D. Scott Brandt

# Librarians as partners in e-research

## Purdue University Libraries promote collaboration

The Association for Research Libraries recently organized an e-Science Task Force to investigate new paradigms for library engagement. Among other things, it is looking at new roles for librarians in an e-research environment, as well as new types of collaborations that can form from them. Purdue University Libraries' 2004 initiative to engage in multidisciplinary research has identified and begun to engage similar issues.

While the term *e-science* is sometimes used to refer specifically to the digitally enabled environments of science and engineering research, it is also synonymous with e-research, a broader net that encompasses all domains of research that are also challenged with exploring new ways of doing research in computationally enhanced and networked environments.

Long recognized as organizers, enrichers, and disseminators of information, librarians often have been pioneers in developing systems, process, and approaches to the delivery and use of information—from developing finding aids for archival materials, to cataloging and classifying print sources, to developing indexes and taxonomies for databases.

Given the highly distributed on-demand nature of the global environment that defines the Information Age, Purdue Libraries' initiative seeks to apply library science knowledge and expertise to multidisciplinary research to provide organization, enrichment, and dissemination of e-research. The goal of this multidisciplinary research initiative is to align with the university's strategic mission to develop preeminence in interdisciplinary research.

A particular topic demanding attention in many universities is the so-called "data deluge" problem. Issues with research data have been greatly highlighted in a variety of forums—from 2005's National Science Board report discussing the nature of long-lived massive data sets to the Microsoft 2020 report focusing, in part, on the need for a distributed environment of data. It is clear that as computationally assisted capabilities to create, process, and use data continue to grow, issues related to curating research data will likewise grow.

Since the arrival of Dean James L. Mullins in 2004, the Purdue Libraries have been actively promoting and seeking collaboration between librarians and disciplinary faculty in engineering, science, and technology on sponsored research projects. Through conversations, workshops, and seminars, it has become obvious that there is an urgent need for the knowledge that librarians have: the ability to collect, organize, describe, curate, archive, and disseminate data and information. When researchers were asked if they had data discovery, management, and organization needs, similar responses were repeated over and over again:

- not sure how or whether to share data,
- lack of time to organize data sets,
- need help describing data for discovery,
- want to find new ways to manage data, and
- need help archiving data sets/collections.

D. Scott Brandt is associate dean for research at Purdue University, e-mail: techman@purdue.edu

© 2007 D. Scott Brandt

The libraries are investing time to investigate and engage in opportunities to collaborate in e-research areas, where increasingly information-related needs have applications to library science. Purdue librarians are currently working on projects such as helping developing digital collections that link documents and data, enhancing distributed information systems and repositories, designing access via middleware to Web-based systems, and integrating information and technology literacy for end user education. From 2005 to 2006, 11 librarians participated in nine multidisciplinary proposals, including National Institutes of Health (NIH), National Science Foundation, and locally funded grants. It appears likely that more library science expertise can be leveraged to meet cross-, multi- and interdisciplinary research needs.

Purdue is not alone in investigating such problems—many libraries are engaged in related activities. From a library science perspective, curation is an aspect of organization that can help facilitate discovery, access, dissemination, and archiving of e-research. It is something that librarians or archivists have done for hundreds of years. In the Information Age, curation can be defined as essential activities and systems that facilitate access, dissemination, and archiving of e-research. At a practical level it can include protocols and tools that provide descriptive analysis of digital collections and objects to augment discovery, management, use, reuse, and preservation. These protocols can take the form of schemas to describe digital objects, systems to facilitate discovery of collections of objects, and middleware to resolve problems of interaction between systems and applications. On the one hand, curation is about policies and consultation, and on the other it is about tools and systems.

## Purdue's experience

Purdue University is heavily weighted toward engineering, science, and technology, and so initial forays into multidisciplinary research have centered there. For instance, the libraries are working with one collaborator in analyzing water quality data files straight off a sensor (a water logger that records flow, elements in the water, etc.) in the agronomy department. An NIH proposal for a biology resource required having a plan for archiving data, and librarians were asked to be part of the grant. An engineering project sought library science expertise in building metadata for an ontology that allows tracking of data through the workflow. Most recently, a research group working on space exploration came to the libraries to ask if they could "park their research" with the repository until new funding becomes available. They all view the libraries as "trusted" partners.

Several opportunities have resulted in technology-oriented projects, which are beginning to have an impact on how the Purdue Libraries operate. The first was conceptualization and proof of concept of a distributed institutional repository (DIR) framework that supports discovery and access to digital objects of e-research, including data and documents in various forms, formats and locations. The distributed framework interoperates with disparate information systems and repositories through an Open Archive Initiative (OAI) architecture, providing an institutional, regional, or disciplinary context for collecting, tracking, and disseminating e-research. The architect of this framework at Purdue, Michael Witt, is an example of a librarian who came out of the systems world and is a strong advocate for bridging technology and user needs.

The most visible result of the DIR so far is Purdue e-Scholar,[1] an umbrella service that includes several repositories: e-Pubs (a document repository), e-Archives (special collections repository), and soon, e-Data (a federation of data repositories). The e-Scholar Web portal is only one of many applications that have been built upon the DIR framework. Some applications include a metadata harvester and a Search and Retrieve via URL interface for searching across the e-Scholar repositories.

Another example of an application is OAISRB, middleware that allows metadata

from objects residing on a Storage Resource Broker (SRB) data grid to be harvested using a standard protocol, the OAI-PMH.[2] This enables metadata from research datasets from computational sciences to be included in more conventional digital library applications. The libraries are working with Purdue's Rosen Center for Computing and the Cyber Center on similar collaborations. Next, the libraries will have to determine how to integrate the repository framework with the other infrastructure of the library for uniform access and dependable sustainability.

