1-1-2010

E-science, Cyberinfrastructure and the Changing Face of Scholarship: Organizing for New Models of Research Support at the Purdue University Libraries

Jake R. Carlson
Purdue University

Jeremy R. Garritano
Purdue University, jgarrita@purdue.edu

E-science, Cyberinfrastructure and the Changing Face of Scholarship: Organizing for New Models of Research Support at the Purdue University Libraries

Jake R. Carlson*

Data Research Scientist

504 W. State Street

Purdue University Libraries &

Distributed Data Curation Center

West Lafayette, IN 47907

jrcarlso@purdue.edu

Jeremy R. Garritano

Acting Head, Chemistry Library

Chemical Information Specialist

Assistant Professor of Library Science

504 W. State Street

Purdue University Libraries – CHEM

West Lafayette, IN 47907

jgarrita@purdue.edu

* To whom correspondence should be addressed.

Introduction

A revolution in scientific research is being driven by the proliferation and wider availability of high performance computing, the development of visualization, simulation and other sophisticated analysis tools, and the increasing capacity to store massive amounts of data. As a result, science is beginning to shift away from traditional experiment-based practices and towards computationally driven models of research, in which massive data sets are used to test hypotheses. These developments in scientific practice, collectively known as e-science, are leading to massive changes in how science is conducted. Under e-science, questions that were once relegated to being purely theoretical such as:

- What impact does species gene flow have on an ecological community?
- What happens to space-time when two black holes collide?
- What are the key factors driving climate change?

are now within the ability of researchers to explore and answer using cyberinfrastructure.[1]

According to the seminal 2003 report to the National Science Foundation (also known as the "Atkins Report"), cyberinfrastructure consists of the multiple layers of distributed computer, information and communication technologies upon which e-science models and practices are built. These layers of cyberinfrastructure house the information, data, storage, standards, personnel, policies, tools, services, and social practices that enable e-science to function. As stated in the Atkins Report, "If *infrastructure* is required for an industrial economy, then we could say that *cyberinfrastructure* is required for a knowledge economy."[2]

An example of the potential of cyberinfrastructure to revolutionize how science is practiced is the National Virtual Observatory (NVO). The NVO provides a centralized portal for discovery and access to distributed catalogs of astronomical data gathered from telescopes all over the world. Rather than having to secure access to a high-powered telescope and schedule a limited amount of time to obtain an individual data set, professional and amateur astronomers

alike can now easily discover and access high-quality catalogs of data sets at their convenience through the NVO. Although astronomy data collections are gathered and stored at installations across the world, the NVO is able to bring these data sets together in a centralized interface through its use of standard protocols for registering the existence and location of data with NVO and the use of community based metadata standards to enable the interoperability of the registered data sets. The ready availability of a multitude of interoperable data sets enables astronomers to search for data that match a particular set of criteria and analyze patterns and to better study the complexity of astronomical systems in ways that simply were not previously possible. The NVO also offers a software library and a suite of statistical, visualization and other tools for the analysis of astronomical data. The NVO promotes community involvement in the development of its services and resources by enabling astronomers to publish their own data or tools through the NVO.[3]

E-science is expanding into a broader paradigm of e-scholarship as other academic disciplines are beginning to create their own cyberinfrastructures for their fields of practice. For example, applying the tools and capabilities of cyberinfrastructure to research in the Liberal Arts has the potential to change how scholars make sense of the human record.[4] This potential is illustrated in *Rome Reborn* , a digital 3-D model of the city of Rome as it existed in late antiquity. Produced by the Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia and its collaborators, the primary purpose of *Rome Reborn* is to communicate the current understanding of the urban topography, infrastructures and individual structures of ancient Rome. In addition to providing a means to address theories of how the city may have looked, IATH and its partners envision *Rome Reborn* as a means to explore questions

that would otherwise be difficult or impossible to address such as how the design of the city and its buildings affected ventilation, illumination or the movement of people.[5]

This chapter will briefly highlight two key components of cyberinfrastructure as identified by the National Science Foundation (NSF) – data curation and preservation, and interdisciplinary research and virtual organizations – and address the challenges and opportunities they pose for academic libraries. A review of new ideas for library organizational structures and staffing models to meet the changing needs of library users in the digital age will follow. Next, this chapter will describe the approaches taken by the Purdue University Libraries to adopt aspects of these ideas and put them into practice to enable librarians to become directly involved in the development of cyberinfrastructure and to provide support for e-science research. The Purdue model includes creating a new organizational structure within the Libraries to support librarians active engagement in working directly with faculty to research and develop solutions to address the problems of data management, organization, dissemination and preservation, and creating new positions to coordinate the libraries research efforts and to help leverage the existing skills and relationships of subject librarians. Finally, a case study describing a collaboration between two librarians, the information technology department, and faculty in the Chemistry and Food Sciences departments will be presented to illustrate the Purdue Libraries' model of engagement with faculty.

Data Curation and Preservation

Under the paradigm of e-science, the scientific method shifts from "hypothesize, design and run experiment, analyze results" to "hypothesize, look up answer in a database."[6] In this environment, data are the lifeblood of scientific practice, and access to data becomes as important to the scientist as was the microscope to a traditional laboratory. The problem is that

researchers' capability to generate or manipulate data through e-science experiments has far surpassed their ability to manage, organize, or make their data easily accessible. E-science experiments both generate and require massive amounts of data, and the rate of growth in data production is expected to increase as better, faster, and smarter technologies are developed and deployed. The physical and social infrastructures that are needed to manage this "data deluge" have not developed at the same pace as the ability of researchers to produce it.[7] This disparity between the growth of production capability and the lack of tools, infrastructures, workflow systems and collaboration to address distributed and complex data sets is recognized as a barrier to realizing the full potential of e-science.[8]

In its 2007 cyberinfrastructure report, the NSF articulated a vision of the future "in which science and engineering digital data are routinely deposited in well-documented form, are regularly and easily consulted and analyzed by specialists and non-specialists alike, are openly accessible while suitably protected, and are reliably preserved."[9] Enacting this vision will require much more than just collecting it and putting it on a Web server for access or on a backup hard drive for storage. Data have to be curated and preserved to retain their value and to remain accessible over the long term. Data curation is defined as: "the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and re-use."[10] Data that are not curated will become irretrievable or indecipherable and thus lose their value and utility as an information resource.

However, the curation and preservation of data present an array of difficult challenges. For example, research data come in all different types, formats, and sizes, even within individual academic disciplines. The heterogeneity of research data, as well as the diversity of the social and cultural practices of researchers producing the data, means that "one size fits all" solutions

will not work.[11] This situation is further exacerbated by the lack of agreement (or even awareness) in many communities on the standards, practices and technologies to employ for curation and preservation work. The infrastructure supporting data curation and preservation activities must be flexible, scaleable and extensible to accommodate current and future needs. Data sets need to be accompanied by appropriate metadata to provide the context for understanding and using the data. Metadata are required so data sets can be discovered, preserved, administered, and made interoperable with one another. Appropriate and fair intellectual property rights for the owners of research data sets and fair-use exceptions for would-be users will need to be identified, conveyed and enforced. The preservation of data requires an ongoing commitment of resources, which, in turn, mandates the need for viable economic and technology sustainability plans.[12]

Interdisciplinary Research and Virtual Organizations

The scope of the questions being addressed by e-science goes far beyond what could be addressed in a single academic discipline or even by a single institution. Research performed under the e-science paradigm typically requires the formation of interdisciplinary research teams composed of researchers from multiple disciplines and who are distributed across multiple types of institutions and locations. In addition, the complexities of this new research environment and the amount of support needed to develop and maintain the necessary cyberinfrastructure mandate close collaboration between researchers and computer scientists, software developers, network engineers, data managers, and other IT providers. Other types of professionals possessing an array of diverse skills and abilities will be needed to contribute to an environment in which e-science can flourish. Librarians and archivists, for example, have been identified as having the knowledge necessary to potentially help researchers address data curation and preservation

issues.[13] Economists could help craft economic sustainability plans for supporting cyberinfrastructure. Copyright attorneys may be needed to sort out intellectual property issues over the ownership and fair use dissemination of digital objects.

