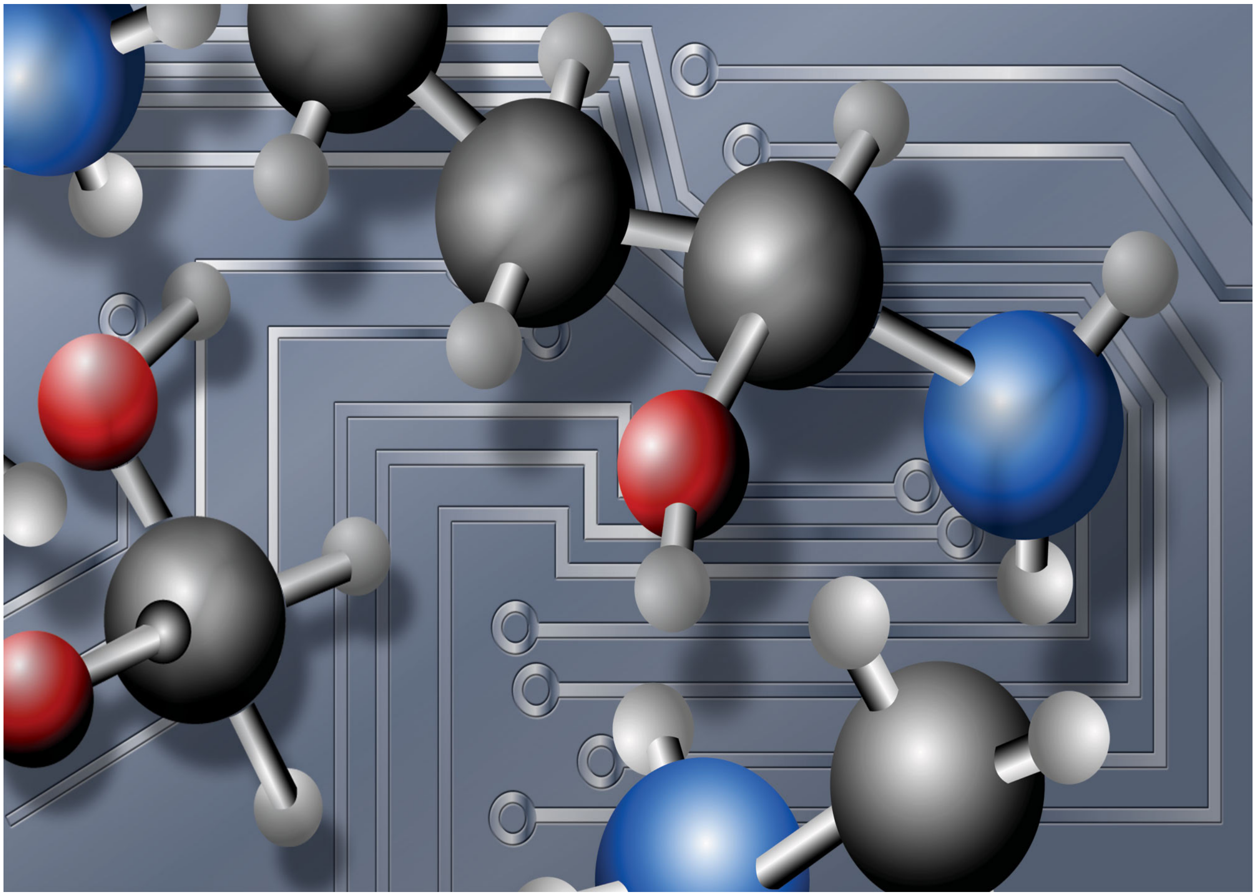
E-Science and Data Support Services

A Study of ARL Member Institutions

August 2010



Catherine Soehner, University of Michigan
Catherine Steeves, University of Guelph
Jennifer Ward, University of Washington

E-Science and Data Support Services:
A Study of ARL Member Institutions

August 2010

Catherine Soehner, University of Michigan
Catherine Steeves, University of Guelph
Jennifer Ward, University of Washington

