Proceedings of the

ASIA-PACIFIC CONFERENCE ON LIBRARY &
INFORMATION EDUCATION & PRACTICE 2011
(A-LIEP 2011)
Issues, Challenges and Opportunities

Putrajaya, Malaysia, 22-24 June 2011
URL http://fim.uitm.edu.my/a-liep2011

Organized by
Faculty of Information Management, Universiti Teknologi MARA
Shah Alam, Selangor, Malaysia

In collaboration with
Tun Abdul Razak Library, Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia

Edited by
Fuziah Mohd Nadzar
Mohd Sharif Mohd Saad
Szarina Abdullah

Shah Alam, Selangor, Malaysia
2011

Data Needs and Data Services: A Case Study

MINGLU WANG, minglu@rutgers.edu
KA-NENG AU, au@rutgers.edu
Media & Digital Library Services, John Cotton Dana Library, Rutgers University Libraries

ABSTRACT

Our library's traditional reference service is effective in helping users discover and gain access to data, while our new data services focus on management, analysis, and presentation of statistical data. This paper describes the process of how we gained a better understanding of the variety of users' data needs, and how we gradually established some new data services based on our current capabilities. How were we to understand and evaluate our users' needs of data services? What are our users' unspoken data needs? How would we extend traditional data reference services? How should we plan programs to meet these needs? What was within our means and what was beyond our capacity? 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. Typical examples are described and analyzed, and then general concepts, classification, and conclusions are given to inspire the design and implementation at other libraries. 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. Traditional data reference services are still the fundamental part of a reliable data service, although many reference librarians are doing the job without using the title of data services. Data services, even computing data services, can be provided at different levels based on the users' needs and library's resources. Computing data services can involve group training and statistical analysis assistance, but it can be extended to data project management and data management as more and more research is required to have formal and accountable procedures when dealing with data.

Keywords: Data services; data reference; data analysis; data management; user needs.

INTRODUCTION

When thinking about the future of academic libraries and the role of librarianship, one major expectation is for the library to be involved in deeper levels of academic research. For data services specifically, traditional library reference services have focused mainly on data access and discovery and have done quite well by pointing users to the right data resources among the enormous and continually changing information pool. However, many disciplines involve large scale data collection in the laboratory or field. New forms of research collaboration, publishing, and applications have developed not just within academe, but also in all kinds of institutions, in order to facilitate evidence-based decisions. How then can a library be involved one step further into the research process? There seem to be opportunities and challenges at the same time.

The Dana Library serves the Newark campus of Rutgers, The State University of New Jersey, and is part of the state-wide Rutgers University Libraries. Although data collection and repository services are handled by a library on the main New Brunswick campus for the whole library system, the reference librarians at our local campus library are very competent at providing data access services. These services 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 to learn the use of statistical databases; and online resource guides for both undergraduate and graduate courses.

But when the active researchers and centers on the Newark campus began to realize and articulate their new data related needs, which went beyond just data source identification, and approached our library for possible support, our library responded by submitting 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 under the direction of an experienced reference librarian and to start the process of venturing into the new service of data computing areas, including data management, analysis and presentation.

We started from the point where we 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 a year of continuous interaction between the librarians and the different user groups on campus to understand the general data computing needs that had been articulated, and to know what we were able to do and what we were not able to respond to. Through the analysis of our own experience during the process, we will provide some classification and conceptual tools for other libraries that have or will start computing data services to better understand and design their local services.

LITERATURE REVIEW

Geraci, Humphrey, and Jacobs (2010) developed a very useful theoretical framework for a library's data services, especially the classification and possible scope of the data services into different aspects and levels. *Data collection and access services* have been conducted by reference librarians as a natural part or an extension of their subject expertise. *Data repository services* is a completely new institutional role for libraries, which had been collecting information both in print and digital formats and now believe that they should naturally take on the responsibility of helping researchers with the storage, management, and sharing of their research data too. *Computing data services* is usually defined as helping researchers with their actual usage of the data after that data has been obtained, which requires the librarian to be able to understand the statistical analysis of the data and be familiar with the technical tools that are appropriate for the analysis. It is closely related with reference services, but goes beyond the initial stage of identifying the data source, and requires technical skills that are necessary to deal with data related issues.

There are not many empirical studies on data services due to the recent emergence of the field. Among the current research on libraries' new data services (Delserone, 2008; Gold, 2007), most focused on data curation and data repository support. A few studies (Duke, Vogel, & Wilson, 2007; Garritano & Carlson, 2009; Gerhan, 1999) discussed data access and discovery support in light of the library's reference service and made suggestions about how to better prepare the reference librarians with further knowledge about actual data usage in analysis, how to collaborate with faculty members to improve access to data resources and collections, and how to participate in the whole process of a networked, data-driven science project with subject expertise. Another set of researchers (Mooney & Silver, 2010; Read, 2007) shared their success stories about improving data services' awareness and usage through various venues, such as awareness surveys and outreach to faulty members.