The high performance networks, advanced data storage capacities, distributed computational tools and other components that make up cyberinfrastructure enable the formation of virtual organizations to carry out e-science projects. Cyberinfrastructure capabilities have grown to the point where it is now possible for these virtual organizations to carry out all of their work in an online environment. Through shared access to online tools, services, and data, research teams can share information, design experiments, operate scientific instruments remotely, run simulations, and analyze or visualize data in order to conduct their experiments and work together as a cohesive unit. The resources and services used by virtual organizations can be distributed across multiple locations and made centrally accessible in real time reducing the burden of providing support for any single institution. The creation of viable virtual organizations eliminates the barriers of geography enabling the creation of cross institutional teams of researchers in the United States and internationally.[14]

Supporting interdisciplinary research and collaborative cyberinfrastructure is a significant challenge to universities as academic culture and structures are largely geared towards supporting and rewarding the work of individuals. The traditional “cottage industry” approach to research, in which research is driven by individuals, resources are obtained for singular purposes and deployed locally in the department or lab runs counter to the goals of e-science: to make cyberinfrastructure and the data, tools and other resources available to a wide audience for multiple research and educational purposes. The lack of an overarching approach to cyberinfrastructure makes it difficult for universities to integrate technology resources and open

them up to broader access in order to achieve necessary economies of scale and to make the best use of scarce resources.[15] Changing the current environment and realizing the full potential of a sustainable cyberinfrastructure will require innovative thinking, creative approaches, and new synergies within and between academic units and institutions.

Challenges for Libraries

The changes in how research is done under the e-science paradigm will have an effect on how the library carries out its mission of supporting the research and information needs of the university. The nature of scholarly communication, for example, is already undergoing dramatic change in response to technological advances, and the spread of e-science research models will only accelerate the pace of these changes. Research data are already shifting from being a disposable by-product of research into an important outcome, if not the most important outcome, of the experiment.[16] To reflect this shift, some have suggested that scientific publications may transform into databases themselves. The data from these publications would be peer reviewed and available for researchers to re-use or repurpose in order to create new science.[17] Moreover, researchers participating in e-science will increasingly need more than static sources of information in the form of books and journals to satisfy their information needs. As the effects of "Web 2.0" continue to influence scholarly communication and the sharing of information, libraries will be increasing called upon to capture, curate and preserve dynamic streams of raw, loosely-structured communication streams.[18]

Beyond simply reacting to the changes in research brought about through e-science libraries have an opportunity to become actively involved in developing cyberinfrastructure and in addressing the issues and challenges of e-science. Libraries have already realized the need to protect the investments they have made in building digital collections and to ensure their

availability for the long term by actively seeking solutions to the many conundrums that surround the preservation of digital materials [19] Many within the library community are increasingly interested in developing roles for librarians in curating and preserving the digital data generated by research faculty, believing that librarians possess abilities and expertise that will be needed in this area. Librarians bring a long-term perspective to their work and understand the value of maintaining primary documents for the historical record as well as the planning and effort preservation requires. Further, librarians are experienced not only in navigating complex information environments, but in understanding the architecture of these environments and how they connect to the needs of research communities. The function of the librarian has always centered on creating, maintaining and employing logical systems of organizing and describing information in order to enable its discovery and retrieval for the appropriate audiences. Librarians have long recognized the need for standards to manage heterogeneous sources of information and the benefits of commonly accepted and uniform terminologies, rules, and structures in disseminating information. Finally, librarians have already invested resources to digitize materials and house, preserve and disseminate digital collections of materials in institutional repositories and developed the expertise needed to manage these repositories.[20]

Recognizing the potential transformative impacts that e-science may have on the role of research libraries, the Association of Research Libraries (ARL) formed a task force on library support for e-science with a mandate to shape an agenda for developing e-science capacity in libraries. Its 2007 report articulated several focal areas for libraries to explore: data and new forms of scholarly communication, support for virtual organizations, and policy development. The outcomes recommended by this task force included an increased level of interaction with e-science communities and a greater understanding of how libraries can contribute to the continued

development and deployment of cyberinfrastructure and e-science. At the organizational and staffing level, the ARL report advocated the development of knowledgeable and skilled library professionals who could undertake new roles to help libraries conceive and implement new services and resources in support of e-science. The ARL report also recommended enabling "research libraries [to be] active participants in the conceptualization and development of research infrastructure, including systems and services to support the process of research and the full life cycle of research assets."[21] To achieve these outcomes, libraries will need to adopt and support new types of organizational and staffing models that encourage innovation and risk taking by librarians and enable librarians to explore possibilities for applying their knowledge and skills outside of the traditional boundaries of the library.

In conjunction with NSF, ARL also held a workshop to examine the role of libraries and other partners in the stewardship of science and engineering research data. The report that followed urged NSF to "facilitate the establishment of a sustainable framework for the long term stewardship of data."[22] This framework would include: support for the training and education of a new workforce in data science; support for research and development in understanding what is needed to curate and preserve data for the long term effectively; and support for librarians working on data curation and preservation issues as members of research teams.[23]

Although librarians may be seen as a natural and logical partner in addressing the challenges of e-science, there are many significant roadblocks to overcome if librarians are to play a meaningful role. Some of the larger challenges include:

- existing library systems and infrastructures are primarily set up to support a text-based environment and will not translate easily, if at all, to managing and supporting research data;
- a steep learning curve to gaining the level of fluency needed to understand e-science issues and practices;

- librarians generally lack the training and background needed to understand the technical issues involved in data management, curation and preservation;[24]
- recent assessments of researchers indicate that libraries are generally not viewed as being connected to the research infrastructure needed to support interdisciplinary research;[25]
- libraries are institution centered, whereas most scientific data repositories are subject based and designed to serve interdisciplinary audiences;[26]
- it is likely that efforts in data curation and stewardship will require libraries to make significant investments in infrastructure and people.

Rethinking the Organizational Structure and Staffing Models of Libraries

Just as the traditional organizational structures and the conservative culture of academia pose barriers to supporting e-science research, the traditional organizational structures and culture of academic libraries pose barriers to the library becoming more actively involved in building cyberinfrastructure and supporting e-science. However, librarians have recognized that the organization and staffing models of libraries need to be rethought and adjusted to become more in synch with developments in research, teaching and learning in the digital age. Susan Gibbons has written about the need to cultivate an "R&D mind-set" in librarians and to foster an environment in libraries that enables and rewards innovation. Gibbons defines an "R&D mind-set" in libraries as a culture in which all staff are expected to stay connected with developments not only within academic libraries, but in higher education, technology, management and other fields in order to bring in new ideas from these different perspectives to the library. In this culture the exploration and development of new ideas is actively encouraged by the library administration, and staff are provided with reasonable amounts of time, resources and support to experiment without being stigmatized if their ideas do not pan out. Gibbons believes that a library fostering an "R&D mind-set" will be more agile, flexible and able to respond more effectively to change than a library with a more traditional mind-set.[27]

Wendy Pradt Lougee, chair of the ARL e-science task force, has written about the evolution of library roles and foresees libraries and librarians becoming more "diffused" through working directly as collaborators with stakeholders within and beyond their home institutions. Lougee identifies four key shifts in libraries:

- from emphasizing the value of collections to emphasizing the value of librarian's expertise;
- from supporting information description and access to taking responsibility for greater information analysis;
- from serving as a support agency to serving as a collaborator; and
- from a facility-based enterprise to a campus-wide enterprise.

