya, Malaysia.

Abstract:

Purpose

This paper describes the process of how we gained a better understanding of the variety of library users' data needs, and how we gradually established some new data services based on our current capabilities.

Design/methodology/approach

This paper uses a case study of the new data services at the John Cotton Dana Library, at Rutgers, The State University of New Jersey, Newark campus, to demonstrate the possible ways to extend data reference services and provide data computing services. A content analysis of our services records shows how each user group falls into our multiple data services levels and subcategories.

Findings

Library users can be classified into many different categories, and each of these may have different needs. Research centers might have big projects involving data gathering and applications where we can mainly provide consultation, while an individual faculty member or student might need the librarians as research partners, with help for their specific problems. Computing data services can involve group training and statistical analysis assistance, where researchers need emergent help. Data librarians can take various opportunities for data management education, thereby gradually raising awareness and cultivating better research habits among researchers.

Originality/value

Library data computing services can make unique contributions to faculty and students' research and study. Institution, library and users' interaction determines the levels and extent of data services is generalized from the description and analysis of typical data service examples. Classic concept of data services levels is applied to a concrete case of data services program, and sub-categories of each data services level and user types are developed based on our services record.

Keywords: Data services; Data reference services; Data computing services; Data management; Academic Libraries; US

INTRODUCTION

When thinking about the future of academic libraries and the role of librarianship, one major expectation is that the library will be involved in deeper levels of academic research (Lewis, 2007). For data services specifically, traditional library reference services have focused mainly on data access and discovery and they have done quite well in pointing users to the right data resources from within the enormous and continually changing information pool. However, research in a greater number of disciplines now involves data collection in the laboratory or field. New forms of research collaboration, publishing, and applications have developed, not just within academia, but also in all types of institutions, with the aim of facilitating evidence-based decisions. How then can a library become involved one step deeper in the research process? At present, opportunities and challenges seem to coexist. Some have even argued that data librarianship will be a model for the future library services, with a greater role in pre-publication scholarship (Jacobs & Humphrey, 2004; Gold 2007).

In this article, we will describe and analyze our own experience of establishing library data services on our campus, to shed light on some of the possible ways that academic libraries could become more involved in the teaching and researching process. Unlike most existing practices of either data reference services or data curation services within academic libraries, our current data services are provided mainly at the level of assisting researchers' data computing processes; for example, data analysis and presentation, which is a middle part of the research process, after researchers have gathered the data and before publication of their results.

The Dana Library serves the Newark campus of Rutgers, The State University of New Jersey, and is part of the statewide Rutgers University Libraries. Before the new data services librarian was hired, data resources acquisition and the research data repository portal were centralized and operated on the main New Brunswick campus for the whole library system. The reference librarians at our local campus library were very competent at providing data reference services. These services continue to include reference assistance to identify and locate subject-specific statistics for students, staff, and faculty; information literacy classes and instructional workshops for students and local entrepreneurs who wish to learn the use of statistical databases; and online resource guides for both undergraduate and graduate courses.

However, when the active researchers and centers on the Newark campus began to realize and articulate their new data-related needs, these needs clearly went beyond just data source identification or acquisition. When our library was approached for possible support, the response was to submit a proposal to a University-wide challenge grant program. The proposal was successful and the grant enabled the library to hire a data services librarian to work under the direction of an experienced reference librarian and to start the process of venturing into the new service of conducting institutional research and assisting faculty and research centers with data management, data analysis, and presentation.

We started from the point where we only knew that data services were going to help with data analysis and presentation – the computing portion of data services – but we were not exactly clear about the scope of the needs and the concrete forms of delivering the service. It took about two years of continuous interaction between the librarians and the different user groups on campus to develop an understanding of the general data computing needs that had been articulated, and to know what we were able to do and what services we were not able to provide. Here, through the analysis of our own experience during the process, we provide an additional concrete example of how to implement data computing services. We share our thoughts and unique discoveries for other libraries that have or are about to start data computing services, so that they can better understand and design their local services.

LITERATURE REVIEW

Geraci, Humphrey, and Jacobs have been team-training new social science data services librarians through the Interuniversity Consortium for Political and Social Research (ICPSR)'s Summer Institute Program, which focuses on how to design a social science data services program. They developed a very useful concept and classification of data services levels to describe the possible scope of data services (Geraci *et al.*, 2010). These levels of data services were adopted by later researchers when they introduced data services to general social science reference librarians (Kellam and Peter, 2011). The concept has also been extended into more tiers, to incorporate new types of data management, curation, and preservation services by other researchers, when recommending a more comprehensive and forward-looking data services program for an academic library (Bennett, 2010).

Geraci, Humphrey, and Jacobs categorized data services into three major levels. *Data reference services* can be understood as an analogy of general library reference services that include data identification and retrieval assistance and instruction. *Data collection services* are usually conducted by subject librarians as a natural part or an extension of their subject expertise; these services typically include functions such as selecting and acquiring data sets and building local data collections. *Data computing services* are usually defined as helping researchers with their actual usage of the data after those data have been identified and obtained; this requires that the librarian is able to understand the statistical analysis of the data and is familiar with the technical tools that are appropriate for the analysis. *Data curation services* are the highest level of all of these data services, and according to Bennett's data services tiers model (which is not included in Geraci's original model), this is a completely new institutional role for libraries (Bennett, 2010). Whereas libraries once collected information both in print and digital formats, they now perceive their role as naturally taking on the additional responsibility of helping researchers with the storage, management, and sharing of their research data.

The development of data reference and data collection programs and services within academic libraries has been a continuous topic since the 1980s, especially in the area of social sciences reference services (Adams, 1982; Rowe, 1982). When research data first started to be available in digital formats in significant amounts, librarians immediately became interested in learning the new knowledge and skills necessary for collecting and cataloguing machine-readable datasets, in order to help users identify and access data. Since then, multiple data library models have been developed and recognized. In 2001, the Association of Research Libraries (ARL) published its SPEC Kit on Numeric Data Products and Services (Cook *et al.*, 2001). According to a survey conducted among ARL libraries, the majority of libraries were providing data collection and cataloguing services in collaboration with other campus units, especially particular academic departments that were also collecting and maintaining specialized data sets for their own departmental use. Recent studies in this field (Bennett & Nicholson 2004; Duke *et al.* 2007; Mooney & Silver 2010; Hogenboom *et al.* 2011; Kellam & Peter 2011; Read 2007) have provided more detailed evidence of library users' real needs regarding data-related assistance in light of the increasing availability of digital research data. These studies suggested that a library has a unique capacity to break up departmental silos of data collection, to aggregate resources, and to set up convenient data clearinghouses. Multiple ways of promoting this new library service through outreach, marketing, and collaboration with other institutional departments are also recommended.

