What Do Libraries Have to Do with e-Science?

An Interview with James L. Mullins, Dean of Purdue University Libraries

By Svetla Baykoucheva
Chemical Information Bulletin, Spring 2011, Vol. 63, No. 1, p. 45-49

James L. Mullins has been dean of Libraries and professor of library science at Purdue University since 2004. Before that he was associate director for administration of the Massachusetts Institute of Technology (MIT) Libraries. His more than thirty years long career includes administrative positions at Villanova University and Indiana University. He earned BA and MALS degrees from the University of Iowa and a PhD from Indiana University.

Dr. Mullins has served in leadership positions within the American Library Association (ALA) and the Association of Research Libraries (ARL) and presently is an elected member of the ARL board of directors and chair of the e-Science Working Group. Presently he serves on the editorial board of the journal College and Research Libraries. He is also on the board of directors of the International Association of Scientific and Technological University Libraries (IATUL), Center for Research Libraries (CRL), and a delegate to the Science and Technology Section of the International Federation of Library Associations (IFLA). Last June, Purdue was host to the 2010 IATUL Conference, which focused on the role of libraries in e-science. He was a signatory to the formation in December 2009 of DataCite, an international consortium assigning digital object identifiers (DOI) to datasets for citation.

Dr. Mullins is a frequent contributor to the professional literature, speaks at national and international conferences, and consults with research libraries and universities internationally on challenges facing research communication and dissemination. He has served on National Science Foundation (NSF) panels, including one in 2006 recommending that data management plans be required for NSF research funding.

Svetla Baykoucheva: *The new buzzword in academic libraries is "e-Science." It is also called "eScience." We are seeing job announcements for e-Science librarians, conferences on e-Science being organized, the Association of Research Libraries (ARL) publishing a white paper on it, and NSF introducing new requirements for data management. What is e-Science?*

James L. Mullins: In 1999, John Taylor, the Director General of the United Kingdom's Office of Science and Technology, created the term to describe computationally-intensive science that draws upon large data sets and, through modeling and algorithms, test assumptions. In today's world, scientists rarely use the term e-Science since computational methodologies have become so embedded in the research process that it hardly warrants distinctive nomenclature.

SB: *Last year you organized a conference on e-Science. What were the topics discussed at this conference? Could you point to some future conference on e-Science?*

JM: Purdue was host to the 31st Annual Conference of IATUL (International Association of Scientific and Technological University Libraries); the theme of the program was: "The Evolving World of e-Science: Impact and Implications for Science and Technology Libraries." The intent of the conference was to start with the broadest concept—what is e-science/computational science, what is the role of data in computational science and how are scientists coping (or not) with managing data? The keynote speaker was Dr. Dan Kleppner of MIT who co-chaired a task force for the National Academies on issues related to data. In addition, Dr. Arden Bement, who had stepped down as director of the NSF a few weeks before the conference, spoke about the interest the funding agencies have in ensuring that data generated through sponsored research would be available generally to researchers. Dr. Bement assumed the position of executive director of the Global Policy Research Institute at Purdue earlier that month, so his interest was twofold: the management of data and the need to create a global policy on data management to facilitate research. The majority of the program was focused on how data can be managed and what the role can or should be for librarians; so it wasn't just a theoretical discussion, as it provided an opportunity for librarians to gain knowledge of the processes that could assist them in developing e-science programs in their institutions. Rather than having me provide a complete summary of the program, it would be easy for readers who are interested in the topics to go to the website: http://blogs.lib.purdue.edu/iatul2010/program/.

There are many organizations that have a focus on e-science/data management within the international library community, especially the Digital Curation Centre (DCC) in the United Kingdom: http://www.dcc.ac.uk/events. In the United States, the Distributed Data Curation Center (D2C2) at Purdue is a research center focused on exploring and researching ways in which data can be accessed and archived. Further description can be gained at the link: http://d2c2.lib.purdue.edu/index.php. The Coalition for Networked Information (CNI) at its twice annual briefing sessions often has papers focused on e-science and data management. Also, on the CNI website (http://www.cni.org/regconfs/) there is a list of upcoming conferences and workshops that include ones on e-science/data management

Finally, the Association of Research Libraries (ARL) and the Digital Library Federation (DLF) are in the early stages of developing an e-science institute planned for fall, 2011. Initially the Institute will be open to sponsoring libraries (ARL/DLF members), but the intent is that it will be repeated for the broader community in 2012.