These shifts mark a path of development that focuses on adopting distributed and open models that enable libraries to take on more varied roles. In the final phase of this evolutionary path, the "diffuse library" has become both more broadly and more deeply ingrained not only in the dissemination of knowledge, but in its creation as well. In addition to its traditional role of collecting research outputs, the professionals housed in, or associated with, the diffuse library work alongside knowledge producers as a part of the research process. Traditional library functions and roles such as collection development and information access are expanded outward to form new paradigms such as the library as a publisher of information, or using metadata to enable new access strategies and techniques for content being used or developed by research communities.[28]

As academic libraries reconsider their organizational systems to better meet the needs of faculty, staff and students in the digital age, they are also reconceptualizing staffing models for the library and reexamining the skill sets needed by library personnel. James Neal has written about the rise of "feral professionals" in libraries. Neal defines two types of feral professionals. The first type are individuals without a Master's degree in Library Science but who possess skill

sets, experience, or other credentials that allow them to effectively assume traditional library positions, such as subject specialist or cataloger. One example might be a PhD chemist who migrated to an information center within industry and then was hired as an academic chemistry librarian. The second type are professionally trained librarians, that have positions or responsibilities that lie outside of the traditional library core, such as those related to fundraising, publishing, or instructional technology. These types of positions are designed to allow libraries to offer additional services, increase the library's capabilities, or to move the library in new directions. These professionals are considered "feral" not only due to the non-traditional nature of their backgrounds or positions, but because they often carry a different set of values, expectations, outlooks and opinions than a more traditional librarian does.[29] Although "feral professionals" have challenged the status quo of libraries, their non-traditional expertise have helped libraries become more in touch with the core academic mission of the university.[30]

A second new staffing model at academic libraries has been conceptualized by Steven Bell and John Shank. Bell and Shank have observed that the use of information technologies is a disruptive force in the support and delivery of instruction at colleges and universities, changing the way faculty teach and students learn. They note that librarians have not kept pace with developments in information technology and are at risk of being overshadowed or eclipsed as the adoption of information technologies in the classroom and curriculum continues to gain traction and acceptance. Bell and Shank propose the role of "blended librarian" as a remedy. A blended librarian is one who combines the traditional skill set of a librarian with an information technologist's skill set in hardware and software along with an instructional designer's knowledge of pedagogy to understand when and how to apply technology in the teaching and learning process appropriately. The combination of these skill sets places the blended librarian,

and the library as a whole, in a better position to offer the kinds of services and resources needed in a new age of technology enabled teaching and learning.[31]

A third innovative staffing model is centered on librarians working outside of the library setting and more directly with faculty in the classroom or "in the field" to support the faculty's teaching or research. Several university libraries have been experimenting with this type of model. The Community College of Vermont and the University of Rhode Island have embedded librarians into online courses to enable them to connect more directly with students and better address their information needs.[32] Virginia Tech has implemented a college librarian program in which librarians are placed outside of the centralized university library and distributed within the colleges they serve, thus allowing them to be more immersed in all aspects of a researcher's or student's information needs.[33] Building on the Virginia Tech model, the University of Michigan has launched a field librarian program in which the librarian is not only housed with faculty researchers but offers technology skills in addition to subject domain expertise.[34] The underlying goals of these "embedded librarian" programs is for librarians to immerse themselves in the department's environment and to better understand and respond to the individual faculty members needs. Ideally deeper understanding enables librarians to build trust relationships with faculty and to become more of an active partner in faculty research and teaching.

The Purdue University Libraries' Approach to E-science

Dr. James Mullins became the Dean of the Purdue University Libraries in 2004, in the midst of major efforts launched by Purdue to actively respond to the changes brought about by e-science and to adapt the university to accommodate the development and expansion of cyberinfrastructure. Similar to most other research universities, e-science figured prominently in

Purdue's strategic for 2001-2006. The characteristics of Purdue's vision and goals under this plan included:

- Model interdisciplinary and collaborative partnerships in the university community.
- [An] active role of all disciplines in contributing their disciplinary and interdisciplinary strengths to Purdue's vision.
- Collaboration with public and private enterprise in Indiana, the United States and abroad as a model for pursuing common objectives.
- A stimulating and supportive state-of-the-art infrastructure that includes informational, technical, facility and human resources.[35]

Again like many research universities, Purdue has put the e-science based conceptual goals articulated in its strategic plan into practice. Purdue's notable accomplishments in supporting e-sceince have been the creation of Discovery Park and the NanoHUB. Discovery Park was created at Purdue in 2001 and serves as a home to Purdue's interdisciplinary research centers through providing space for researchers from different disciplines to work together, and centralized access to resources including cutting-edge scientific instruments and facilities.[36] The NanoHUB is designed to support a virtual community in the interdisciplinary field of nanotechnology. This is accomplished through NanoHUB's online gateway which provides access to simulation and other computational tools, research materials, resources for teaching and learning, and virtual meeting space and other communication tools. The NanoHUB is designed to be used by a wide audience of researchers, teachers and students. The complexities of the cyberinfrastructure behind the NanoHUB are hidden from the user and all of the tools and resources can be accessed using nothing more than a Web browser.[37]

As part of Purdue's research infrastructure, the Libraries were given a mandate from the president to define its roles in supporting interdisciplinary research and enabling collaborative partnerships. Early on Dean Mullins met with almost every department head on campus to better understand the research needs of Purdue faculty. What he heard from these meetings echoed the

challenges reported in the literature. Common refrains from faculty were that they are having difficulty managing, organizing, sharing, and archiving their research data and that they lacked the time and the skills to address these problems. Many faculty realized that their data has value beyond their original purpose and expressed a desire to share their data with others, but were not sure how to share their data effectively. In addition, funding agencies were beginning to talk about requiring data management plans to ensure the availability of the data to others outside of the project after a reasonable amount of time and beyond the duration of the project's funding cycle. More than just making data sets available, funding agencies would expect the data to be organized and described well enough so that these data could be mined or repurposed by other researchers. Although faculty were aware that these requirements are coming they were not sure how to prepare for them.[38]

Dean Mullins saw an opportunity for the Purdue University Libraries to play a direct role in the University's push towards interdisciplinary research by working with faculty to meet the challenges of e-science, focusing on data curation and preservation in particular. He quickly recognized that this endeavor would require transforming the Libraries' traditional roles, resources and services to better suit the information needs of researchers in this new era.

The organizational changes made by the Purdue University Libraries to support librarian involvement in interdisciplinary research have been based upon several assumptions made by the Libraries' administration. These assumptions have been articulated in presentations made by Dean Mullins and the Libraries' Associate Dean of Research.[39] The first set of assumptions reflect one of the recommendations of the ARL e-science task force: *in order to be effective in addressing data curation and preservation issues librarians will need to reexamine the role they play in the research process and modify their practices accordingly*. Working with faculty and

IT professionals to develop effective means for the curation and preservation of data requires that librarians be intimately involved with the project early on in the research process, ideally at its conception.[40] This level of involvement would mean that librarians would need to be fully embedded within the research project, similar to the way that librarians are now becoming embedded within courses and academic departments. The librarian embedded within a research project would generate ideas, solicit partners to work with, participate in the grant application process, conduct the research, and write up and report out the results just as any other researcher would do.

The second set of assumptions is that librarians would be accepted as research partners in developing solutions to address data curation and preservation issues by faculty researchers and information technology professionals. The barriers to data curation and preservation identified in the literature: heterogeneity, lack of standards, the need for metadata, copyright, etc., pose serious challenges to researchers as they generally do not have expertise in data management, nor do they typically have many resources for effective data management at their disposal. Furthermore, most researchers would rather focus their energies on conducting their research than on managing their data. Information technologists are needed to build the technical systems and infrastructure to support data curation and preservation; however they are likely to be more interested in the technology aspects of the problem, often overlooking the data management and data description tasks necessary to enable discovery, access, or preservation. Under this set of assumptions, the goals and objectives of the researcher for their data would be difficult to achieve without collaborating with librarians. Thus, librarians would be viewed as problem solvers with applicable and needed skills by faculty researchers and information technology professionals.

A third set of assumptions is that, as no readily apparent solutions to the data deluge exist at this time, resolving these issues will call for a substantial investment in time, effort, and infrastructure, beyond what libraries and other parties could reasonably provide on their own. Enabling these kinds of investments requires resources, specifically the availability of funding through grants or other sources. Government agencies and other funding organizations have already expressed a desire to protect the investments they are making in research by ensuring that research outputs are accessible to others and sustainable for the long term. For example, the 2003 statement on data sharing from the National Institute of Health affirms its support of sharing research data and puts forth an expectation that researchers requesting more than $500,000 in a single year will design a plan for sharing their data.[41] The National Science Foundation's (NSF) 2005 grant policy manual also states that grantees are expected and encouraged to share their primary data and other supporting materials with other researchers.[42] In late 2007, the Office of Cyberinfrastructure within the NSF released a solicitation to fund the creation of organizations dedicated to developing and sustaining new methods, technologies, and management structures to address the challenges of the data deluge. These organizations, dubbed "DataNet partners" in the solicitation, will be driven by the needs of researchers but will include representatives from the library and archival science communities as well as the cyberinfrastructure, computer science and information science fields.[43] Therefore, it seems likely that the NSF, NIH and other agencies will continue to offer support and funding for efforts to address data management, curation and preservation issues as a component of their solicitations, or as solicitations in and of themselves.