## Getting started

A common question is, "How do you get started in these collaborations?" It helps that Purdue librarians are tenure track faculty and are seen as peers on campus. But when the new dean started, he set out on a campaign to talk to all deans and department and center heads on campus to tell them what library science, and thus libraries faculty, could do.

Generally the conversation started with the dean asking what goes on in the school, department, or center, and eventually the topic of data would come up. A simple inquiry into organization or archiving needs resulted in the understanding that librarians have been handling digital information and data for more than 20 years, and that library science had principles and practices that could help their needs. Since librarians started attending grant call-outs, workshops, and seminars and began mentioning some of the libraries' interactions and projects, they have been seen as potential collaborators on research.

In addition to a new role in multidisciplinary research, the libraries are involved in looking at scholarly communication in new ways, as well. A variety of factors have led to a "deconstruction" of traditional scholarly communication, and it appears that the role of librarians (library science) can contribute in an active and positive way, not only differently, but at different points in the process than previously understood or imagined. Whereas librarians once interacted with scholarly communication at the final or published end of the spectrum, it appears that interaction with alternatives to traditional publication or with prepublication points along the spectrum find librarians dealing not only with "born digital" material, but in some cases things that were not necessarily meant to be published, such as learning objects, departmental reports, and data. And even data has a spectrum in which librarians are investigating points of and levels of interaction—from working with raw to processed to analyzed to publishable data. Interestingly enough, if librarians can play a role in helping to organize and describe digital objects at the very early end of the e-research spectrum it will facilitate data mining in the future, which could truly facilitate new paradigms of research. Purdue Libraries are creating a data curation center to investigate and address such data needs and problems.

## Looking forward

As libraries and librarians move forward in these areas, there are still many questions to investigate. What is the role for librarians in facilitating scholarly communication? Some libraries, like those at Cornell and Pennsylvania State University, are using repositories and university presses to support and publish scholarly works. How involved should they be in the research process? Other libraries are partnering with research groups to explore this role, such as Johns Hopkins University. How do we train people to work in these realms? Library schools, such as the one at the University of Illinois, are engaged in and investigating the territory.

Some of these roles address new problems, but in many ways they do so using time tested methods of organization and sound library science practices. Chris Miller, Geographic Information Systems (GIS) specialist to the Purdue Libraries (whose undergraduate degree is in English and History), notes, "Any time one begins throwing around terms like *information* and *data* and *interdisciplinary*

*(continues on page 396)*

*in the Rye* and *To Kill A Mockingbird* in high schools and universities. The novel, which earned accolades from Eudora Welty and Walker Percy, won the Sue Kaufman Prize for First Fiction of the American Academy and Institute of Arts and Letters and the Louis D. Rubin Writing Award. The book's sequel, *The Life All Around Me, by Ellen Foster,* was released last year.

**Actor Harold "Hal" Gould's papers have** been acquired by the University at Albany's Libraries. Gould has been recognized with five Emmy nominations, an Obie, a Los Angeles Drama Critics' Circle Award, and an ACE Cable TV Award for his varied roles. Gould's career includes Broadway roles in such plays as John Guare's *House of the Blue Leaves,* Neil Simon's *Fools,* and Jules Feiffer's *Grown Ups.* He is most widely known for his roles on *The Mary Tyler Moore Show, Rhoda,* and *The Golden Girls.* He also had roles in films such as *The Sting* (1973) through *Freaky Friday* (2003), and *English as a Second Language* (2005).

**The architectural archive of Pierre Koenig,** the internationally celebrated architect whose work helped to define modern architecture, has been acquired by the Special Collections of the Research Library at the Getty Research Institute (GRI). This archive, containing more than 3,000 objects, including drawings, models, photographs, slides, and documents, will enable scholars to study a significant chapter in post-war American domestic architecture. Among the architectural gems documented in the collection are Koenig's Case Study Houses #21 and #22, which were both executed as part of Case Study House Program of 1945–1963 for John Entenza's *Arts & Architecture* magazine. Koenig was one of the youngest architects included in the program, which promoted modern, indoor-outdoor California living through innovative steel-frame design and construction. Born in San Francisco in 1925, Koenig became interested in the structural possibilities and advantages of steel residential construction, while a student at the University of Southern California's School of Architecture in the 1950s. Confronted with the skepticism of his professors, who questioned the applicability of steel to residential architecture, Koenig proved them wrong by designing and building his own steel home at a cost lower than that of a traditional wood frame structure. This innovative structure earned him the American Institute of Architects' House and Home Award of Merit. Upon graduation, he opened his own architectural practice in Los Angeles. Throughout his career, which spanned five decades, he never relinquished his goal of producing prefabricated homes for the masses. Koenig died in 2004.

---

*("Librarians as partners...," continued from page 367)*

and *access,* one is more often than not talking about libraries.

"GIS is necessarily, intrinsically interdisciplinary, as are library systems. I'm proud of the fact that having a librarian's universal, abstract perspective on the methods and modes and behaviors of information allows me to see much more clearly how GIS can act as a dumb (but very, very intelligent) meeting place for information and data from disparate sources."

As Purdue Libraries embark on this new initiative to investigate issues and problems of applying library science knowledge and expertise to multidisciplinary research, it intends to share findings and insights with the library community so that others may benefit from and build upon these projects.

## Notes

1. Purdue e-Scholar, e-scholar.lib.purdue.edu.
2. See OAI-PMH www.lib.purdue.edu/research/oaisrb/ for a description of this application.