Discussions about data curation services, the highest level of data services, have blossomed in the last few years. If data reference and collection services are closely aligned with traditional library services, data curation support is then more closely tied to innovative digital library services. Data management support in that case becomes revolutionary as it requires, ideally, that librarians participate throughout the whole research process, contributing both their subject expertise and information organization skills (Gold 2007; Gold 2010; Garritano & Carlson 2009; Soehner *et al.*, 2010). User studies in terms of researchers' data curation and management needs (Bennett, 2010; Borgman, 2010; Feijen, 2011; Jahnke *et al.*, 2012), however, do not indicate enthusiastic interest among data practitioners, who are more concerned with being able to safely store their data, manipulate data sets, and publish quickly than they are with sharing data or the long-term preservation of research data. In this situation, viable data management and stewardship seem more likely to be accepted and effective, either when a university-wide collaboration exists among the library, IT department, research office, and active departments/research centers (Delserone, 2008) or when librarians are able to build deep relationships with researchers and become collaboratively involved at the earliest stages of supported research projects (Garritano & Carlson 2009). In 2010, ARL published a new survey of the e-science and data support services of ARL member institutions (Soehner *et al.*, 2010). The survey revealed that data curation and management strategies are still young and evolving, but even given current difficult budgetary times, continuous investments are made in the library's special affinity with the bigger trend of e-science research. A library cannot complete this task by itself, however, and institutional and even national/international collaboration on both infrastructure and service delivery is increasingly necessary.

One particular level of library data services – that of data computing services – has not yet been well documented or discussed in the data services literature. On the one hand, according to a researchers' user needs survey (Hogenboom *et al.*, 2011; Read, 2007), data analysis does not seem to be the most urgent issue facing faculty and students, when compared with data identification and data access needs. On the other hand, a campus commonly has an IT support department, statistical consulting services, or even departmental staff who are capable of serving the data computing needs of researchers (Cook *et al.*, 2001; Bennett & Nicholson 2004; Delserone 2008). However, by browsing

the job advertisements that are archived on International Association of Social Science Information Services and Technology (IASSIST)'s website, data computing experience, especially with respect to knowledge and usage of statistical and GIS software is now a very common requirement. Statistical software assistance is also apparent on the existing data services websites of many academic libraries (Bennett & Nicholson 2007). These observations point increasingly to the importance of understanding what exactly a library's data computing services can do to assist and participate in research activities, given a local context where users' needs have not been met and data librarians' expertise could be used. This research is now aimed at exploring this area, and focuses especially on topics that investigate: 1) through what specific assistance that this middle level data service can become a connection between the data reference and collection services at one end and the data curation services at the other end; 2) how a library's data computing services add value to pure technical support; and 3) how dynamically the institutional and library context and resources, data librarians' expertise and initiatives, and users' needs interact and negotiate with each other at each level of library data computing services.

Outside of the library research literature – for example, in the field of information systems and software training – many studies address how to design effective training session features to enhance the result of training, especially during demonstration-based trainings (Olfman & Bostrom 1991; Yi & Davis, 2003; Rosen *et al*., 2010). The recommendations include helping users retain declarative knowledge about how to use particular software through their pre-training preparation, during training-applicable exercises, by symbolic coding, through cognitive rehearsal of what is demonstrated, and by performing after-training retrospective tasks. In another essay (Wang, 2011) and a poster presentation (Wang, 2013), we reflected on the additional unique values of library data computing supports, comparing these to the value of regular technical support that is usually provided by campus IT departments; that is, our training sessions could include more external and broader information resources that users would be able to access in the future when they are conducting new types of data analysis by themselves that may be beyond the content included in the training session. We will discuss this again later in the article.

METHOD

We have maintained an inventory of the various data services activities performed since October of 2009, when the new data services first began operating. It is a text file that documents the users' information and the detailed data services that we provided to them. To better understand the scope and focus of data services, we conduct a simple content analysis (Franzosi, 2010) of our data services inventory, using a coding scheme composed of both data services levels and user categories to analyze the inventory. The results reveal the scope and focus of our data services. This also allows us to explore possible relationships or patterns between these two dimensions; for example, whether certain groups of users are more likely to have specific data services needs. In social sciences research, the coding schemes for content analysis are usually derived through an interactive process that includes identifying research interests, studying texts, using primary and theoretical coding categories, coding, refining categories, and further coding and refining until the researchers are satisfied that the schemes accurately reflect both the original text and research interests (Franzosi, 2010). Our final data services levels and user category scheme combine previous literature frameworks and our own services practices, so that we can easily refer our unique local services back to the classic data reference/collection/computing/curation services levels, and at the same time show how different user groups are distributed among these different levels of support on our local campus.

We then present our library's experience in establishing a data computing service as a typical case of extending existing data reference services and establishing new service formats and programs. In his book, *Case Study Research: Design and Methods*, Yin (2003) defined a case study as most relevant when "how" or "why" questions are posed and when inquiring about a contemporary phenomenon within its real-life context. In this paper, examples of our particular services are concrete cases, where we examine the context of our campus and library capacity in detail, analyzing the different groups of users and their different needs and explaining our data services decisions regarding our support levels and extent in each situation as results of different methods of communication between our library and the faculty, students, and research centers. Detailed analysis of these services cases will also demonstrate the unique value of the library's assistance with the research process, when compared with purely technical training.