Finally, a fourth set of assumptions is that the Purdue Libraries' involvement in developing solutions in data curation and preservation will affect many, if not all, librarians at

Purdue. New library professionals willing to help guide and lead the Libraries' efforts in data curation and preservation will be needed. However, the scope of what is required for the Libraries to support data curation and preservation effectively will go beyond what is possible for individuals in data curation positions to provide on their own. An effective program in data curation and preservation will require multiple skill sets and the involvement of many different librarians. For example, public service librarians could conduct reference interviews to uncover the data related needs of the researcher. Librarians with collections expertise could serve as data stewards if the library agrees to take responsibility for data sets. The expertise of technical services or metadata librarians could be brought in to identify appropriate taxonomies or ontologies that could be employed to meet the needs of the researcher. Librarians supporting technology infrastructures or services could be able to design and implement data repositories to provide appropriate discovery, access and preservation functions.[44] These librarians, while continuing to provide traditional library services, will need to expand their capabilities, knowledge and skill sets with data curation and preservation to become a different kind of "blended librarian," one with a knowledge and understanding of the production and use of research data instead of instructional design.

The Reorganization of the Purdue Libraries

Working from these assumptions, the Purdue Libraries began reorganizing to create a "diffuse library" with an "R&D mind-set" in which librarians would collaborate with faculty and others through involvement in interdisciplinary research projects, and would apply their skills as librarians in new ways to address data curation and other issues in e-science. Realizing that someone would need to take ownership and lead the charge to investigate and assess the Libraries' involvement in campus research, Dean Mullins approached D. Scott Brandt, then

Technology Training Librarian, to take on this role. In 2004, Brandt became the first Interdisciplinary Research Librarian.

It quickly became apparent that the library needed to realign its organizational structure to more closely resemble that of an academic department so as to gain recognition and acceptance as a legitimate research organization. Most Colleges at Purdue have an Associate Dean of Research who fosters and facilitates research initiatives, and serves as liaison to the Office of the Vice-President for Research (OVPR). Many important decisions are made at the OVPR's Strategic Research Initiatives group meetings and so it was important for the Libraries to have representation on this group. Attendance at group meetings is limited to Associate Deans and the heads of research centers, however, and so Dean Mullins proposed the creation of a formal research department within the Libraries which would be headed by an Associate Dean of Research. The Associate Dean position would coordinate the Libraries' research initiatives and give the Libraries access to people and resources, such as membership on the Strategic Research Initiatives group, that would otherwise be difficult to obtain. The Libraries' proposal for the creation of a research department was approved by the provost in November 2005 and Brandt was appointed as the first Associate Dean of Research.

The mission of the Libraries Research Department is to support the Libraries' research initiative in applying the knowledge and expertise of librarians to provide organization, enrichment, and dissemination of e-science.[45] This mission is currently accomplished in several different ways. First, research department faculty and staff help librarians in more traditional library roles to identify and resolve access, organization, dissemination and other issues in data curation and e-science. Second, the research department sponsors activities that support and promote research, such as hosting informational and brainstorming sessions on research or

funding opportunities for librarians. Third, the research department oversees the investigation of the Purdue Libraries' Distributed Institutional Repository (DIR) system. The DIR system will consist of multiple repositories that will enable the dissemination of research at Purdue as well as its curation, storage, and preservation as appropriate for the nature of the content, available infrastructure and the needs of the users. A hallmark of the DIR will be the federation and interoperability of metadata between these repositories to enable cross disciplinary discovery and research. The DIR consists of three interconnected repositories: one for publications, one for archives and special collections, and one for research data (in development). Finally, the research department seeks out grants and other opportunities for the Libraries to obtain the funding necessary to pursue their research objectives and assists librarians in navigating the process of applying for these grants. Research is also supported by the Libraries Research Council, which is charged with addressing issues and policies related to the Libraries' research activities.

In 2006, the Libraries created the Distributed Data Curation Center (D2C2) at Purdue as another means to further the diffusion of the Libraries into the research culture and infrastructure of Purdue. The goals of the D2C2 are to enable the Libraries to align itself with the university's strategic mission and to make connections with researchers in ways that they and funding agencies would understand and accept.[46] Faculty and administrators are not used to working with the Libraries as research partners; but are used to collaborating with research centers, so creating a research center has given the library greater access to resources and key individuals throughout the Purdue community and elsewhere. The board of directors for the D2C2 is comprised of several Deans, a director of a research center at Purdue, Purdue's Chief Information Officer and others. These board members enable the D2C2, and by extension the Purdue Libraries, to stay abreast of important developments within the University.[47]

The D2C2 seeks to address the challenges presented by the increasingly data-intensive and highly networked nature of academic research. Its mission is "to investigate and resolve curation issues of facilitating access to, preservation of, and archiving for data and data sets in complex and distributed environments" through the application of the practice and principles of library science.[48] This mission is primarily accomplished through working directly with researchers and embedding librarians within their research projects to add value to their data sets. The D2C2 defines data sets broadly to include a wide range of digital collections or objects. Adding value to data sets may include such tasks as helping to identify and apply appropriate metadata standards and ontologies to data sets, designing models and workflows to better enable the access, curation, and preservation of data, and ensuring that data are available for the long term by capturing provenance information or stewarding data in the Libraries' e-data repository. The D2C2 also pursues its own teaching and research agenda which includes addressing the higher scope of data curation through the development of best practices, teaching others about data curation theory and application, and building community based curation profiles and policies to define appropriate access to and use of data.

New Staffing Models at the Purdue Libraries

The reorganizations at Purdue, the new strategic plan, and the "diffusion" of the library across campus through the creation of the Libraries' research department and the D2C2, have all been undertaken with an overarching purpose in mind: to enable librarians, not only to adopt an "R&D mind-set," but to think of themselves as academic researchers in their own right. Purdue's approach has been to assert that librarians have what Dean Mullins refers to as "a natural research domain" in the management, organization, dissemination and preservation of information. However, as is often the case, many librarians at Purdue are focused on the

operations of a library and have only limited experience in applying their skills on projects or research outside of a library setting. Working to design and build solutions to address problems in data curation and preservation, and partnering with faculty in their research projects directly is not only a major paradigm shift to the organizational models of libraries but a significant change in how librarians view themselves and their place in the university. Therefore, enabling librarians to become involved in interdisciplinary research requires more than reorganizing the structure of the Libraries, it requires librarians, along with faculty and others outside of the profession, to reconceptualize the role of a librarian.

The Purdue University Libraries approach to staffing its research initiative is two-fold. First is to adopt the "feral professional" approach by bringing new personnel into the library to focus directly on building the Libraries' capacity and expertise in researching data curation and related areas to support the new paradigm of e-science based research. Second is to adopt the "blended librarian" approach through the reapplication of the specialized skill sets of traditional librarians at Purdue towards providing the new services and resources that are needed to meet the information needs of researchers practicing e-science. This approach entails a not only a role for librarians in assisting in the Libraries' research efforts, but operational roles as well. One of the desired outcomes of this approach is to develop the data management and curation skills of librarians to the point where they become a natural part of the core skill set of librarians at Purdue. Ultimately, it is envisioned that librarians will build and maintain collections of digital research data sets as they would build and maintain a collection of electronic books or journals.