Bennett & Nicholson (2004) conducted research about business libraries and research data services, and the result showed that current practices of data services, in terms of the location, existence, awareness, and usage, are varied and lack a consistent pattern across academic institutions. Therefore, under this situation, more empirical studies about the concrete process of establishing new data services and about the interactions between libraries and users struggling to figure out the scope of needs and services will be beneficial for the whole library community to better understand and plan their own local services. Research about computing data services will be able to demonstrate typical scenarios that are representative of possible deliverable services.

METHOD

In his book, *Case Study Research: Design and Methods*, Yin (2003) defined that a case study is most relevant when "how" or "why" questions are being posed and when inquiring a contemporary phenomenon within its real-life context. In this paper, we will study our library's experience of establishing the new data service, as a typical case of extending existing data services and establishing new service formats and programs. We will examine the context of our campus and library in detail, analyzing the different groups of users and their different needs. Some typical examples of interactions will be described to demonstrate the process and the different ways of communication between our library and the faculty, students, and research centers. We will offer a classification and conceptual framework for future theoretical or practical application by libraries providing data computing services.

WHO ARE THE USERS, WHAT DO THEY NEED, AND WHAT CAN WE DO

Based on our experience, it is very useful to divide our users into different categories, and they all have different kinds of data needs that need to be analyzed in detail. Individual faculty members or students have their own specific research topic on which to focus and need help with concrete data analysis or presentation. Research centers, on the other hand, may have a big project that involves data resources on multiple subjects, or that is related to issues of building a data warehouse.

Some of the needs for library data services were clearly stated by the users, department, or research centres, and were easy for us to determine the scope of the service to provide, but 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 the issues related to data analysis. We will analyze how we determined the needs and tried to provide the services accordingly in several typical cases organized by the different user groups of our library.

There were instances where we restricted our services to the reference and consultation level, without being further involved with the researchers' project, due to the limit of our capacity in terms of data computing techniques and staff resources. There were cases where we provided information that helped researchers make effective use of data based on their research questions, characteristics of data, and the statistical packages that they were using. And there were situations where we were deeply involved with the research projects and assisted the researchers or administrators to actually analyze the data and produce deliverable results and visualizations according to their specific needs.

Research Centers' 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 communicate what they needed 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. They 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 up such a data bank, they approached our library from the planning stage of the project, with 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 found out that their specific need was to collect data and continuously manage a data portal. We could see that the work actually required a full time data manager to maintain the data portal and provide customized data extraction services for the researchers. It is not a one-time effort to build up a data bank, and the skills of dealing with the data goes beyond basic statistical data analysis knowledge and techniques. Establishing and maintaining a data server require additional skills in computing and networking. We concluded that these expectations were not in 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 of establishing a data server and computing platform for depositing new data and for extracting data to do research. These findings were then shared with this research collaborative for guidance in implementing their own data server.

In another case, the project was to create a web based data bank for the general public to have easy access to all levels (local, county, and state) of government performance and community indicators data. The principal investigator of the project approached our library at the beginning stage too. He believed the 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 having 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 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 found out that we were able to provide reference services to their project, by identifying additional data resources as needed, and acting as a consultant to help review their project grant application and make recommendations based on our reference librarians' expertise.

Institutional Statistical Analysis

The Rutgers-Newark Writing Program was among those early contacts with our library for data services. This office had data of 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 find out whether the placement test that was being used was effective as a tool to filter the incoming students into the different levels of the English writing program. The 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 do 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 they could understand 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 the appropriate writing class.

In another case, our assistance was more limited to advice on data analysis and presentation. The Diversity Research Center obtained student enrolment data from the campus Office of the Registrar in order to determine the racial and ethnic diversity of the student body. We helped the research assistants perform basic statistical analysis with Microsoft Excel and then derive suitable tabular and graphical displays of the results for a meeting presentation. The results confirmed that Rutgers-Newark students are racially and ethnically diverse and that the international students come from over seventy countries around the world.

Graduate Students

Our reference librarians frequently conduct information literacy sessions and 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. So we decided to start such workshops as well, based on our data services librarian's expertise in using multiple software packages to do data analysis.

In order to understand graduate students' specific needs about classes for learning about statistical software packages, we conducted an online survey to gather data. We asked about their interests in the following software packages: SPSS, Stata, SAS, ArcGIS, and GoogleEarth. We received 161 responses, of which 142 expressed their interest in learning these packages through introductory workshops offered by our library (see Table 1).

Table 1: Graduate Students' Interests in Library Workshops by School

Schools	Students Interested in Introductory Workshops Count and Percentage within Each School						
	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

The survey was distributed to all the graduate students on the Rutgers-Newark campus. There were variations among different departments and schools in terms of which software they were interested in learning. For example, 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.

So we planned workshop sessions on these specific packages, selected computing labs located close to the school that had the major interest, and prepared the exercises to be related to the topics that were familiar to them. We also invited the data librarian from the main campus to offer a workshop in his area of expertise. Over the course of three semesters, thirteen workshops were provided and about 130 students attended the sessions. Our data services librarian plans to learn more about GIS data and analysis using different software packages, so that we will be able to offer classes in that category in the future.

It was 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 (Womack & Wang, 2011a). These resources included video and online tutorials, help documents, programming notes, and information on user groups.