Except for the user categorizations, which are consistent with the content summary of our data services inventories, detailed case descriptions do not always fall into a particular cell of the cross tabulation of services subcategories because we often provide multiple levels of data services to one

particular user or user group. That is also a reason for using the two methods of analysis, to complement each other in describing different perspectives of the services that we provide. The content analysis answers the "what" questions by summarizing our services' scope and focus; the case descriptions tell "why" and "how" some of the typical services were decided and implemented. When discussing librarians' conducting of qualitative research, Farmer (2011) pointed out that this kind of nuanced data is especially good at capturing the reality of a library's everyday interaction and the reason for that interaction, and these unique cases can make significant contributions to the library literature and can be a data points for meta-analysis in the future.

We also report the results of our survey that we sent out by email to graduate students on campus to determine their interests in introductory workshops on statistical and GIS software. The students were asked to complete an online questionnaire and to tell us whether they would be interested in participating in these workshops if the library offered them. In 2009, the campus had about 3,000 graduate students and we sent out email surveys to all of them through each of the major schools' administrative offices; 161 of the students quickly replied and indicated great interest in many of the workshops that we thought we would be able to offer. Therefore, starting the following year, we provided these data computing and data visualization workshops. The statistics for the registration and attendance in these workshops are also reported as a comparison with the original survey. An additional voluntary feedback survey was sent to students who attended the workshops, to ask their opinion about the usefulness of different instructional features and contents. The items that were included in the survey questions varied due to the different content of the workshops for each kind of software package, but many features were the same for most workshops, so we are able to combine all the feedback results from different workshops within different time periods into one table; only the number of responses differs for each evaluated item.

CONTENT ANALYSIS OF DATA SERVICES RECORDS

When coding our data services records in terms of data services activities, and especially for the major categories, we started from the basic data services levels developed by previous researchers: data reference/ collection/ computing/ curation services. Two changes were then made to adjust the framework to fit our actual practices.

Data services literature talks about "data collection services" from the perspective of acquiring data resources products (for example, Social Explorer and PolicyMap) for all the library users' access. As we mentioned before, in the current Rutgers library system, the data services librarian at the main campus library is in charge of this type of data collection development, a process that we do not perform on our local campus, although our campus has access to the data resources that are acquired centrally. Instead, we help our research center and departmental users to collect their own data for institutional research or departmental use as one of our established services; for example, by conducting surveys, providing consultation on survey instruments, and developing data collecting systems. Our opinion is that this data collecting assistance should be one of our major service levels between data reference services and data computing services. This is because this level of service goes beyond merely identifying data sources and it is a necessary step for actual data usage.

The main campus library's digital curation services department is developing an institutional data repository, and has organized and educated a group of subject librarians who now participate in a data management support team. The vision is to provide data preservation minded data management support, ranging from data management plan writing support and metadata support up to the final research data archiving in our research data repository. Although we are not directly involved with the repository's infrastructure development, our data services librarian and our two science librarians are part of the team and are learning how to help faculty with data management issues that come before final research data curation and are essential for guaranteeing the quality of the final data. Therefore, data management (including data management plan consultation, data management workshops, and library guide development), rather than comprehensive data curation services, is our last service level.

Subcategories under each major data services level were first summarized based on each case of our service inventory, and then we normalized the vocabulary, coded the cases again, and counted the cases under each subcategory for each user group. Based on our experience, dividing our users into different categories is very useful, since they all have different kinds of data needs that must be analyzed in detail. Our user groups are classified into three categories: 1) teaching and research faculty; 2) campus departments, research centers, and offices; and 3) graduate students.

Table 1 shows the results of the coding and counting cases based on data service levels and user categories. The data services level scheme itself reveals our unique focus of the new data services (data computing services), which has an extensive range of tasks that can be summarized as

statistical software support, data analysis services, and data visualization services. Our data computing services are very comprehensive, ranging from instruction and tutoring on software usage to writing analysis/programming codes, generating specialized graphs and maps, and actually conducting data analysis. This also reflects our new data services librarian's expertise. Only the data reference services that involve our new data services librarian are recorded in our inventory, although all of our subject reference librarians also provide excellent and substantial amounts of assistance for users who want to identify the best research data for their projects. The Newark campus of Rutgers has a relatively small number of faculty (about 500), we have only advertised our data services to graduate students (about 3000), and many of these users have been unaware of our new data services (only 2 years); these factors all help to explain the overall small number of services requests that we received. Nevertheless, some of the projects have been time consuming due to the deep level of involvement and sometimes required data services librarians' self-learning of new skills, which have already generated a quite large workload for us.

The case counts for each user category showed that each user group had different needs, as reflected by the services that we offered to them. Faculty members who teach and who conduct their own research projects (either themselves or with the help of research assistants who conduct data analysis for them) interacted frequently with three levels of our data services, except for the data collecting services. These faculty members are also the major target group of our data management services. Research centers and other campus departments and offices are also aware of our data services. The research centers themselves usually have data computing personnel but they need our help with identifying data sources. We take this opportunity to introduce them to some data management concerns. Campus departments and offices usually do not have data experts of their own, and once they receive help with evidence-based research, they then express a need for more help, ranging from data collecting to actual data analysis. Graduate students, both in groups and individually, became enthusiastic data computing services customers, especially in terms of our statistical software training and support and our data analysis services.

Table 1: Cross Tabulation of Data Services Levels and User Categories
October 2009 – September 2012
(counts in cells are data services cases in our service records)

Data Services Levels	Teaching and Research Faculty	Campus Departments, Research Centers, and Offices	Graduate Students
Data Reference Services			
Identify data sources	3	3	
Identify data analysis resources			1
Data retrieval	1	1	
Data Collecting Services			
Conduct surveys		1	
Consult on survey instruments		1	
Develop data collecting system		1	
Data Computing Services			
Statistical Software Support			
• **Statistical software instruction through workshops**			259
• **Assist statistical software usage in research method classes**	3		
• **Statistical software tutoring**	2	1	4
Data Analysis Services			
• **Data format conversion**	1		
• **Write programming steps for advanced data analysis**	2	1	3
• **Conduct advanced data analysis**		1	1
Data Visualization Services			
• **Create charts and graphs for numeric data**		2	
• **Map Census data using ArcGIS**	1		
Data Management Services			
Data Management Plan consulting	1		

SELECTED CASES OF DATA SERVICES FOR EACH USER GROUP

Some of the needs for library data services were clearly stated by the users, departments, or research centers and we were easily able to determine the scope of the service to provide. However, other data needs were vaguely presented because the project was still under development. In some cases, the needs were not even recognized because the users had not asked our library for prior help on issues related to data analysis. We analyzed the way that we determined these needs and how we endeavored to provide appropriate services in several typical cases that we have organized according to the different user groups of our library.