The Data Research Scientist Position

The Data Research Scientist is a new position created by the Purdue Libraries to help drive its efforts in developing the Libraries' capabilities and services in data curation and

preservation. The idea for this position came out of a report by the National Science Board (NSB): "Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century." This report describes the roles and responsibilities of individuals in curating research data including a position they labeled "data scientist." The role of the data scientist differs from the role of the data manager, as the data manager would be directly responsible for the operation and maintenance of databases. The role of the data manager roughly corresponds to the role envisioned for Purdue's Librarians as stewards of collections of data sets. In contrast, the role of the data scientist centers on enabling others to conduct research and educational activities using collections of digital data through consultation, collaboration and coordination. The NSB notes that individuals who are serving as Data Scientists hail from a variety of backgrounds including the technology and programming fields, disciplinary science fields and library/archival fields. No matter the background, the feature uniting Data Scientists is their ability to enable others to get more out of their data than they could have otherwise on their own. This is accomplished by the Data Scientist developing a deep understanding of the project's data and workflows, obtaining an intimate knowledge of the needs of the researchers who are using the data, and having the creative capacity to create the resources, tools and services that will connect researchers with the data in ways that would not have otherwise been possible.[49]

The idea of a data librarian is not a new one. There are academic librarians who provide data services and resources, particularly in the social sciences. However, although some position descriptions do include data management and archival responsibilities, data librarian positions are often more centered on incorporating data into traditional library services: reference, instruction, collection development, etc.[50] What makes the Data Research Scientist position different is the lack of public service and other traditional librarian responsibilities and the

primary focus on building interdisciplinary research initiatives in data management, curation, and preservation and related areas. At a high level, the Data Research Scientist is charged with developing an awareness and knowledge of available data management, access and preservation tools, and services and standards as well as understanding how they could be employed effectively to address specific needs and enrich research outcomes. The Data Research Scientist applies this knowledge in order to add value to digital data collections being generated or used for research at Purdue through consultation, collaboration, and coordination with researchers. On the ground level, the Data Research Scientist is responsible for helping the Libraries move research strategically forward through increasing awareness and visibility of Purdue Libraries' research on campus and elsewhere, fostering interactions between librarians and research faculty on campus, leveraging interdisciplinary research collaborations and co-investigations (particularly on grant funding) and increasing the Libraries' capabilities and opportunities to work in data related research and applied activities.

New Roles for Existing Positions – The Subject Specialist

In addition to building an expertise in data curation and preservation, the Data Research Scientist serves as a resource for subject specialists needing to develop the skills and abilities necessary to become a "blended librarian" and include data curation in their own work. Perhaps the largest responsibility of the Data Research Scientist is to act as a catalyst in initiating and building connections between researchers with data management, curation, dissemination, and preservation or other related needs and librarians who possess the knowledge and expertise to develop solutions to address these needs. These connections are built through the Libraries' having an understanding of the research being conducted at Purdue and reaching out to faculty who are generating and/or using data as a component of their research activities.

Collaborations between the Data Research Scientists and subject librarians to identify faculty researchers with data needs are critical. Subject librarians have already invested time and effort in building connections with faculty, learning about their particular area of research and understanding their information needs. These existing connections can also help the Data Research Scientist identify which faculty may be receptive to working with the library to address their issues and needs with regards to their research data. The relationship between a subject librarian and a faculty member can open the door to a discussion about the researcher's data needs and how the library may be able to address these needs.

However, the role of the subject librarians in the Libraries' data curation initiatives goes far beyond simply making introductions for the Data Research Scientist. As specialists in the information environment of a particular discipline, subject librarians are a vital part of any data curation project undertaken by the Libraries and need to be actively involved in every step of the process as a collaborator with the Data Research Scientist if the project is to succeed. Their understanding of a subject includes such things as knowledge of the information sources that exist for the discipline and how they are constructed and used along with an awareness of the metadata standards, thesauri and taxonomies or ontologies that are used to organize and manage information in their field. Many of the subject librarians at Purdue have advanced degrees in their field, and, even if they do not, they have gained an understanding of the cultural practice of research and the research process in their respective disciplines. Subject librarians can use their knowledge of a discipline to identify what data sets might be of use to others in the field and can articulate what description, documentation, or other information will be needed to enable the data to be used beyond a particular faculty's research project. Finally, subject librarians have

connections to professional societies within their discipline and have an understanding of the trends and the direction the discipline is likely to go in the future.

Once a researcher or research group with possible data management, dissemination, preservation or other related needs has been identified, the Data Research Scientist, with the help of the subject librarian, may contact the researcher to arrange a face to face meeting. The purpose of the meeting is typically to conduct a needs assessment to learn more about the specific data needs of the researcher or research team, and then to convey how the library may be able to address these needs.

The needs assessment is often done through a data interview with the faculty researchers and others who are connected to the data and is conducted by the Data Research Scientist and/or subject librarian. The questions asked will vary according to the nature of the research being conducted and the data being collected; however, there are some fairly standardized broad categories of inquiry. Beyond learning about the nature of the data (its format, size, type and other attributes), the interviewer seeks to uncover the unmet needs of the researcher or group generating or using the data and to obtain specific details about these needs. For example, if researchers state that they would like to make the data accessible to others beyond their own research group, then the interviewer asks about what data should be available to whom, at what point in the research cycle (immediately after the data has been processed, in conjunction with the publication of the results, 6 months after publication, etc.), and how the data should be made available. A sample of the types of questions that are typically asked during the data interview is provided in Figure 1.[51]

Figure 1: Sample Data Interview Schedule

- What's the story of the data?
  - Asked to obtain background information, context, and insight into the value of the data.
- What's the expected lifespan of the data?
  - Asked to obtain information about provenance and preservation needs.
- Who are the potential audiences for the data?
  - Asked to determine how the data might be used by others as well as if any restrictions on accessing need to be created.
- Who owns the data?
  - Asked to identify and clarify intellectual property issues.
- Does the data include any sensitive information?
  - Asked to identify any confidentiality issues that may have to be addressed.
- How should the data be made available?
  - Asked to determine needs for disseminating the data and for sharing it with others.

As the needs of the researcher are articulated, the Data Research Scientist and subject librarian consider and communicate ways in which the library could potentially address these needs as appropriate. In these instances, the work of the library is driven by the needs of the researcher and so the exact nature of the Libraries' proposals will vary accordingly. For example, if the researchers need a data management plan to satisfy a requirement of a funding agency we may assist in creating such a plan. If the researcher needs to develop a means of sharing data with others, we may propose developing metadata based on the community accepted standards to enable the data to be shared with the intended audiences. If the researcher wants to preserve his or her data for the long term, we may discuss what decisions, documentation and resources would be required to do this effectively and then develop a preservation model for their data. As a part of this communications stage, the Data Research Scientist may ask for a sample of the researcher's data. Having a sample of the data allows the Data Research Scientist to become

more familiar with the nature of the data under consideration. The sample can also serve as a test bed for the efficacy of the ideas being generated.

After the possible role(s) of the Libraries in addressing the needs of the researchers are discussed and solidified, the Data Research Scientist will often write up a summary of the articulated needs of the researcher and the initial work the Libraries propose to do in response. This document spells out the resources that are needed, the responsibilities of everyone involved in the project, provides an estimate for the amount of time allotted to the project, and describes the expected results. This written summary ensures that both sides have a shared understanding of the expectations and the work involved in achieving these expectations.

These initial collaborations are often smaller "pilot" projects conducted to test out the ideas, demonstrate the benefits of what the library has to offer, and get a sense of the time, effort and resources that would be required in a larger scale project. Frequently, however, the stated expectation is that the smaller pilot project will lead to a significant collaboration with the researcher to build and implement a full-scale project. These larger projects usually require more resources than the library is able to provide on its own, and so grant funding is sought. The pilot projects also serve to demonstrate to funding agencies the feasibility of the work being proposed in the grant application.

Faculty will typically seek grant funding in order to obtain the resources necessary to conduct their research projects. E-science projects are typically conducted in teams with each team member bringing a different set of skills and expertise to the project. In pursuit of grant funding, the principal investigator of the project will seek out partners who can address the requirements of the funding agency, contribute to the project in meaningful ways, and increase the likelihood of success. Ideally, the work done by the Data Research Scientist and the subject

librarian and their relationship with the principal investigator would lead to an invitation for them to participate in the grant proposal as co-principal investigators or senior personnel. As a co-PI or senior personnel, the Data Research Scientist and/or subject librarian collaborate with the other faculty and staff in the grant building, negotiating and writing process. They work as a part of the research team in all of the tasks that go into creating a grant proposal including: defining the specifics of the project and proposal, contacting the program officers of the grant, writing up the proposal following the terms and guidelines specified in the grant, negotiating a budget for the project, soliciting letters of support, and creating or gathering other needed documentation.

Collaboration in Practice: The CASPiE Project

One example of how the Purdue University Libraries Research Department is working to support e-science is the collaboration between the Libraries and the faculty and staff of the Center for Authentic Science Practice in Education (CASPiE). This collaboration enabled the Data Research Scientist and Chemical Information Specialist to embed themselves within the CASPiE project and to work closely with CASPiE administration.

