What is E-Science?

Whether you spell it "eScience" or "e-Science", the concept of e-Science has been cropping up in scientific research circles quite a bit over the last few years. I like to call it "big data science" myself, because e-Science deals with fields of research that handle massive amounts of data, whether these data come from astronomical sky surveys, supercolliders or genome wide association studies. A related term is "e-Research", which is a more general concept of research using digital technology (computing, networks, digital data) in fields including science but also social science and the humanities. *Science* magazine even devoted an entire online issue to the problems that arise from analyzing and storing large volumes of data.

In January 2011, the National Science Foundation (NSF) altered its data sharing policy to require data management plans in future grant proposals to the agency:

> *"All proposals must describe plans for data management and sharing of the products of research, or assert the absence of the need for such plans. The attachment name must include the words "Data Management Plan". NSF will not permit submission of a proposal that is missing a Data Management Plan."*

So how do scientists deal with massive amounts of data? The *Science* issue states that most researchers store their data locally (in their labs' computers) and may not have a backup plan for their data. In addition, many scientists are not able to analyze large data sets themselves, because they lack either the software or expertise required for analysis. Fortunately, many institutions have researchers or core facilities that provide data analysis consultations.

Libraries and librarians at academic research institutions are keenly aware of these issues and are prepared to make contributions to solutions for e-Science at their respective institutions. Toward this end, the Association for Research Libraries and the Digital Library Federation have teamed up to create the E-Science Institute. The Institute is designed to give librarians tools they can use to assist researchers in crafting Data Management Plans and dealing with other issues concerning e-Science and e-Research. The institute's sponsors include Northwestern University Library (NUL), along with more than 80 other academic libraries. The Institute's training initiative began in July and will continue through the end of 2011.

Northwestern University has four librarians attending the Institute:

- Claire Stewart - Head of Digital Collections, NUL*
- Cunera Buys - Science Librarian, Mudd Science & Engineering Library*
- Heidi Nickisch Duggan - Associate Director, Galter Health Sciences Library^
- Pamela Shaw - Biosciences & Bioinformatics Librarian, Galter Health Sciences Library^

(*official team representatives to the Institute; ^contributing attendees)
Additionally, contributions to Northwestern's team are being made by Anna Ren of the Mudd Science & Engineering Library.

At the Feinberg School of Medicine, Heidi and Pamela will be gathering information on how researchers are managing their data and what data storage and retrieval solutions are available to researchers, such as the Enterprise Data Warehouse at NUCATS. We will then look for ways in which Northwestern University libraries, including Galter, can assist researchers in dealing with their data.

If you are interested in being interviewed or sharing your thoughts on Galter Library's role in data management, please contact me using my email link at the end of this blog post.

Pamela Shaw
Biosciences & Bioinformatics Librarian
email Pamela

The Biosciences & Bioinformatics Blog highlights new tools and news items of interest to the biosciences research community at Northwestern University.