SB: *How do you see the role that librarians could play in this new area? What kind of expertise will be required from them?*

JM: Working in the area of data management draws upon the principles of library and archival sciences. Our ability to see structure to overlay on a mass of disparate "parts," as well as the ability to identify taxonomies to create a defined language for accessing and retrieving data is what is needed from us. The challenge will be for librarians to understand that we have collections that we cannot see and may not actually understand the importance of, but that we will have a responsibility to steward and preserve for researchers now and in the future. Archival science is important since there are requirements and expectations from investigators that there will be limited access to data that will require that an embargo be in place. Just as a person can give their personal papers to archives with an expectation that access will be limited to specific researchers or closed for a period of time, researchers may similarly want to protect their intellectual property by creating an embargo. For librarians this would be normally be unacceptable, while for archivists this is standard procedure. I also think it helps us to think about our present print archives as being raw bits of data, until a researcher (typically a humanist or social scientist) "mines" them to answer a research question, which is similar to a scientist or engineer consulting digital data in their research.

SB: *Will e-Science change the way academic libraries function? Will it change the infrastructure and the services libraries provide?*

JM: Many of our librarians (even those working in scientific and engineering disciplines) often have humanities or social sciences backgrounds. However, the trepidation that many librarians may have about sitting down with researchers and discussing their data management needs shouldn't be a controlling factor. Once a librarian has the experience of talking with researchers about their research and the challenges they have with managing data, it becomes clear that the most important factor is not our subject expertise (although some subject understanding is needed) but rather the librarian's knowledge of metadata and taxonomies. In the old days we would have said that this is "cataloging and classification," but today, to convey that we have morphed into a new role, it is best to use the more technical terminologies since it may help identify our "new" role as a cutting edge initiative and not be encumbered with past misperceptions. In fact, a few times I have seen researchers frustrated by librarians with significant subject expertise, who more or less intrude their subject knowledge into what the investigator is researching, while what investigators want is the library/archival science contribution to their team. We need to

remember that and be proud of the special expertise that we as librarians bring to the research team.

The impact for libraries in the broadest sense is the recognition that we have an important role to ensure the archiving and preservation of important data sets that initially may not be apparent to the researchers or us. We need to be able to think of treating these data sets as important collections, which is not that dissimilar to how we have stewarded our print book and serial collections or our archives. Responsibility for digital data brings new challenges and cost models—ones that we will need to work through with our university administrations and develop further collaboration with our colleagues in research administration and information technology.

SB: *What kind of problems do you see for librarians to be able to get involved in e-Science? Will faculty be willing to share raw data with outsiders and how could this potentially affect intellectual property rights?*

JM: I have touched on some of the problems for librarians to become involved with e-Science; so I will focus on the second part of your question. And the simple answer, from my perspective, is, "it depends." The one thing we have learned from the work we have done so far with disciplinary faculty and their research is that no two disciplines have identical policies or principles guiding them about sharing data. When we at Purdue embarked on this work six years ago, we thought it was going to be simple to help researchers manage and share their data. However, that naïve assumption was soon disproved. Some disciplines share data through a central database available to all, while others keep their data "close to the vest" while the research is being undertaken and are willing to share it only when it is needed to document findings in a published research article.

The mandate by the NSF and the likelihood this will be adopted by other funding agencies will trump, possibly, the traditions of data sharing (or not) within a field. It will take some time before it becomes an accepted, required step of the process. The NSF mandate is a start, but ultimately it will gain acceptance when researchers themselves begin to see benefits of sharing data beyond what they have done in the past.