CASPiE is a multi-institutional NSF-funded undergraduate research center headquartered at Purdue University.[52] It provides authentic research experiences to first and second-year students in order to increase student interest and retention in the sciences. Rather than simply answering textbook questions or conducting experiments in which the results are already known, students enrolled in CASPiE courses investigate a real-world research problem to support the work of a faculty member. The nature of the research problem is explained through a course module developed by a faculty member (subsequently referred to as the module author). After students complete a series of introductory lab sessions, they then develop their own hypotheses

and design their own experiments. The module author then receives the results of the experiments to investigate whether anything of interest has been discovered by the students.

The students in CASPiE courses have access to a system of networked high-quality scientific instruments to gather research-quality data for their experiments if required by the module. These highly advanced instruments – including a Raman spectrometer, a liquid chromatograph with array detector, and a gas chromatograph – are not typically available for use by undergraduate students. CASPiE investigators collaborating with Information Technology at Purdue (ITaP) developed a means to network these instruments and make them available for students through the Internet while monitoring how these instruments are accessed and used. This network allows many students at different geographic locations to access the instrument network hosted by Purdue and to run their samples remotely.[53]

The Libraries' involvement with CASPiE began in April 2007 when the Data Research Scientist and the Chemical Information Specialist attended a seminar given by the Director of Instrumentation Networking for CASPiE. Although the seminar focused more on the technical aspects of CASPiE's instrumentation network, it was apparent to the librarians by the end of the seminar that data management was a growing problem for the project. CASPiE and ITaP had built a very powerful and secure high throughput system of remote instrumentation, but they lacked the means to manage this data as effectively as desired. It was clear that as more instruments came onto the network, as more institutions participated in the CASPiE project, and as more CASPiE courses were created, the data being generated would quickly become unmanageable.

The Data Research Scientist made contact with the Director of Instrumentation Networking to introduce the Libraries' interest in supporting data curation, and to request a

meeting to learn more about the CASPiE project and their needs in managing and archiving their data. Over the course of several meetings, these librarians met with CASPiE staff involved with the management and operation of the instrumentation network. CASPiE did not have a data management plan that reflected what they wanted to accomplish with their data or the infrastructure in place to support the effective use, access, or preservation of the data they were generating. Furthermore, there was a lack of metadata describing the data which made the data difficult to use. From these meetings the librarians began to understand the workflow outside the instrumentation network, including how students generated additional data through in-class experiments, how students recorded additional information during and after lab in their notebooks, and how the final data and conclusions were forwarded to the module author for review and future exploration.

After these meetings the Data Research Scientist and Chemical Information Specialist met to design a pilot project that would address the issues articulated by CASPiE's administration. In consultation with the Libraries' Associate Dean for Research, the Data Research Scientist and Chemical Information Specialist worked out what the library could offer CASPiE at this stage, defined the boundaries of the pilot, obtained the necessary resources and worked out release time for the Chemical Information Specialist to work on the pilot. The Libraries' initial offer was to provide 200 hours of staff time for a pilot project to design a prototype model for managing and archiving the data from one of CASPiE's modules based on its existing workflows.

For the Libraries, the expectation was that building this prototype would serve two purposes. First, the prototype might serve as an example solution that could potentially be extended to a full data management and archiving model for the entire CASPiE program and

eventually be applied to other similar research projects or situations. The second expectation was that the prototype could be used to leverage additional grant funding for the Libraries and CASPiE that would provide the necessary resources to undertake the development of the full-scale system. CASPiE administration understood these expectations and identified a suitable module to serve as the basis for the pilot project. The librarians worked step-wise through the module identifying where students and instruments were generating and recording data.

In order to be successful in our efforts with this pilot project, the Data Research Scientist and Chemical Information Specialist had to become more familiar with the:

- interests of the module's author (the faculty researcher) with the data, with the particular lab module itself and with the purpose of each of the analytical methods used;
- workflow of the students and CASPiE staff as they implemented the module and generated data;
- size, format, type and other attributes of the data;
- desired services and outcomes for the data for all parties involved; and
- metadata and other documentation that would be required to enable these services and outcomes.

A data management plan for the selected CASPiE module was developed over the course of several months. As the Data Research Scientist and the Chemical Information Specialist gathered the information necessary to create the plan from CASPiE's staff and from reviewing the data generated by the students three distinct stages emerged. First, in the "educational" stage the data is generated by the students enrolled in the course and reviewed by the instructors for evaluation. The instructor teaching the course (who may or may not be the module author) needs to have access to the data and sufficient documentation to complete the educational aspects of the CASPiE project. Second, in the "research" stage the module author needed access to the data in order to analyze it further and to apply the results towards answering the research question he or she had set forth. This would require sufficient metadata for the module author to understand

enough of how and why the data had been generated to be able to trust in its quality and to reproduce the data if necessary. Access to the data at this point would be determined by the module author. Finally in the "archive" stage, once the module author had exhausted the value of the data for his or her own purposes and published the results, the data could be ingested into the Libraries' e-data repository to be preserved for the long term. The data would be freely available and its metadata would be in openly accessible formats and structured according to community standards that would enable the re-use of the data by others.[54]

The final report for the Libraries' pilot project was presented to CASPiE administration in early February of 2008. The report contained a description of the proposed data management model, a discussion of the types of metadata that would be needed to enable desired functionality, details of what the challenges might be in implementing this model, and a list of recommendations for actions and future directions to overcome these challenges.

Knowing that implementing this model would require additional funding, the librarians, CASPiE and ITaP began pursuing sponsored funding in late Fall of 2007. Once a suitable call for proposals was identified, the first step was to define and articulate the role the Libraries could and should play in the overall solicitation. The approach was to take what had already been learned about the data management needs of CASPiE from the work that had done, and to adapt that knowledge to meet the requirements of the grant proposal. Collaborating with the other co-PIs on the grant also required the Libraries Research Department to negotiate for the resources it would need to accomplish its part of the proposal, while still remaining within the overall budget of the grant. The Data Research Scientist was made one of the co-PI's and designated to coordinate the Libraries efforts in the process of data capture, curation and preservation for the project. The Chemical Information Specialist was named as senior personnel in the grant to

identify and translate research and instrument methodologies into data curation and archive functions. If this grant is awarded, the Libraries will be an integral part of the research team driving the CASPiE project. If this grant is not awarded, the Libraries and the administration of CASPiE are still committed to partnering with the Libraries in seeking out the sources of funding needed to build data management, curation and preservation tools for the CASPiE project.

The partnership between the Data Research Scientist and the Chemical Information Specialist was essential in designing and developing the prototype model to address CASPiE's data needs and in applying for grant funding to continue this work. The Data Research Scientist had little previous experience in Chemistry and needed assistance in understanding the nature of the instrumentation and the processes used to generate the data. This project was the Chemical Information Specialist's first direct involvement with data management and curation issues and he needed guidance in navigating the many aspects involved in addressing data issues effectively. The Data Research Scientist served as the project manager and conducted the interviews to identify the specific needs of the CASPiE administration and the module author with regards to their data. He reviewed CASPiE's processes and scientific workflows to determine how data management, curation and preservation functions could best be incorporated into their existing systems and investigated the metadata that might be needed to enable discovery and preservation. The Data Research Scientist also helped to guide the Chemical Information Specialist through the process of applying for grant funding. The Chemical Information Specialist researched what descriptive, administrative and structural metadata standards might be applicable to address CASPiE's needs for their data. He tutored the Data Research Scientist in the research techniques that were used in the CASPiE module and provided insight as to how the data might be used by the researcher or repurposed by others. The work on

the CASPiE project was a team effort and would not have succeeded without the collaboration between the Data Research Scientist and the Chemical Information Specialist.

Skills Needed for Support E-Science Research

The collaboration between research faculty and staff of CASPiE, ITaP, and the Purdue University Libraries is only one illustration of how librarians at Purdue are defining roles for themselves in the era of e-science. The Purdue Libraries' model of building the capacity and capabilities of librarians to participate in interdisciplinary research as equal partners with academic faculty and IT professionals has pushed librarians at Purdue to adapt their skill sets to this new role. Different types of librarians at Purdue are continuing to build up or rework their specific abilities as appropriate to their particular specialties in the library; however, from our experiences we have already begun to see that all librarians who will play an active role in interdisciplinary research will need to develop a set of shared skills as well. The need for librarians to develop strong technical skills should not be underestimated, but technical skills alone are insufficient for success. Soft skills such as communication, creativity and flexibility, and risk taking are equally important. These skill sets are interrelated, but require some explanation individually.