Acknowledgements

The authors are grateful to a wide variety of individuals who willingly shared their expertise with them. In particular, James L. Mullins, Dean of Libraries, and D. Scott Brandt, Associate Dean for Research at Purdue Libraries; Luc Declerck, Associate University Librarian, Technology Services and Ardys Kozbial, Technology Outreach Librarian, in the UCSD Libraries, and David Minor, Head of Curation Services, San Diego Supercomputer Center; Gail Steinhart, Research Data & Environmental Sciences Librarian, Medha Devare, Bioinformatics and Life Sciences Librarian, Dianne Dietrich, Research Data & Metadata Librarian, and Dean Krafft, Chief Technology Strategist, at Cornell University; Sayeed Choudhury, Associate Dean for Library Digital Programs, at Johns Hopkins University; Dr. Robert J. Sandusky, Assistant University Librarian for Information Technology, at the University of Illinois at Chicago; and MacKenzie Smith, Associate Director for Technology, and Katherine McNeill, Data Services and Economics Librarian, at the Massachusetts Institute of Technology.

Cover Image: Cyber-enabled Chemistry

Credit: Nicolle Rager Fuller, National Science Foundation
Chemistry meets computer, data and networking technologies in the National Science Foundation's first Cyber-enabled Chemistry Program awards. NSF's chemistry division developed the program to explore how researchers and educators in that field can fully exploit the potential of cyberinfrastructure.
Courtesy: National Science Foundation

Published by

Association of Research Libraries
Washington, DC 20036
www.arl.org

Contents

Preface

The dramatic evolution of technology in the late 20th century gave rise to a new age of science. Variously described as e-science, cyberscience, or the "fourth paradigm,"[1] the emergent era of scientific discovery distinctively exploits technologies for computation, data curation, analysis and visualization, and collaboration.

The fundamental shifts in scientific practice are, not surprisingly, also affecting the institutions and organizations committed to advancing science and supporting scientists. Recognizing the impact and opportunities for research libraries, the Association of Research Libraries launched an E-Science Task Force in 2006. The group defined the domain of e-science as those new methods that are large-scale, data-driven, computationally intense, and often engaging research teams across institutional boundaries. The Task Force's first report[2] outlined the challenge: e-science fundamentally alters the ways in which scientists carry out their work, the tools they use, the types of problems they address, and the nature of the documentation and publication that results from their research. Consequently, e-science requires new strategies for research support and significant development of infrastructure.

The E-Science Task Force was followed by an ongoing Working Group with a charge to develop member understanding of the changing requirements for professional skills and infrastructure and to address evolving policies and protocols for data stewardship and a new, data-enriched publishing genre. The recognition of the different approaches being undertaken by individual institutions prompted a survey of the ARL membership in 2009. The survey sought baseline data about institutional and library engagement with these issues, targeting information about planning structures, projects, programs and services, and the library workforce. This report explores the survey results and provides a more in-depth analysis of institutional models.

The survey data revealed a professional community experiencing a not-unexpected diversity of approaches and investments. The data provide a picture of institutional infrastructure in transition, analyzing options and leveraging organizational capacities in ways that reflect the institution's unique culture and assets. Not unlike the early days of digital library development, the degree of library investment and capacity building at this early stage varies considerably.

The emergence of the "fourth paradigm" will have profound impacts on science and will prompt equally profound examination of library roles and information infrastructure. This report, a portrait of emergent models, should provide significant background to inform the dialogue and enable development within the research library community.

Wendy Pradt Lougee, University of Minnesota
Chair, ARL E-Science Task Force and Working Group

[1] Jim Gray's characterization of phases of scientific discovery is captured in: The Fourth Paradigm: Data-Intensive Scientific Discovery, Edited by Tony Hey, Stewart Tansley, and Kristin Tolle. http://research.microsoft.com/en-us/collaboration/fourthparadigm/

[2] ARL Joint Task Force on Library Support for E-Science has released its final report, an "Agenda for Developing E-Science in Research Libraries." http://www.arl.org/bm~doc/ARLESciencefinal.pdf

Executive Summary

The Association of Research Libraries E-Science Working Group developed a survey to "build an understanding of how libraries can contribute to e-science activities in their institution" and "identify organizations and institutions that have similar interests in e-science to leverage research library interests." The August 2009 survey gathered 57 responses to the survey from the 123 ARL member libraries in the United States and Canada. Twenty-one respondents report their institution provides infrastructure or support services for e-science, 23 institutions are in the planning stages, and 13 do not provide support for e-science.