In some instances, we restricted our services to the reference and consultation level, without being further involved with the researcher's project, due to our limited capacity in terms of data computing techniques and staff resources. In other cases, we provided information that helped researchers make effective use of data based on their research questions, the characteristics of their data, and the statistical packages that they were using. In still other situations, we became deeply involved with the research projects and assisted the researchers or administrators in actual data analysis and the production of deliverable results and visualizations according to their specific needs.

The library system's large structure and the local campus environment have always influenced the interaction between data services librarians and their users and the levels and extent of the ongoing data services delivery. In our case, the main campus library has taken over the responsibilities of data collection and has developed the infrastructure for a data repository, which serves as our local data services resource and backup. Our local campus IT services provides data analysis facilities by installing statistical software packages on all computers in public computer labs. However, they do not provide further assistance or training support for the software usages. This then became our opportunity to fill in a gap on campus. Based on a previously established deep relationship between our engaged library instructional services and one of the research office senior staff members, the research office on our campus began to collaborate with and support us in terms of data management education for our faculty members.

Interestingly, in many cases, multiple levels of data services were provided jointly to a particular user category or in a particular project. On occasion, users needed multiple levels of support, while at other times, data librarians consciously educated the users with additional data management awareness, even if the users were not aware of, or were not interested in, the topics at that time.

Research Center Data Bank Projects

Several research centers on the Rutgers-Newark campus were actively interacting with our library concerning their data-related projects, and they eagerly sought help about potential data issues, although they did not know exactly what they really needed at the time. It took many meetings to understand their needs and to determine how we were able to help.

In one case, a newly created research collaborative was expecting to obtain access to restricted research data on urban schools and students. The collaborative planned to host the data, and at the same time, to integrate census demographic data at the related geographic levels, so that researchers within the center could easily access and analyze the data for a series of analyses on the topic. Since the project investigators did not have experience in building this type of data bank, they approached our library at the planning stage of the project, in the hope that the library might have some resources or support to help with the implementation of the project.

After several meetings with members of the research collaborative, we determined that their specific need was to collect data and continuously manage a data portal. We recognized that the work actually required a full time data manager to maintain the data portal and to provide customized data extraction services for the researchers. Building this type of a data bank is not a one-time effort, and the skills required to deal with the data go beyond basic knowledge of statistical data analysis and techniques. Establishing and maintaining a data server requires additional skills in computing and networking. We concluded that these expectations were not within our data librarian's scope of work.

We eventually helped the research collaborative by selecting and compiling a list of online resources of available and relevant data and statistics. We also contacted several other research collaboratives and surveyed their strategies for establishing data servers and computing platforms for depositing new data and for extracting data for research. These findings were then shared with our research collaborative to provide guidance in implementing their own data server.

In another case, a social science research school was to create a web-based data bank that would provide easy access for the general public to all levels (local, county, and state) of government

performance and community indicators data. The principal investigator of the project also approached our library at the beginning stage. He believed that our library was collecting and hosting some ready-to-use governmental and institutional data; in addition, he perceived the library as an experienced data information curator that had valuable expertise on the issues of locating and collecting all kinds of data.

This project quickly started a pilot website with a departmental web server and found a software developer to design the prototype of the public interface, while a group of graduate students began collecting the widespread data that needed to be integrated. We determined that we were able to provide reference services to their project, by identifying additional data resources as needed and by acting as a consultant to help review their project grant application and make recommendations based on our reference librarians' expertise.

Although our formal data services for these two research center data bank projects were limited to the data reference services level, we did try to cultivate an ongoing relationship and conversation with their data managers and to obtain continuous updates on their progress. We have also tried to plant the ideas of implementing a well-designed data documentation system and considering long-term data preservation. So far, these suggestions are not being applied due to pressures of project completion or lack of immediate need. However, our active discussions with them about their data management issues and concerns actually resulted in a collaborative conference presentation on the topic of their data management infrastructure design, the difficulties of aggregating data from multiple resources in multiple formats, and the innovative technical solutions that they found for web delivery of data information. Our data services librarian was able to lead the data managers of these research centers to reflect on their everyday work and to synthesize their common difficulties, efforts, and strategic solutions.

For both of these two cases, our data services were not able to go much beyond traditional data reference assistance for the users. Even though we have endeavored to direct the project managers toward thinking about data management and preservation, no concrete implementations were taken into action, mainly because their own data managers had no immediate pressures to solve these long term data issues, nor did our data services librarian have the skill or time to fully participate in the establishment of their local data infrastructures or their everyday data management work. It is still very meaningful to have all the conversations with these research centers – first, as librarians, to provide perspectives from information preservation perspective, and second, as a cross-disciplinary observer, to facilitate each singular project's awareness of similar other projects on the campus, and the similar or different issues that they face, and possible multiple solutions to the problems that they encounter.

Institutional Statistical Analysis

The Rutgers-Newark Writing Program was among the early contacts with our library requesting data services assistance. This office had data on students' scores on placement tests and class exams but the administrators of the Program did not have the statistical analysis skills to use the collected data. Specifically, the administrators needed to determine whether the placement test that was being used was effective as a tool for filtering the incoming students into the different levels of the English writing program. This particular need matched up with our data librarian's data analysis expertise, the dataset was relatively small, and the research questions could be operationalized directly with the available data, so we decided to provide support by obtaining the data and conducting the analysis for them.

To prepare for the analysis, our data services librarian did a thorough literature review about the topic, in terms of methodological issues, and then chose the appropriate statistical strategies to conduct the analysis. During the whole process, our data services librarian had several meetings with the administrators to identify the exact research questions, the specific variables in the dataset, the operation of the placement test, the hypothesis they had in mind, and finally, the most effective way for the results to be presented without statistical jargon so that the administrators could understand them and not be overwhelmed by too much detail. In the end, the data analysis results were provided in a way that enabled the administrators of the Writing Program to evaluate the effectiveness of the placement tests and identify better indicators for students' placement into appropriate writing classes.