First, a strong set of communication skills is a key component in enabling librarians to participate in interdisciplinary research and collaborate with faculty and IT professionals. This skill set includes the ability to actively listen, understand, and empathize with the information needs of the researcher within the context of the particular research project. The librarian must also be able to articulate what the library has to offer and translate library science concepts so that all parties can understand them and see how they would add value to the project. The hard work of meeting on common ground helps build strong relationships, which is critical as the

researchers must get to know the librarian personally and be able to trust that the librarian will be able to deliver if the librarian is to become part of the research team. Negotiation skills are also important as librarians will need to ensure that enough resources are allocated to support their role in the research project.

Second, success in interdisciplinary research rests on creativity and flexibility. Librarians need to be willing and able to conceptualize how their skills could be applied in new ways outside of the traditional domains of the library. This requires the creativity to be able to see possibilities for the application of library science and the involvement of librarians in a new and relatively undefined frontier. In order to lay their claim to this new ground, librarians must expand their knowledge base beyond the library science field and become familiar with the discussions that are taking place in other fields as Gibbons noted in her description of the "R&D mind-set." Obtaining this familiarity is essential in bridging the "language barrier" that may exist ("curation," for example, means different things to different people) and in conceiving possibilities for librarians to become more directly involved. Finally, events and circumstances in data projects often change or evolve over time as things progress or as more information becomes known. In this type of environment, librarians will need the flexibility to react accordingly and reframe their role in the project to reflect any changes that occur.

Third, entering this new frontier involves a willingness to take risks. As has been illustrated in the previous skill descriptions, embedding oneself in an interdisciplinary research project entails leaving the comfort zone of the well established practices and procedures of library science and venturing into unknown territories. Even beyond leaving one's comfort zone, investing time and effort in interdisciplinary research, such as data curation and preservation projects, is a risky endeavor for librarians. The amount of effort required just to design a project

and get it off the ground demands a high degree of commitment and investment from all team members, especially when it comes time to locate, write and submit a grant proposal. Unfortunately, even with this investment, the logistical challenges of putting a project together and the nature of grant funding are such that a lot of effort may not lead to any tangible results. Despite the prospect of failure however, the Purdue Libraries have found it worthwhile to take the risk. Enough of the projects that we have been involved in have received support and funding, thereby justifying the investment of librarians time and effort. Furthermore, even in situations where the project did not take off as planned, the experience of having been a part of a research team as a full-partner is a valuable learning experience for librarians, because it better prepares us for future opportunities, and helps us forge deep relationships with faculty and IT professionals that otherwise would not have developed.

Conclusion

The Purdue University Libraries initiatives in supporting e-science through the active involvement of librarians in collaborations in interdisciplinary research projects have been quite successful thus far. Since the initiative began in 2005 more than 22 librarians have participated in more than 47 multidisciplinary grant proposals including proposals from the National Institute of Health, the National Science Foundation, the National Endowment for the Humanities, the United States Department of Agriculture as well as regional and local grants. Fourteen of these grants have been awarded.

The creation of a Libraries research department and related positions such as Associate Dean of Research and Data Research Scientist have enabled individual librarians as well as the Purdue Libraries as a whole to expand their participation in research support in the age of e-science and cyberinfrastructure. This new organizational structure can coordinate the Libraries'

efforts in contacting potential collaborators, seeking sponsored funding, and educating and assisting subject librarians in becoming more comfortable with enacting data services for faculty and research groups. The Purdue Libraries are now in the process of disseminating and sharing our experiences in this area through such events as participating in the University of Illinois at Urbana-Champaign's summer data curation institute and hosting a Committee on Institutional Cooperation (CIC) conference on libraries and e-science.[55]

To be sure, the Purdue University Libraries have only just begun to explore new roles for librarians and there are still many challenges yet to face in defining what the role of librarians could or should be in curating or preserving data, or in supporting e-science in general. However, as e-science continues to provide new capabilities – and even requirements – for exploration and research, librarians and the services they provide must also evolve to meet the information needs of their patrons in emerging data driven research environment of the 21st century. Furthermore, data curation and preservation activities give librarians an opportunity to reclaim our status as the central provider and steward of research information, no matter in what form it is captured. By so doing, the profession will add value to e-science as it evolves, equipping its discoveries to persist and to remain accessible into the years ahead.

Works Cited

1. Cyberinfrastructure Council, "Cyberinfrastructure Vision for 21st Century Discovery." National Science Foundation (2007). Available online from http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf [Accessed 18 July 2008].
2. Blue-Ribbon Advisory Panel on Cyberinfrastructure, "Revolutionizing Science and Engineering through Cyberinfrastructure." National Science Foundation (2003). Available online from http://www.nsf.gov/od/oci/reports/atkins.pdf [Accessed 18 June 2008].
3. G. Sayeed Choudhury, "The Virtual Observatory Meets the Library." *Journal of Electronic Publishing* 11, no.1 (2008). Available online from http://hdl.handle.net/2027/spo.3336451.0011.111 [Accessed 22 September 2008]; Sayeed Choudhury, Tim DiLauro, Alex Szalay, Ethan Vishniac, Robert Hanisch, Julie Steffen, Robert Milkey, Teresa Ehling, and Ray Plante, "Digital Data Preservation for Scholarly Publications in Astronomy" *The International Journal of Digital Curation* 2, no.2 (2007): 20-30. Available online from http://jhir.library.jhu.edu/handle/1774.2/32796 [Accessed 22 September]; U.S. National Virtual Observatory, http://www.us-vo.org/index.cfm [Accessed 22 September 2008].
4. Commission on Cyberinfrastructure for the Humanities and Social Sciences, "Our Cultural Commonwealth." American Council of Learned Societies. Available online from http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf [Accessed 18 July 2008].

5. Institute for Advanced Technology in the Humanities at the University of Virginia, "Rome Reborn 1.0." (2007). Available online from http://www.romereborn.virginia.edu/ [Accessed 18 July 2008].

6. Stephen Emmott, ed. *Towards 2020 Science*. Cambridge, England: Microsoft Research Ltd., 2006.

7. Tony Hey and Anne Trefethen. "The Data Deluge: An E-Science Perspective." In *Grid Computing: Making the Global Infrastructure a Reality* edited by Fran Berman, Geoffrey Fox and Anthony J.G. Hey, 809-24. Chichester: Wiley, 2003.

8. Ewa Deelman and Yolanda Gil, "Workshop on the Challenges of Scientific Workflows: Executive Summary." Information Sciences Institute, University of Southern California (October 16, 2006). Available online from http://vtcpc.isi.edu/wiki/images/7/71/NSF-Workfow-Summary.pdf [Accessed 18 June 2008].

9. Cyberinfrastructure Council, "Cyberinfrastructure Vision," 24.

10. Philip Lord, Alison Macdonald, Liz Lyon, and David Giaretta, "From Data Deluge to Data Curation." UK e-Science All Hands Meeting (2004). Available online from http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/150.pdf [Accessed 18 June 2008].

11. ARL Workshop on New Collaborative Relationships, "To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering." Association of Research Libraries. Available online from http://www.arl.org/bm~doc/digdatarpt.pdf [Accessed 18 June 2008].

12. Michael Witt, "Institutional Repositories and Research Data Curation in a Distributed Environment." *Library Trends* 57, no. 3 (forthcoming).

13. Christopher L. Greer, "A Vision for the Digital Data Universe." (2007). Available online from https://www.nanohub.org/resources/2291/ [Accessed 19 July 2008].

14. Cyberinfrastructure Council, "Cyberinfrastructure Vision," 31.

15. Mark C. Sheehan, "Higher Education IT and Cyberinfrastructure: Integrating Technologies for Scholarship." EDUCAUSE Center for Applied Research (2008). Available online from http://www.educause.edu/ir/library/pdf/ers0803/rs/ERS0803w.pdf [Accessed 18 July 2008].

16. Nicholas Joint, "Data Preservation, the New Science and the Practitioner Librarian." *Library Review* 56, no. 6 (2007): 450-5.