After analyzing the survey results, the authors identified a small set of respondents (Purdue University, the University of California, San Diego, Cornell University, Johns Hopkins University, the University of Illinois at Chicago, and the Massachusetts Institute of Technology) for interviews to further elaborate their activities. The resulting six case studies synthesize the interviews with the corresponding institutions' responses to the survey. The cases further illuminate programs and services mentioned only briefly in the survey and allow some interesting patterns to emerge from interviewees' reflections on faculty connections, staffing levels, and organizational structure and culture.

This report presents a summary of the survey results and the six cases studies. It also includes a bibliography of related articles, reports, and Web sites, along with the survey instrument and a selection of recent research library position descriptions with significant e-science support components.

Survey Results

As stated in the survey introduction, "e-science is defined broadly not only as big computational science, but also team science and networked science. It includes all scientific domains, as well as biomedicine and social sciences that share research approaches with the sciences." This broad definition of e-science means that support for e-science necessarily incorporates a broad range of activities and services. Rather than a single service, e-science requires the development, co-ordination, and synthesis of a range of investments to create a support system. Data curation, preservation, access, and metadata are areas of e-science where libraries find a natural affinity, and consequently dominated the survey responses. However, a few institutions also provided evidence or commentary on institutional initiatives and activities that reflect the broader definition offered by the ARL E-Science Working Group. Some respondents also considered broad scholarly communications efforts and digital repository activities as part-and-parcel of e-science efforts.

Approaches to E-Science in Institutions

Institutionally, four major approaches to creating an e-science support system were reported.

1. Institution-wide or Centralized Response

Of the 44 respondents that provide or are planning support for e-science, only four (9%) indicated that their institution had or was planning to rely mainly on an institution-wide group or task force to advance e-science planning and policy development.

2. Unit-by-unit or Decentralized Approach

Eleven respondents (25%) describe their organizational culture as being highly decentralized with the focus of e-science activities in particular subject areas or interdisciplinary institutes. These institutions developed infrastructure and policies related to individual units' (i.e., departments, colleges, schools, etc.) e-science needs and lack an overall centralized response to e-science.

3. Hybrid of Both Decentralized and Centralized Efforts

The majority of the survey respondents (27 or 61%) indicated that their institution had or was planning a hybrid structure that included both institution-wide and unit-specific efforts to advance e-science planning and policy developments.

4. Multi-institutional Collaborations

Many institutions are collaborating with one another to tackle various aspects of e-science and are often the result of partnering on grants. Approximately half of the respondents (20 of 41) reported that their institutions were involved in a collaborative program with another institution. This response to e-science is anticipated as becoming the fastest growing response simply due to the size of the issues involved.

Data Support and Services in Institutions

While the majority of respondents (23 of 42) indicated that there were no designated units to provide data curation and support for scientific research data on their campus, 19 institutions (45%) did identify the presence of designated unit(s) for this purpose. The units included data centers, disciplinary informatics centers and institutes, statistical analysis and academic computing centers and services, library data and institutional repositories, digital research and curation centers, campus information technology units, and high-performance computing and cyberinfrastructure centers and institutes. Although not a part of this survey, additional information about incentives and policies for the use of centralized data centers would be a useful component to understanding and creating successful centralized services.

While many institutions have yet to conduct an evaluation of service needs among their researchers, more than a third of respondents (16 of 42) reported conducting assessments relating to data services and more were planning assessment activities.

Approaches to E-Science in Libraries

Approximately 73% of the respondents (29 of 40) indicated the library was involved in e-science support at their institutions. Leadership of these efforts was primarily through a team effort (15 of 31 respondents or 48%) or some combination of individuals, units, and teams working together (13 or 42%). When asked about campus collaborations, 87% (27 of 31) noted that their e-science services were provided in collaboration with at least one other unit on campus, with a majority of those being the centralized IT unit.

All but a few of the responding libraries are providing e-science consultation and reference services, such as finding and using available technology infrastructure and tools, finding relevant data, developing data management plans, and developing tools to assist researchers. A few described more advanced services such as "archiving relevant data and curating it for long-term preservation and integration across datasets" and "providing curatorial and data stewardship services" as part of data management plans.