In other cases, our assistance was more focused on data collecting and data visualization. The Diversity Research Center on campus used a nationally comparable survey instrument to conduct a survey on the graduating senior students. Our data services librarians were fully involved with the whole process of the survey implementation, from applying for institutional IRB approval to ordering the email survey services. Once the survey reports were available, we downloaded and saved them, compiled them together with previous years' data, and conducted a data visualization focused on the topic of diversity for the center's director to present to leaders on other campuses. The results helped

confirm that Rutgers-Newark students are racially and ethnically diverse compared to our peer institutions, and that diversity experience is apparent, as are potential gains for students' intellectual and civil engagement developments.

Research universities usually have a dedicated institutional research office that collects university level data and on some general topics, for example, the demographic components of the faculty and students body and students' overall learning and living experiences. However, rich data resources exist in each unit of the university, and very specific research questions and issues are of interest to each unit within the university. Here, a lack of data professionals equipped for each department becomes an opportunity for our library's new data services to make contributions. These small-scale and short time durations of data collection, analysis, and presentation projects can feasibly fit into a data librarian's workload, and thus could represent a successful services model, especially when the department has clearly defined concrete research questions and the data gathered are sufficient to provide answers.

Statistical Software Workshops for Graduate Students

Our campus computer labs have all the basic statistical and GIS software installed for students to use, and some software is even available through a remote server so that students and faculty can conduct data analysis remotely through their own desktops. However, training and assistance for use of this software is not available through the campus IT services. Our reference librarians frequently conduct information literacy sessions and they have been very successful at educating the users about the existing data resources for their specific study areas. Our peer libraries have also been offering workshops on statistical software packages as a part of their data computing services. Consequently, we decided to start these types of workshops as well, based on our data services librarian's expertise in using multiple software packages to perform data analysis.

In order to understand graduate students' specific needs for classes to learn about statistical software packages, we conducted an online survey at the end of 2009 to gather data. We asked about their interests in the following software packages: SPSS, Stata, SAS, ArcGIS, and GoogleEarth. From the 3000 graduate students on campus, we received 161 responses, of which 142 expressed their interest in learning how to use these packages through introductory workshops offered by our library (see Table 2). Variations were found among different departments and schools in terms of which software the students were most interested in learning. For example, by examining the percentage of students within each school who are interested in each workshop, one can see that within the Business School, the majority of respondents were interested in learning SAS; in the School of Public Affairs and Administration, most students were interested in the use of Stata; while the students in the Graduate School were most interested in learning SPSS.

Table 2: Graduate Student Interests in Library Workshops by School

	Count of Students Interested in Introductory Workshops (Percentage of Students within Each School Interested in Each Workshop)						
Schools	**Intro to SPSS**	**Intro to Stata**	**Intro to SAS**	**Intro to ArcGIS**	**Intro to GoogleEarth**	**Other Class Interests**	**Total Count**
Graduate School	37 (66%)	21 (37%)	36 (64%)	16 (29%)	23 (41%)	9 (16%)	56
Business School	14 (45%)	13 (41%)	24 (77%)	5 (16%)	11 (36%)	4 (13%)	31
College of Nursing	14 (77%)	1 (6%)	3 (17%)	3 (17%)	6 (33%)	3 (17%)	18
Division of Global Affairs	20 (80%)	14 (56%)	16 (64%)	11 (44%)	10 (40%)	3 (12%)	25
School of Public Administration	1 (25%)	3 (75%)	2 (50%)	0 (0%)	1 (25%)	2 (50%)	4
School of Criminal Justice	3 (50%)	1 (17%)	1 (17%)	3 (50%)	2 (33%)	2 (33%)	6
Other	2 (100%)	1 (50%)	1 (50%)	1 (50%)	1 (50%)	0 (0%)	2
Total	91	54	83	39	54	23	142

As this was not a strongly advertised online survey from the library and it provided no incentive for responding, we believed that the over one hundred quick responses was a good indication of the potential needs from the graduate students for these workshops. Therefore, we planned workshop sessions on these specific packages, selected computing labs located close to the school that had the major interest, and we prepared exercises that were related to the topics that were familiar to the students. We also invited the data librarian from the main campus to offer a workshop on the statistical software package R, which is his area of expertise. Over the course of two years or more, our data services librarian has been adding more workshops whenever new statistical software is available or when new online database and mapping tools are added to our library collection.

Table 3 shows the workshops that we have offered so far, and it also shows the statistics for student registration and actual attendance. We do send our email reminder before each workshop is conducted, and we try to schedule the workshops during campus "free periods" on weekdays when there are no regular courses scheduled, but we always have more students register than attend on the workshop day. According to some email explanations of these absences, many are due to unexpected conflict with classes or projects during the middle of each semester.

Table 3: Data Services Workshop Registration and Attendance

Workshop	Number of students registered	Number of students attended
2010 Spring		
Introduction to SPSS	25	12
Introduction to SAS	26	15
Total	51	27
2010 Fall		
Introduction to SPSS	26	13
Introduction to SAS	32	10
Introduction to Stata	24	9
Total	82	32
2011 Spring		
Introduction to SPSS	46	14
Introduction to Stata	40	13
Introduction to SAS	44	10
Total	130	37
2011 Fall		
Introduction to SPSS	31	24
Introduction to Stata	44	22
Introduction to SAS	35	13
Introduction to R	38	24
Mapping Census Data	35	12
Total	183	87
2012 Spring		
Introduction to SPSS	43	19
Introduction to Stata	34	9
Introduction to SAS	32	14
Introduction to R	28	9
Mapping Census Data	44	14
Using PolicyMap & Social Explorer	32	5
Data Management	50	6
Total	263	76

We also found it interesting to discover that a certain number of graduate students were actually interested in learning multiple statistical software packages. Therefore, our data services librarian prepared class material about how to choose among the different packages. In addition to the workshop material and class notes, other online resources were compiled and made available as part of one of our library research guides. These resources included video and online tutorials, help documents, programming notes, and information on user groups. Students who are not able to attend the workshops receive an email with a link to these resources so that they can study by themselves according to their own schedules and needs.