17. Emmott. “Towards 2020,” 19.

18. Richard E. Luce, "A New Value Equation Challenge: The Emergence of Eresearch and Roles for Research Libraries." Council on Library and Information Resources (2008). Available online from http://www.clir.org/activities/registration/08FebR21/Luce.pdf [Accessed 23 July 2008].

19. Andy Guess, “At Libraries, Taking the (Really) Long View.” *Inside Higher Ed* (2008). Available online from http://www.insidehighered.com/news/2008/07/23/preservation [Accessed 31 August 2008].

20. Joint Task Force on Library Support for E-Science, "Agenda for Developing E-Science in Research Libraries." Association of Research Libraries (2007). Available online from http://www.arl.org/bm~doc/ARL_EScience_final.pdf [Accessed 18 June 2008].

21. Joint Task Force. “Agenda for Developing,” 17.

22. ARL Workshop. “To Stand,” 12.

23. Ibid , 12-13.

24. Anna Gold, "Cyberinfrastructure, Data and Libraries, Part 1: A Cyberinfrastructure Primer for Librarians." *D-Lib Magazine* 13, no. 9/10 (2007). Available online from http://www.dlib.org/dlib/september07/gold/09gold-pt1.html [Accessed 11 June 2008].

25. Research Information Network and Consortium of Research Libraries, "Researchers' Use of Academic Libraries and Their Services." (2007). Available online from http://www.rin.ac.uk/files/libraries-report-2007.pdf [Accessed 29 July 2008]; University of Minnesota Libraries, "Sciences Assessment." (2005). Available online from http://www.lib.umn.edu/about/scieval/ [Accessed 29 July 2008].

26. David G. Messerschmitt, "Opportunities for Research Libraries in the NSF Cyberinfrastructure Program." Association of Research Libraries (2003). Available online from http://www.arl.org/resources/pubs/br/br229/br229cyber.shtml [Accessed 9 February 2008].

27. Susan Gibbons. *The Academic Library and the Net Gen Student: Making the Connections*. Chicago, IL: American Library Association, 2007.

28. Wendy Pradt Lougee, "Diffuse Libraries: Emergent Roles for the Research Library in the Digital Age." Council on Library and Information Resources (2002). Available online from http://www.clir.org/pubs/reports/pub108/pub108.pdf [Accessed 23 July 2008].

29. James G. Neal, "Raised by Wolves: Integrating the New Generation of Feral Professionals into the Academic Library." *Library Journal* 131, no. 3 (2006): 42-44.

30. Stanley Wilder, "The New Library Professional." *The Chronicle of Higher Education* (2007). Available online from http://chronicle.com/jobs/news/2007/02/2007022001c.htm [Accessed 18 July 2008].

31. Stephen J. Bell and John D. Shank, "The Blended Librarian: A Blueprint for Redefining the Teaching and Learning Role of Academic Librarians." *C&RL News* 65, no. 7 (2004): 372-75.

32. Victoria Matthew and Ann Schroeder, "The Embedded Librarian Program: Faculty and Librarians Partner to Embed Personalized Library Assistance into Online Courses." *EDUCAUSE Quarterly* 29, no. 4 (2006): 61-65; Karen M. Ramsay and Jim Kinnie, "The Embedded Librarian: Getting out There Via Technology to Help Students Where They Learn." *Library Journal* 131, no. 6 (2006): 34-35.

33. Nancy H. Seamans and Paul Metz, "Virginia Tech's Innovative College Librarian Program." *C&RL* 63, no. 4 (2002): 324-332; Jane E. Schillie, Virginia E. Young, Susan A. Ariew, Ellen M. Krupar, and Margaret C. Merrill, "Outreach Through the College Librarian Program at Virginia Tech." *The Reference Librarian* 71 (2000): 71-8.

34. Brenda L. Johnson & Laurie A. Alexander, "In the Field: An Innovative Role Puts Academic Librarians Right in the Departments They Serve." *Library Journal* (2007). Available online from http://www.libraryjournal.com/article/CA6407750.html [Accessed 1 September 2008]; Wendy Pradt Lougee. "Diffuse Libraries," 19.

35. "The Next Level: Preeminence: Strategic Plan 2001-2006." Purdue University (2001). Available online from http://www.purdue.edu/strategic_plan/index.html [Accessed 18 June 2008].

36. "Discovery Park: Home." Purdue University (2008). Available online from http://www.purdue.edu/dp/index.php [Accessed 23 July 2008].

37. Network for Computational Nanotechnology, Purdue University, "nanoHUB." (2008). Available online from http://www.nanohub.org/ [Accessed 19 July 2008].

38. James L. Mullins, interview by Jake R. Carlson, June 9, 2008.

39. D. Scott Brandt and James L. Mullins, "Building an Interdisciplinary Research Program in an Academic Library." Coalition for Networked Information (Apr. 4, 2006). Available online from http://www.cni.org/tfms/2006a.spring/abstracts/handouts/CNI_Building_Mullins.ppt [Accessed 23 June 2008]; James L. Mullins, "Enabling International Access to Scientific Data Sets: Creation of the Distributed Data Curation Center (D2C2)." Purdue University (2007). Available online from http://docs.lib.purdue.edu/lib_research/85/ [Accessed 18 June 2008].

40. Joint Task Force, "Agenda for Developing E-Science."

41. "Final NIH Statement on Sharing Resarch Data." National Institutes of Health (2003). Available online from http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html [Accessed 18 June 2008].

42. "Grant Policy Manual: Chapter VII - Other Grant Requirements - Section 734." National Science Foundation (2005). Available online from http://www.nsf.gov/pubs/manuals/gpm05_131/gpm7.jsp [Accessed 18 June 2008].

43. Office of Cyberinfrastructure, "Sustainable Digital Data Preservation and Access Network Partners (DataNet)." National Science Foundation. Available online from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141 [Accessed 18 June 2008].

44. Brandt, Mullins, “Building an Interdisciplinary Research Program in an Academic Library”; Mullins, “Enabling International Enabling International Access to Scientific Data Sets: Creation of the Distributed Data Curation Center (D2C2).”

45. D. Scott Brandt, "Librarians as Partners in E-Research: Purdue University Libraries Promote Collaboration." *C&RL News* 68, no. 6 (2007): 365-7, 96.

46. For a more complete account of how the D2C2 was created see – Mullins, “Enabling International Enabling International Access to Scientific Data Sets: Creation of the Distributed Data Curation Center (D2C2).”

47. James L. Mullins, interview by Jake R. Carlson.

48. D. Scott Brandt, "Vision for the Distributed Data Curation Center." Purdue University. Available online from http://d2c2.lib.purdue.edu/vision.php [Accessed 18 June 2008].

49. National Science Board, "Long-Lived Data Collections: Enabling Research and Education in the 21st Century." National Science Foundation (2005). Available online from http://www.nsf.gov/pubs/2005/nsb0540/start.htm [Accessed 18 June 2008].

50. Michael N. Cook, John J. Hernandez, and Shawn Nicholson. *SPEC Kit 263: Numeric Data Products and Services*. Washington, D.C.: Association of Research Libraries, 2001.

51. Michael Witt and Jake R. Carlson, "Conducting a Data Interview " Purdue University (2007). Available online from http://docs.lib.purdue.edu/lib_research/81/ [Accessed 24 June 2008].

52. "The Center for Authentic Science Practice in Education." CASPiE (2004). Available online from http://www.purdue.edu/dp/caspie/ [Accessed 24 June 2008].

53. Fred E. Lytle, Gabriela C. Weaver, Phillip Wyss, Debora Steffen, and John Campbell, "Making Instrumentation a Secure Part of the Cyberinfrastructure." (working paper, 2008).

54. Jake R. Carlson and Jeremy R. Garritano, "Preserving Undergraduate Research Data." Purdue University (2007). Available online from http://docs.lib.purdue.edu/lib_research/82/ [Accessed 18 June 2008].

55. Center for Library Initiatives, "CIC Library Conference: Librarians and E-Science: Focusing Toward 20/20." Committee on Institutional Cooperation (2008). Available online from http://www.cic.uiuc.edu/programs/centerforlibraryinitiatives/Archive/ConferencePresentation/Conference2008/home.shtml [Accessed 24 August 2008].