Libraries reported using a mix of strategies to create a workforce with the skills and capacity to provide e-science services and programs. Considering the current economic climate, it is not surprising that a majority of libraries (18 of 28 respondents) are reassigning existing staff or providing training to existing staff as part of an overall strategy to incorporate e-science responsibilities into their current portfolios. In addition to reassigning existing staff, libraries have hired or plan to hire staff specifically to provide e-science services as part of an overall strategy. This investment of resources even during budget cuts indicates the level of commitment to e-science services by many of the respondents to the survey. The MLIS degree was listed in a majority of positions that libraries have or are planning to have to provide e-science programs and services. There is some evidence in more recent relevant job postings that the focus on the MLIS is in flux.

Pressure Points

The top three areas identified by survey respondents as pressure points include a lack of resources, difficulty acquiring the appropriate staff and expertise to provide e-science and data management or curation services, and the lack of a unifying direction on campus. While libraries identified a significant list of pressure points, an overall enthusiasm for new roles in the academic research process was evident throughout the survey responses and in the case studies.

Observations

Collaborations are essential to address even modest support of e-science. The survey revealed successful and frequent collaborations on all levels: between libraries of different institutions, between libraries and the departments they serve, between various departments to address interdisciplinary subject areas, and between institutions. Because the data sets created by modern scientific methods are often very large, the resources required to manage such data must also be extensive. The services to support research often require capabilities from different units. Since these services need to manage costs but also tap diverse expertise, collaborations will continue to be an important method to address the enormity of the challenges posed by e-science.

Faculty interest and institutional support at the administrative level are important for success of library services in this area. Without faculty and institutional engagement, libraries will find themselves preaching about the importance of data curation, preservation, and access without making an impact. All of the case studies described in this report reinforce this point.

The Master of Library and Information Science degree has a place in this new area of librarianship. Since reassigning staff is a major strategy for resourcing roles in e-science in libraries, it is not too surprising that the MLIS was reported in a majority of data positions. Although science degrees and expertise are also highly valued in these new roles, the cooperative and team-oriented values of most MLIS degree holders, along with the understanding of the role and importance of metadata and preservation, makes the combination of the MLIS along with an advanced science degree useful and successful.

The fact that investments in e-science activities are being made even during difficult budget times demonstrates that this is a priority for libraries. As research activities become more data intensive and as faculty and institutions become increasingly concerned about the preservation and access to that data, libraries have an opportunity to demonstrate their expertise and relevance to their institutions and are taking advantage of opportunities to move forward into data management activities. Collecting and preserving information for researchers is a recognized capability of libraries and, therefore, libraries are obvious partners to provide the expertise to assist the university with a centralized plan for data management.

Strategies for data curation, management, and preservation are still young and evolving. DataNet grants through the National Science Foundation as well as other externally and internally funded programs will provide substantial models and information that will help guide decisions over the next few years.

Conclusion

The results of this survey indicate that research institutions are quickly rising to meet the challenge of managing data, especially in light of the anticipated federal government requirements for data plans as part of grant proposals. There is great diversity in the strategies employed by institutions to address the needs of their researchers, but this diversity of response reflects the needs and culture of the institutions. The trend toward centralization of these services through campus committees and the development of central data centers reflects the growing understanding of researcher needs and an appreciation for benefits that can only be gained through a centralized service.

Collectively, the respondents show an increasingly sophisticated view of data management skills, services, and resources. Promising strategies for engagement in these activities are becoming clear and are often highly collaborative. Many respondents to the survey seem to have made substantial progress toward achieving e-science support systems. However, there are still substantial gaps in their support capabilities. Institution-wide activities are often policy or planning-focused, with service delivery still somewhat fragmented and underutilized.

The key challenges facing research libraries as they consider this new area of service to their communities are both tangible and social in nature. Many survey respondents cited lack of money and resources as the most obvious physical limitations to quick mobilization around these issues. However, just as compelling were pressures from lack of faculty interest and common direction on campus. Within libraries, the pressure of a work force not originally trained to manage data and the need to prepare them to take on these duties is a hurdle that many libraries have already begun to address. Collaborations between libraries at different institutions to initiate grant funding might alleviate the budgetary issues for a short time, but further discussions on sustainable budgetary models will be important for the future.