Individual Faculty Member's Book Project

We had an opportunity to work with a faculty member on developing his new book project. The faculty member intends to apply a theory of mutuality as a new foundation for people to understand the social, economic and political issues of inequality. He is also interested in empirically and visually showing that where racial concentration is getting worse, poverty is deepening as well. In addition to pointing him towards possible data resources, we have been helping him to map related census data, and along with his concept development, to refine the indicators and mapping them accordingly.

Again, this service delivery involved continual communication through emails and meetings. At the beginning of the project, the researcher himself needed to experiment with visualizing some of his hypothesis, 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 deep statistical analysis.

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 doing 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 too. But some students still needed a one-on-one tutorial when they first started to use software, or just psychological support when dealing with a totally new package. With experience with the different software packages, our data services librarian was able to work individually with graduate students, and showed them the tricks, warnings, and lessons that were especially valuable for a new user.

There were also times when students got stuck during the process of doing data analysis and 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 students by providing programming suggestions and even contacted a software package developer on behalf of the students about the usage of a certain package, because the data services librarian knew not only whom to contact, but also knew better about how to frame the question to ask.

CONCLUSION

Thus we gradually established 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 that approached us for support, in addition to understanding better about the specific project plan, we realized there is a need for a common vocabulary that the library and users can use to communicate more clearly about needs and capacities. This paper has provided some of the classifications and concepts that will be beneficial for other libraries to plan and design their data services. Being aware of the possible complex project issues related to data besides direct data identification and 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 the staff capacity. It will also reveal the possible skill sets or areas of knowledge about which the data librarian needs to be trained and updated frequently.

Data services librarians should also expect a longer circle 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 the issues related to data needs and services. Communication needs to exist throughout the process, from identifying data resources, to strategic selection of data for analysis, to data results presentation. If the researchers keep adjusting their original research plan, the delivery of the services might need to go through cycles of definition and redefinition of the variables to look for, or mapping and remapping the results according to the changes.

There were some other issues that came to our attention through our first year of the data services, and we have just started to try to address them. For example, we found out that users are not well prepared during their research design stage, when data is first planned to be collected. Some projects lack the thinking about future analysis when they design the data collecting tools and that causes potentially serious problems for data analysis in a later research period. Data storage is another issue that can be better addressed if the users are provided with some training on data file organization, backup, and version control. Data management related to data preparation for research, documentation, metadata creation, and eventually curation and archiving are new needs among many researchers and research centers even though they

themselves haven't fully recognized it. Some peer libraries and data archiving institutions have developed guidelines and best practices for data management and archiving. Our data services librarian participated in compiling an online guide (Womack & Wang, 2011b), which describes these new resources for our users' reference. All these emerging issues will be new challenges and also opportunities for data services librarians to better support their data users, and should be at the forefront when data librarians are communicating with users.

REFERENCES

Bennett, T. B. & Nicholson, S. (2004). Interactions between the academic business library and research data services. *Portal: Libraries and the Academy*, 4, 105-121.

Delserone, L. M. (2008). At the Watershed: Preparing for Research Data Management and Stewardship at the University of Minnesota Libraries. *Library Trends*, 57(2), 202-210.

Duke, L. M., Vogel, K. & Wilson, S. (2007). Faculty-library collaboration in dataset and statistics management. *Technical Services Quarterly*, 24(4), 43-54.

Garritano, J. R. & Carlson, J. R. (2009). A Subject Librarian's Guide to Collaborating on E-Science Projects. *Issues in Science & Technology Librarianship*.

Geraci, D., Humphrey, C. & Jacobs, J. (2010). Data Basics: An Introductory Text. ICPSR Summer Program.

Gerhan, D. R. (1999). When Quantitative Analysis Lies behind a Reference Question. *Reference & User Services Quarterly*, 39(2), 166-176.

Gold, A. (2007a). Cyberinfrastructure, Data, and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians. *D-Lib Magazine*, 13(9/10). Retrieved from http://www.dlib.org/dlib/september07/gold/09gold-pt1.html

Gold, A. (2007b). Cyberinfrastructure, Data, and Libraries, Part 2: Libraries and the Data Challenge: Roles and Actions for Libraries. *D-Lib Magazine*, 13(9/10). Retrieved from http://www.dlib.org/dlib/september07/gold/09gold-pt2.html

Mooney, H & Silver, B. (2010). Spread the news Promoting data services. *College & Research Libraries News*, 71(9), 480-483.

Read, E. J. (2007). Data Services in Academic Libraries: Assessing Needs and Promoting Services. *Reference & User Services Quarterly*, 46(3), 61-75.

Womack, R. & Wang, M. (2011a). *Rutgers University Libraries Research Guide: Data Services*. Retrieved from http://libguides.rutgers.edu/data

Womack, R. & Wang, M. (2011b). *Rutgers University Libraries Research Guide: Data Management*. Retrieved from http://libguides.rutgers.edu/datamanagement

Yin, R. K. (2003). *Case Study Research: Design and Methods* (3rd ed.). Thousand Oaks, CA: Sage.