

Serving the Future E-Science Researchers: Library Computing Data Services for Graduate Students on the Rutgers Newark Campus

Rutgers University has made this article freely available. Please share how this access benefits you.
Your story matters. <https://rucore.libraries.rutgers.edu/rutgers-lib/37135/story/>

This work is an **ACCEPTED MANUSCRIPT (AM)**

This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.

Citation to Publisher Wang, Minglu. (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* 7, 7-9.

Citation to this Version: Wang, Minglu. (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* 7, 7-9. Retrieved from [doi:10.7282/T3BZ64FM](https://doi.org/10.7282/T3BZ64FM).

Terms of Use: Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

Article begins on next page

Serving the Future E-Science Researchers
Library Computing Data Services for Graduate Students on the Rutgers Newark Campus

(Published in *Synergy: News from ARL Diversity Programs*, Issue 7, September 2011)

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

This is an interesting era when both the research world and library world are experiencing big changes; in the research world, networked data intensive research is fast becoming the new scientific paradigmⁱ, and in the library world, librarians are trying to partner with researchers deeper into their research processⁱⁱ. As the researchers are using and sharing more and more data, libraries are starting to provide data curationⁱⁱⁱ, access, and computing support^{iv}; thus creating a new niche of library data services.

At Rutgers, the main campus library has taken charge of the institutional data curation responsibilities, and the John Cotton Dana Library at Rutgers' Newark Campus, one of the Rutgers University branch libraries, has many experienced reference librarians who have been very competent in helping users to identify and access research data. In order to address the emerging data computing needs, the Dana Library has created a new Data Services Librarian position.

It's very exciting that several research centers and institutional departments on the campus have approached the library to express their data computing needs. A continuous conversation is going on between them and the library to try to define our role and capacity to help them with their needs. Although the needs of the centers and departments are being addressed, it is more challenging to address the data analysis needs of graduate students, a big user group, who are not as aware of the new services provided by library and whose data analysis needs are not specifically voiced to the library. My expertise in multiple statistical software packages has helped me to get to know the graduate students' data analysis needs and to help them through different venues. When working with these students, I always keep in mind that these are the future e-researchers, who are the main players in a new academic world where networked data intensive practices are the normal way of conducting research. With this in mind I try to integrate more than simply technical support, which are usually offered by the university's lab or computing services departments.

The Rutgers-Newark campus computing labs are equipped with multiple statistical software packages, but instructional support for using these tools is not available other than during students' research method classes, where the main focus of the class is statistics or research methodology and not how to use the software. The library was aware of other universities offering introductory classes on these data analysis packages; we decided to see if starting similar services at Rutgers was something students would be interested in.

In order to better understand students' specific interests and disciplines, I sent an online survey via email to the whole graduate student body. Within just a few days, more than 160 students responded. Of these responses 142 showed interest in library workshops on using data analysis software.

Interestingly, the results of the survey showed that individual students and departments/schools have unique preferences towards the kinds of software they want to learn more about. For example, Business School students are more interested in learning SAS (Statistical Analysis System); students from the School of Public Affairs and Administration are more interested in Stata, and students from departments in the Graduate School look for training in SPSS (Statistical Package for the Social Sciences).

Encouraged by the survey results, I started to offer workshops on using data analysis software SPSS, Stata, and SAS. Some students even sign up to learn all the packages. In response, I created a section comparing different software in terms of their particular advantages and disadvantages in order to help users choose the appropriate tool for their own projects.

In these introductory workshops, I included best practices of doing social science data analysis, especially on issues of documenting the analysis process by using program syntax file, rather than merely using the point-and-click menu options. I discussed the importance of doing research within the mindset of keeping their analysis easily replicable, because it makes their work efficient, especially in a team based collaborative academic environment, and it makes their work honest and potentially testable in long-term scientific procedures in a broader community of researchers. The new e-scientific paradigm needs not only more open data, but also a new research culture and practice that makes deep level of research sharable and replicable. This is why I think that teaching current graduate students, the future researchers, to develop a habit of documenting data analysis, extremely valuable.

Contextualizing data analysis within a broader process of research life cycle is another feature that I have been trying to integrate into my introductory workshops on statistical software packages. Through my experiences of working with students who brought their concrete data analysis difficulties to me, I realized that many problems could have been avoided if they had a better design of their research from the very beginning, even before they began collecting data. Graduate students have difficulties and are guideless about how to organize and clean the raw collected data in a way that turns it into a ready-to-analyze standard dataset. All of these issues require graduate students to have a broader view of the research process, because determining how data can be applied begins at earlier stages. In my workshops, I recommended document with guidelines on preparing research data in case they need it to process their current data and plan for future projects^v. I stress the importance of the holistic view of doing research: having a better plan at the beginning with the future analysis and possible result in mind.

Storing and organizing research data and analysis files are another problematic area that students seldom learn from their classes. During my conversations with the students, I have not heard concerns about data file management and archiving their data for the long term preservation and sharing. It's quite understandable since they are in their early stages as researchers and because even their faculty have just started to learn about these issues or may not even realized that these problems exist. The library world is by far the most enthusiastic group about advocating for data management best practices, and several of the leading library data services have been developing valuable guides for researchers^{vi}. In response I have participated in maintaining a library guide^{vii} on the topic with the data services librarian and will present it in all the workshops that I teach. Student feedback from the workshops has

been positive, and they appreciated the resources in our library guides, although they don't know yet whether these are going to be practically useful for them right now.

The librarian's role as educator has now extended into that of research partner, and within this new role librarians are quickly learning how to best serve the needs of users in this evolving digital and e-science academic world. Data services librarians are essential to providing services that address the new needs and issues related to all aspects of data based research processes. As this new function within libraries emerges we need to continually learn about data services needs, to adapt to users whose expressed needs are often vague and even unrealized. It is so exciting to be able to lead users, especially the future researchers, to think more broadly than their current project's concerns, to develop good documentation habits, and to be prepared for a new research environment.

ⁱ Hey, T. , Tansley, S., & Tolle K. edited (2009). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research. Retrieved from http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf.

ⁱⁱ Carlson, J., & Kneale, R. (2011). Embedded librarianship in the research context. *College & Research Libraries News*, 72(3), 167 -170.

ⁱⁱⁱ Gold, A. (2010). Data Curation and Libraries: Short-Term Developments, Long-Term Prospects. *Office of the Dean (Library)*. Retrieved from http://digitalcommons.calpoly.edu/lib_dean/27

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

^v Inter-university Consortium for Political and Social Research (ICPSR). (2009). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (4th ed.). Ann Arbor, MI.

^{vi} Hswe, P., & Holt, A. (2011). Joining in the Enterprise of Response in the Wake of the NSF Data Management Planning Requirement. *Research Library Issues*, no. 274, 12-17.

^{vii} Womack, R., & Wang, M. (2011). In *Library Research Guide: Data Management*. Retrieved April 11, 2011, from <http://libguides.rutgers.edu/datamanagement>