The investment of resources in e-science even during difficult budget times indicates a strong priority among libraries and institutions. However, the size of the issues involved demands collaboration between institutions and libraries to solve the collective data problems. As Susan Parham of Georgia Tech stated, "This area is very important, but is much larger than a single institution. We need a national framework for addressing the management, re-use and preservation of scientific data."

The Study

Background

As e-science has emerged as a persistent and increasingly large part of the research enterprise, research libraries are exploring new roles, services, staffing, and resources to address the issues arising from this new mode of research. The Association of Research Libraries has recognized the importance of e-science trends and encouraged its members to adopt new roles in this arena.

The ARL E-Science Working Group was formed in 2008 as part of the recommendations from an earlier Joint Task Force on Library Support for E-Science that determined e-science was a continuing issue for research libraries and deserved an ongoing working group. A survey was developed by the E-Science Working Group in answer to their formal charge to "build an understanding of how libraries can contribute to e-science activities in their institution" and "identify organizations and institutions that have similar interests in e-science to leverage research library interests." In addition, the Working Group recruited the authors to deepen the study by expanding the data gathering through the inclusion of case studies of a small group of survey respondents.

For the study, "e-science is defined broadly not only as big computational science, but also team science and networked science. It includes all scientific domains, as well as biomedicine and social sciences that share research approaches with the sciences." This somewhat expansive definition of e-science means that support for e-science necessarily incorporates a broad range of activities and services. Rather than a single service, e-science requires the development, coordination, and synthesis of a range of investments to create a support system. To understand the nature and significance of library contributions to such a multifaceted e-science support system necessitates looking beyond what research libraries are doing to support e-science to the broader institutional and even the inter-institutional context.

Within a broader support system, data curation, preservation, access, and metadata are areas of e-science where libraries find a natural affinity, and consequently these concerns dominated the responses and conversations regarding library contributions. However, a few institutions also provided evidence or commentary on institutional initiatives and activities that reflect the fuller spectrum of the service system. Some respondents also considered broad scholarly communications efforts and digital repository activities as part-and-parcel of e-science efforts.

Method

The Web-based e-science survey was announced to the ARL member libraries in August 2009. Responses were accepted through November 16, 2009. Fifty-seven of the 123 ARL member libraries responded. Twenty-one respondents report their institution provides infrastructure or support services for e-science, 23 institutions are in the planning stages, and 13 do not provide support for e-science. Not every respondent answered every question. Therefore, the percentages relayed in the survey results section of this report are based on the actual number of answers to individual questions rather than as a percentage of the total respondents.

As with most surveys, individuals interpreted questions in a variety of ways. The vocabulary of e-science is still evolving. Research institutions are still settling on terminology, which is reflected in responses to the survey and the way survey questions were interpreted.

After analyzing survey results, six institutions were selected for further study using such factors as size of student population, public vs. private, and level of e-science activity. These institutions were contacted for interviews to further elaborate on their responses to the survey and also to deepen the authors' understanding of the dynamics of developing an e-science support system. Interviews were conducted over the phone with follow up questions answered via e-mail. Highlights of findings from these conversations are included in this report, and cases based on the interviews along with the corresponding institutions' responses to the survey comprise the Detailed Case Studies section of this report.

Findings

Approaches to E-Science in Institutions

A high proportion of survey respondents indicated that their institutions are providing or planning to provide some infrastructure or support services for e-science (44 of 57 respondents or 77%). Institutional approaches to support e-science vary considerably, and institutions are at different stages in developing a response to data-intensive e-science. Some universities are planning to create or already have established formalized institution-wide e-science strategies. They have formed task forces or standing committees to develop policies and central infrastructure and services to support and foster e-science and data. Some institutions primarily see disparate departmental responses emerging, while still others have established hybrid models that provide for both institution-wide and unit-specific efforts. These three patterns of institutional response impact the nature of collaboration and the role of the library in e-science.

Institution-wide and Hybrid Approaches

There are few institutions with a predominantly **institution-wide structure**. Only four respondents (9%) indicated that their institution had or was planning an institution-wide group or task force to advance e-science planning and policy development. Three of these indicated that the groups were composed of staff from information technology, faculty researchers, the office of research, and the library. The other indicated that institution-wide planning and policy development was to be conducted by a body comprised of the CIO and the library.