SB: *How will e-Science affect the way research is performed and reported? What will be the consequences for the science and technology publishing field?*

JM: Some of the effects have been discussed above; so I won't go back over them here. But I will amplify some of the potential impact that may come from the availability of data and the requirements necessary to provide that access. During the past several years, the publishing industry has begun to assign digital object identifiers (DOIs) through the service provided by CrossRef. This has been very successful as it assigns a persistent identifier that will tag this article for retrieval, now and far into the future. The DOI serves somewhat like a barcode or ISBN, a unique tag that provides access to this article. So, with this ability to identify the article, there comes the concurrent need or desire to link relevant data to it. That initiative has been taken on by libraries around the world, through the development of the international organization called DataCite (http://datacite.org/). Its charge is to create a registry available to researchers throughout the world to permanently tag a data set, and provide enough description to allow for access and retrieval, if desired by a researcher. In the United States, the coordination and assignment of DOIs through DataCite is being undertaken by the California Digital Library (CDL), Purdue University Libraries and the Office of Science and Technology (OSTI) of the Department of Energy (DOE).

Creating DataCite and the assignment of DOIs is a major undertaking, not unlike what took place forty years ago with ISBN—the difference being that ISBN was a collaboration between publishers and national libraries, which had the reach and the clout to make it a standard in a short time and which were dealing with a finished product (a book). For DataCite, it is a few international libraries banding together to try to get this elephant headed in the right direction. At this time, the DOI assignment to a data set is not mandatory. There is a possibility, however, with OSTI recently joining DataCite, that the DOI assignment will become a requirement by funding agencies.

SB: *I have done many interviews for the Chemical Information Bulletin, but this is the first time I am interviewing a dean of libraries. And I would like to ask you a question that all academic librarians are asking: how do you see the academic libraries and the work librarians are doing change in the next few years? As dean of libraries in such prominent institution as Purdue, what changes are underway in your own libraries?*

JM: There is a shift from the trend that was happening ten years ago, which was the reduction of the number of librarians and other professionals and the increase in the number of clerical and student staff. In the "post print" world, the effort necessary to acquire, check-in, catalog, bind, and manage print collections has significantly been reduced. However, the work that needs to be done in collaboration with the faculty in the classroom and lab has increased.

At Purdue, librarians are full members of the professorial faculty, and with that comes an expectation that they not only ensure that the Purdue Libraries operate using sound library science principles, but that the latest initiatives be evaluated and integrated if deemed appropriate into the operations and services of the Libraries. However, in order to extend the work of the librarians, it is becoming clearer and clearer that we need to move much of the day-to-day management and such services as reference and cataloging/metadata operations to another tier of professional and clerical staff, trained and able to do these operations. This frees up the librarians to collaborate on information literacy instruction, research team collaboration, and research in the areas of changing scholarly communication models. If anyone came or is coming to librarianship thinking it would be a static, complacent, and quiet place to work, they may want to reconsider!

SB: *On a personal note, could you tell us about something that interests you besides information science and librarianship?*

JM: One of the great advantages of being a librarian is that we have the ability to explore so many aspects of knowledge and to follow the curiosity that I believe is an important trait that all librarians must have. Although I have a great love of travel and a commitment to international librarianship through participation in IFLA and IATUL, I don't consider that as my sideline interest, as it is still, for the most part, professional. I can give you an example of what I am reading for pleasure, pure enjoyment—and that is about the beginning of the Cold War, from the end of World War II and through the 1960's, into the Vietnam Era. Being a child during the 1950's, I remember so well our fear of the Chinese and the Soviets/Russians and the competition that was in place to out-achieve the Soviets in science and technology. We were aware that we could be destroyed any day by nuclear war, but as a child I really had no idea what the reason was. I remember watching as a boy in the 1950's an old WWII movie made during the War, where the sailors on an American ship began cheering when they realized that the planes they saw overhead were Russian and not Japanese. I remember asking my mother how could that be, and her answer was that they were our allies in the War. In the 1950s that seemed inconceivable. A little like today when we think of Iran. Therefore, I am reading about the beginning of the Cold War period and just finished an excellent book, The Lost Peace: Leadership in a Time of Horror and Hope, 1945-1953, by Robert Dallek.

SB: *It is an interesting coincidence that for this issue I also interviewed Dr. Michael Gordin, who has done extensive research on the beginning of the Cold War and has published books on that period. Thank you, Dean Mullins, for discussing e-Science and for your personal insights.*