In addition, we embed data documentation ideas and best practices into these statistical software instructions and tutoring processes; for example, we teach how to document data and research processes using programming files and commenting on data analysis codes, rather than clicking through software menus, as one early step of best practices in data management. While assisting graduate students with their research method class assignment, we discovered that even many professors who introduce data analysis procedures to students when they teach research method classes are not emphasizing the importance of always generating programming documents to record the data manipulation process and thus to enable better collaboration and data sharing and preservation. We believe this is a great opportunity for data librarians to help cultivate the habit of documenting the research process among these future e-research scholars.

Table 4 shows workshop attendants' ratings of the usefulness of these instructional features and contents. Among those who voluntarily provided feedback, over 90% think that the library guide with additional information that they can use in the future, the information on comparison of statistical software packages that we presented in the class, and the hands-on exercise that they performed in the training session are all useful or very useful. Although the usefulness of specific information about data preparation, data analysis best practices, and limitations of statistical software is not clear to some students yet, each of the these features and contents of the workshop were deemed useful or very useful by more than 70% of the respondents.

Table 4: Students' Evaluation on Usefulness of Data Services Workshop Features and Contents

	Rating of Usefulness				Number of Respondents	Average Rating	Percentage of Useful/Very Useful
	Not Very Useful	Don't Know Yet	Useful	Very Useful			
	1	2	3	4			
The Whole Workshop	3	4	23	44	74	3.46	90.54%
Hands-on Exercise	3	3	23	43	72	3.47	91.67%
Other Online Tutorials and Resources Introduced in Class	2	7	18	42	69	3.45	86.96%
Comparison of Statistical Software Packages	1	1	7	20	29	3.59	93.10%
Limitation of Statistical Software as a Tool	3	8	12	23	46	3.20	76.09%
ICPSR's Data Preparation Guide	1	10	9	22	42	3.24	73.81%
Data Analysis Best Practice	1	9	25	35	70	3.34	85.71%

Library Guide for Data Services	0	5	18	48	71	3.61	92.96%
Library Guide for Data Management	1	9	17	45	72	3.47	86.11%

Individual Students' Statistical Analysis

As individual students gradually learned of our data services, especially through the workshops that we provided on the use of statistical software packages, they started to contact us about their specific data computing needs. We provided both face-to-face and email consultations on entering data for statistical analysis, restructuring data and generating new variables, introducing new statistical software, finding the correct program syntax, and performing a specific statistical analysis on a particular platform.

In addition to the workshops that we offered each semester, we made available many print and online resources for students to learn specific software packages. Nevertheless, some students still needed a one-on-one tutorial when they first started to use the software, or just psychological support when dealing with a totally new package. Our data services librarian was able to work individually with graduate students and apply her experience with the different software packages to show the students many tricks, warnings, and lessons that were especially valuable for a new user.

Students would also occasionally get stuck during the process of conducting data analysis. Although they knew the basic techniques, they were not able to proceed because a particular problem needed an additional programming package, which they knew should exist, but didn't know where it was or how to use it. Our data services librarian worked with these students by providing programming suggestions and even contacted a software package developer on behalf of the students about the usage of a certain package. The data services librarian knew not only whom to contact, but also knew better about how to frame the question to ask.

I interacted with two doctoral students who visited our data services librarian frequently during their dissertation writing periods. One doctoral student was dealing with a large national survey data set and had no previous personal experience and not enough support time from her advisors. She started to see our data services librarian to discuss steps for data cleaning and analysis and to seek help with programming with Stata software. The other doctoral student had actually paid for professional statistics consulting services outside of our campus, but failed to obtain satisfactory results because the statisticians lacked expertise in the subject and had limited understanding of her model design problems. She came to our services librarian eventually, and we sat down together to go over her research question, correct the statistical model, study previous research in the field, learn new analysis techniques, and finally complete the necessary data analysis together.

The unique contribution of our data services in these cases is that we are not merely technical experts, but we also provide subject expertise and quantitative research methodology consultation. This enables us to assist graduate students at a deeper level, from data analysis methods choice to actual data analysis and troubleshooting.

Individual Faculty Member's Book Project

Faculty members engaged in their own research projects sometimes experiment with innovative research ideas and would like to incorporate more data visualization to help present and support their arguments. However, they might not have time to learn these new techniques of data visualization. Purely technical support may be lacking on campus, but even when this type of service is available, technicians usually are not familiar with the research topic or the research design and refinement processes. Therefore, our data services librarians again can fill in the gap and work closely with the faculty members to develop their research.

I had an opportunity to work with one faculty member who was developing a new book project. The faculty member intended to apply a theory of mutuality as a new foundation for people to understand the social, economic, and political issues of inequality. He was also interested in empirically and visually showing that where racial concentration is intensifying, poverty is also deepening. In addition to pointing him towards possible data resources, I was able to help him to map related census data, and along with his concept development, to refine the indicators and map them accordingly.

Again, this service delivery involved regular communication through emails and meetings. At the beginning of the project, the researcher himself needed to experiment with visualizing some of his hypotheses, and to define and redefine his ideal indicators after seeing the preliminary results. This particular service required the data librarian to develop skills in a broad way, with the use of GIS software and geographical data visualization techniques, rather than through in-depth statistical analysis.

As greater numbers of researchers begin to make use of new data analysis and presentation technologies, greater opportunities arise for data services librarians to become involved within the research process, to act as the bridge between information technology and the subject knowledge. It is impossible for the librarian to have all the skills and knowledge prepared beforehand. However, a librarian who is good at understanding research subjects through thorough literature research and review and from multiple disciplinary perspectives, and who can learn new techniques quickly, is at a distinct advantage.