The University of California, San Diego's Research Cyberinfrastructure Design Team (RCIDT) and their report "Blueprint for the Digital University" illustrate the strength of this approach, as well as the belief that access to a centralized cyberinfrastructure is essential to e-science and the modern research university. The UCSD case study provides more context for the institution's choices and philosophy in pursuing this e-science support strategy. (See page 28.)

Far more often, survey respondents indicated that their institution had or was planning a **hybrid structure** that included both institution-wide and unit-specific efforts (27 or 61%). These institutions reported similar players for the institution-wide groups, including staff from information technology, faculty researchers, the office of research, and the library. The only notable difference was the rate of participation reported for the campus office of research. Only sixteen of the respondents reporting a hybrid structure (59%) indicated that the office of research was a core member of institutional efforts; all other categories of participants were selected by all but a few of these respondents.

A majority of the hybrid institutions (21 of 27 or 78%) indicated that an institutional group with e-science planning and policy responsibility either existed or was being planned. Respondents described central groups that were either temporary task groups or a standing or permanent committees. They also described research labs, centers, and institutes with a campus-wide e-science or cyberinfrastructure mandate.

Such institutional groups were charged with a variety of responsibilities: seeding and providing central resources in support of e-science and research cyberinfrastructure (RCI) activities; developing plans or proposals for the establishment of institution-wide RCI for e-science; and conducting interdisciplinary research into e-science and cyberinfrastructure problems. Many institutional groups take responsibility for distinct components of e-science and research infrastructure such as data management or high performance computing.

For instance, the locus for e-science planning and services at the University of Washington is the UW eScience Institute,[1] an interdisciplinary and institution-wide coordinating body based in the Office of Research. In addition to coordinating eScience efforts across campus, the eScience Institute assists researchers at the university in applying advanced computer science (CS) methods to their domain science problems and provides access to high performance computing (HPC) platforms. Institute staff include eScientists, who have backgrounds in one or more domain sciences along with knowledge and experience in CS and HPC. eScientists also work with key CS faculty to advance the state of the art in key e-science technologies such as databases, data mining, machine learning, sensor networks, visualization, and parallel computing. Library staff are involved in data curation discussions, planning, and referral with UW eScience Institute staff.

The University of Minnesota's Research Cyberinfrastructure Alliance[2] is jointly sponsored by the CIO, the vice provost for research, and the university librarian. Membership is drawn from those organizations, college information technology service organizations with robust research support, and faculty researchers. Their charge is to assess service infrastructure, policy, budget models, and opportunities to leverage collegiate infrastructure for more enterprise implementations. Johns Hopkins had an e-science task force that has been disbanded and can be recalled at the discretion of the provost. Currently, the Institute for Data Intensive Engineering and Science (IDIES)[3] has become the most visible umbrella organization for e-science activities at Johns Hopkins.

In some cases, multiple institution-level groups tackle different aspects of e-science support. At the University of Utah, two institution-wide groups exist: the Cyberinfrastructure Council and the Knowledge Management Committee.[4] The council is involved in high-performance computing, data centers, and other computing and network issues. The Knowledge Management Committee is more oriented to the content of e-science and data curation, leveraging the intersection with the institutional repository and scholarly communications initiatives. Data curation is integral to data center operations and so in this arena both the Cyberinfrastructure Council and the Knowledge Management Committee share responsibility.

At Purdue University, three key centers of e-science research activity exist: the Computing Research Institute, the Rosen Center for Advanced Computing, and the Distributed Data Curation Center. These centers conduct complementary research and investigate and provide high-performance computing, as well as data storage, curation, and preservation expertise and services. The Purdue case study provides more information about these centers and the context within which they work. (See page 24.)

Few institutions are even attempting entirely institution-wide approaches to e-science planning and policy making. Those that seem to be making the most rapid progress in institutional e-science support efforts are utilizing both institution-wide and unit specific responses. However, institution-wide activities seem mostly to provide either only portions of the needed support system or focus on policy and planning concerns at present. Many institutions that have strongly decentralized organizational cultures may be moving even more slowly toward developing the needed support systems across their institution, although some units may do quite well for themselves. Further complicating the situation are the widespread opportunities for collaborative service development that span institutional boundaries.