Data Management Support to Faculty

The Rutgers Library is one of those pioneering libraries that are developing institutional data repository infrastructure and services. Researchers are not yet widely aware of the needs of data preservation. They are not interested in depositing their data, even when this infrastructure is available. However, we believe that the library has the responsibility to curate at-risk data and that the research community will gradually treat data preservation as a norm. The leaders of data curation services on our main campus have been organizing a group of subject librarians to collectively learn about the practice and issues of managing metadata in preparation for future involvement with researchers' data management processes. Our data services librarian, together with two of our science librarians, are participating in this training.

At the same time, our local campus research office has become our partner in terms of educating faculty about the data management issues and best practices. Initiation of this process occurred thanks to our senior reference librarians' long time collaboration with the research office. So far, our library and the research office have co-sponsored a data management workshop for faculty members, with a good attendance, and we are organizing another one in this semester.

Our own science librarians are not just our partners among the main library's data management support team; they also actively help us outreach to science departments and classes. We were invited to a science department to give a presentation on data management, and all of the graduate students in that department were required to attend. The professor who organized this event was very satisfied with the content that we delivered. We have to admit, however, that no further involvement has taken place with faculty's data management practice yet, except for a very brief assistance with a science faculty member on her data management plan for a NSF grant application. Here, we again collaborated with our science subject librarian and the faculty member was very satisfied with our help.

Reaching out to faculty and research departments – either to raise the awareness of data preservation needs or to actually help them curate their research data – is still very challenging, even though a well-designed institutional data repository is in place. Although we have not yet been involved in any deep level data management projects, one thing we have learned is that we could equip ourselves with updated best practices and tools in the field, thereby preparing ourselves for upcoming opportunities. Another helpful policy is for the data librarian to actually work closely with subject librarians, as they could be more effective as advocates and could help to discover potential needs with their existing well-rooted relationships with various faculty members and departments.

CONCLUSION

We are gradually establishing the boundaries of our new data services at the library, with a focus on data computing services. Thinking back to the first few months, when we were struggling with figuring out the exact needs of different users who approached us for support, we realized that a need existed for a common vocabulary that the library and users could use to communicate more clearly about needs, capacities, and deliverable services. This paper has combined the four classic levels of data services that form the major framework, together with our own concrete data services subcategories. Although our list of levels and categories of data services is not comprehensive – for example, even data computing services could go beyond small-scale training and data analysis to the level of grid computing using much complicated technologies and tools – our scheme will still be beneficial for many other libraries with similar human resources and organizational capacities as they start to plan and design their data services.

Every institutional environment will be different, however. Some might have multiple departments that also provide data related support that could provide useful collaborations, while others might lack certain research support that library data services can fill naturally, as did, for example, our statistical software support services. Each library structure is also unique. Data services, and especially new programs, need to make good use of the existing resources, collaborate with subject librarians on new service delivery, use existing outreach networks developed by senior librarians, and advocate data services to colleagues and thereby to associated groups of users.

Users' needs are varied and many data services cases are uniquely based on users' different levels of data computing capabilities. Some users need hands-on tutoring to alleviate their technology anxiety, while other users need comprehensive support in data model design and concrete data analysis support. However, as we discovered – and consistent with previous literature – the changing institutional environment, the library and librarian's capacity, and the users' backgrounds and needs are the three main factors that dynamically and continuously interact with each other to determine the levels and extent of the data services that are delivered.

Cultivating an earlier and deeper relationship with researchers and research centers is beneficial in the long run. Remaining aware of the possible complex project issues related to data, besides direct data source identification and data analysis (for example, the needs for a dedicated computer server administrator and web developer), will help libraries to make decisions about their support format and level based on their staff capacity. It will also reveal the possible skill sets or areas of knowledge where the data librarian may require training or frequent updating. Sometimes, researchers or research project data managers are unaware of (or uninterested in) long-term data management or may be too occupied with their immediate tasks to be able to address long-term issues. As data services librarians, we could be their partners or someone with whom to have a conversation, and thereby help them to think a bit further, to update them on current data-related topics, and assist them with better research practices.

Data services librarians should also expect a longer cycle of service delivery, which includes communication and interaction in multiple formats: meetings, emails, lab tutorials, and class instruction. These all provide the opportunity to clarify issues related to data needs and services. Communication must exist throughout the process, from identifying data resources to strategic selection of data for analysis and data results presentation. If the researchers keep adjusting their original research plans, the delivery of the services might need to go through cycles of definition and redefinition of the variables or mapping and remapping of the results according to the changes.

Lastly, we want to emphasize the additional values that data services librarians could add to their delivery of data computing assistance to users. Similar to regular IT technical support, we can use demonstrations and hands-on exercise strategies to help users retain the specific knowledge they learned during the session. We can also discuss more topics prior to skill trainings; for example, comparing different statistical software packages in terms of which package is good at which particular analysis and with which kinds of data. Topics could also be introduced on the limitations of statistical software as a tool, as software use still depends on the researchers to command the appropriate analysis strategies and still leaves interpretation of the results up to the researchers. In this way, students understand that software tools are not going to give them all the answers, while on the other hand, they may feel less stressed in learning a technical skill that is not as difficult as a course in comprehensive statistics or mathematics. During the training session, data librarians could introduce best practices of conducting and documenting the research procedure, by emphasizing the values of diligent writing of programming files and detailed annotations, rather than simply recommending the ease of clicking through menu options. After a one-shot workshop, data librarians should provide prepared library guides that aggregate additional resource that allow users to go over the workshop material again and continue learning and solving problems by themselves, and that direct users to groups with additional expertise. Because data librarians are well trained in the research process and research methodology and are able to conduct comprehensive literature reviews to learn research techniques that are effective within a small research subject area, the services that we provide can be comparable to and even more accurate than those provided by professional statistical services.

Above all, we encourage our colleagues in other academic institutions to observe and grasp the local opportunities, whether this is a lack of certain services or established collaborative relationships, and to step further and partner with research centers, departments, researchers, and students on innovative programs and projects. Data librarians have the unique advantage of being able to see from a multidisciplinary perspective, with visions of future research trends and academic environmental changes, and being able to talk the language of both researchers and technicians. All of these capabilities will prove especially valuable in this coming age of e-science, when multi-disciplinary networked and data intensive research is going to be the norm of academic activities.

REFERENCES

Adams, M.O. (1982), "Models of data library development and information system services: an overview", *Library Trends*, Vol. 30 No. 3, pp. 377–381.

Bennett, T.B. (2010), "Research data services at Singapore Management University: engagement summary report", *Research Collection Library*, Paper 5, available at http://ink.library.smu.edu.sg/library_research/5 (accessed 7 September 2012).

Bennett, T.B. and Nicholson, S. (2004), "Interactions between the academic business library and research data services", *Portal: Libraries and the Academy*, Vol. 4 No. 1, pp. 105–121.

Bennett, T.B. and Nicholson, S.W. (2007), "Research libraries: connecting users to numeric and spatial resources", *Social Science Computer Review*, Vol. 25 No. 3, pp. 302–318.

Borgman, C.L. (2010), "Research data: who will share what, with whom, when, and why?", paper presented at the Fifth China-North America Library Conference, 9-10 September 2010, Beijing, China, available at http://works.bepress.com/borgman/238 (accessed 7 September 2012).

Cook, M.N., Hernandez, J.J. and Shawn, N. (2001), *Numeric Data Products and Services,* Association of Research Libraries, Washington, D.C., available at: http://www.arl.org/bm~doc/spec263web.pdf (accessed 10 September 2012).

Delserone, L.M. (2008), "At the watershed: preparing for research data management and stewardship at the University of Minnesota Libraries", *Library Trends*, Vol. 57 No. 2, pp. 202–210.

Duke, L.M., Vogel, K. and Wilson, S. (2007), "Faculty-library collaboration in dataset and statistics management", *Technical Services Quarterly*, Vol. 24 No. 4, pp. 43–54.

Farmer, L. (2011), "Qualitative research and the librarian", in Cook, D. and Farmer, L. (Eds.), *Using Qualitative Methods in Action Research: How Librarians Can Get to the Why of Data*, Association of College and Research Libraries, Chicago, IL, pp. 3–10.

Feijen, M. (2011), "What researchers want: a literature study of researchers' requirements with respect to storage and access to research data", available at http://www.surf.nl/en/publicaties/pages/whatresearcherswant.aspx (accessed 5 September 2012).

Franzosi, R. (2010), *Quantitative Narrative Analysis*, SAGE, Thousand Oaks, CA.

Garritano, J.R. and Carlson, J.R. (2009), "A subject librarian's guide to collaborating on e-science projects", *Issues in Science and Technology Librarianship*, Vol. 57, available at http://www.istl.org/09-spring/refereed2.html (accessed 10 October 2010).

Geraci, D., Humphrey, C. and Jacobs, J. (2010), *Data Basics: An Introductory Text*, Interuniversity Consortium for Political and Social Research, Ann Arbor, MI.

Gold, A. (2007), "Cyberinfrastructure, data, and libraries, part 2: libraries and the data challenge: roles and actions for libraries", *D-Lib Magazine*, Vol. 13 No. 9/10, available at http://www.dlib.org/dlib/september07/gold/09gold-pt2.html (accessed 25 October, 2010).

Gold, A. (2010), "Data curation and libraries: short-term developments, long-term prospects", *Office of the Dean (Library)*, available at http://digitalcommons.calpoly.edu/lib_dean/27 (accessed 4 September 2012).

Hogenboom, K., Phillips, C.M.H. and Henseley, M. (2011), "Show me the data! partnering with instructors to teach data literacy", in *Declaration of Interdependence: The proceedings of the ACRL 2011 conference, March 30-April 2, 2011, Philadelphia, PA*, Association of College and Research Libraries, Chicago, IL, pp. 410- 417.

Jacobs, J. A., & Humphrey, C. (2004), "Preserving research data", *Communications of the ACM*, Vol. 47 No. 9, pp. 27-29.

Jahnke, L., Asher, A. and Keralis, S.D.C. (2012), "The problem of data", Council on Library and Information Resources (CLIR) Report, pub. #154, available at http://digitalcommons.bucknell.edu/fac_pubs/52/ (accessed 9 September 2012).

Kellam, L.M. and Peter, K. (2011), *Numeric Data Services and Sources for the General Reference librarian*, Chandos Publishing, Oxford, UK.

Lewis, D. (2007), "A strategy for academic libraries in the first quarter of the 21st century", *College & Research Libraries*, Vol. 68 No. 5, pp. 418–434.

Mooney, H. and Silver, B. (2010), "Spread the news: promoting data services", *College & Research Libraries News*, Vol. 7 No. 9, pp. 480–483.

Olfman, L. and Bostrom, R. (1991), "End-user software training: an experimental comparison of methods to enhance motivation", *Information Systems Journal*, Vol. 1, pp. 249–266.

Read, E. (2007), "Data services in academic libraries", *Reference & User Services Quarterly*, Vol. 46 No. 3, pp. 61–75.

Rosen, M. a., Salas, E., Pavlas, D., Jensen, R., Fu, D. and Lampton, D. (2010), “Demonstration-based training: a review of instructional features”, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 52 No. 5, pp. 596–609.

Rowe, J. (1982), “Expanding social science reference service to meet the needs of patrons more adequately”, *Library Trends*, Vol. 30 No. 3, pp. 327–334.

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, D.C., available at http://www.arl.org/bm~doc/escience_report2010.pdf (accessed 10 September 2012).

Wang, M. (2011), “Serving the future e-science researchers: library computing data services for graduate students on the Rutgers Newark campus”, *Synergy: News from ARL Diversity Programs*, No. 7, available at http://www.arl.org/bm~doc/synergy1011.pdf (accessed 10 September 2012).

Wang, M. (2013), "Beyond technical instructions", poster presentation at: *14th Annual VALE/NJ ACRL/NJLA CUS users' conference*, January 5, 2013, New Brunswick, NJ, available at http://www.valenj.org/conference/session/p17-beyond-technical-instructions (accessed 15 January 2013).

Yi, M.Y. and Davis, F.D. (2003), “Developing and validating an observational learning model of computer software training and skill acquisition”, *Information Systems Research*, Vol. 14 No. 2, pp. 146–169.

Yin, R. K, (2003), *Case study research: Design and methods*, Sage Publications, Incorporated, Thousand Oaks, CA.