Unit-by-Unit Approach

Virtually all research institutions have large, grant-funded projects that have developed their own infrastructure on a college, departmental, or unit level. This section focuses on the 11 respondents (25%) who selected "At my institution individual units (i.e., departments, colleges, schools, etc.) develop infrastructure and policies related to their own e-science needs" as the best description for how their institution had organized itself to advance e-science planning and policy development.

These respondents tended to emphasize a strongly decentralized campus culture with the focus of e-science activities in particular subject areas or interdisciplinary institutes. While some survey respondents certainly indicated a strong desire on the part of libraries to participate more fully in e-science efforts on campus, fewer than half indicated that their library was providing infrastructure or support services for e-science. The others said that, "e-science infrastructure or support services are in the planning stages." This pattern, while based on a small number of respondents, suggests that there are special difficulties libraries experience when attempting to collaborate with other units to develop support services on campuses that have a decentralized culture. The interviews of the Massachusetts Institute of Technology and the University of Illinois at Chicago revealed that their efforts to provide e-science support on a decentralized campus have been most effective when staff work with individual researchers. Both campuses found that it is too early to develop a suite of e-science related services, although at the University of Illinois at Chicago, the library has received some preliminary recommendations for how it can assume more leadership in e-research on that campus.

Multi-institutional Grants and Collaboration

While the survey largely focused on institutional activities, there were several indicators in the survey responses of multi-institutional strategies. Cooperation between institutions, including between libraries of different institutions, is becoming increasingly necessary and is likely to become a practical method of addressing common issues in e-science support. Approximately half of the respondents (20 of 41) indicated that their institutions were involved in a collaborative program with another institution in support of e-science. Library involvement in these collaborations was not guaranteed, but was frequent among those filling out the survey. Of those 20 respondents, 16 indicated that the library was involved in some way.

When describing their multi-institutional collaborations, some respondents described their connections with other institutions as a result of joint grant proposals, many of which were NSF DataNet grants. There were equal numbers of respondents (19 or 45%) whose institutions were involved in a DataNet proposal and those whose were not. When providing more details about their DataNet proposals, 11 institutions indicated that more than one institution was involved. When asked if the library was involved in the DataNet proposal, 56% (17 of 30) stated that the library was involved in the proposal. For example, the University of California, San Diego listed their "Datapedia of Science" grant proposal as a multi-institutional project that will provide "an innovative platform for the long-term, scholarly publication and preservation of scientific data." Purdue University made an additional comment in their survey response that was significant: "The relationships developed in preparing this proposal were very beneficial in increasing the visibility of the Libraries on campus and its potential involvement with collaborating on data management issues. The credibility and involvement, especially within science and engineering, increased in many areas." The interviews reveal that Purdue's experience was shared by others, and many libraries benefitted from their DataNet activities with recognition from campus researchers of the contribution the library and librarians can make to e-science and data management, preservation, and curation research activities.

Another good example of collaboration strategies across multiple institutions emerged in the Cornell University and Johns Hopkins University interviews (since both institutions are involved in the Data Conservancy NSF DataNet grant). The Data Conservancy "will develop a framework to more fully understand data practices currently in use and arrive at a model for curation that allows ease of access both within and across disciplines."[5] It is interesting to note the details of how the collaborations between these institutions and others in the project came about. According to Sayeed Choudhury, the initial step for identifying potential partners was strictly an intellectual assessment of which institutions nationally could best contribute to the major objectives of the project (whether Johns Hopkins had any previous history of partnering with that organization or not). After that, existing partnerships, or the work of researchers that had become nationally recognized, were also considered. This theme of relationship building between librarians and researchers is highlighted throughout this report as an activity that was most productive in developing future partnerships and collaborations.

Library participation in grant proposals outside of the NSF DataNet grants was also prevalent and included such interesting proposals as the one from the University of Illinois at Urbana-Champaign, "Digging into Data," a grant focusing on earthquake engineering, and another proposal from the University of Oregon, in which molecular biology faculty requested funding for a bioinformatics center for genomics research that will include space for a librarian. UCSD and its partner institutions received funding from the Library of Congress National Digital Information Infrastructure and Preservation Program (NDIIPP) for the Chronopolis project to build a national center for the management and long-term preservation of digital assets. The UCSD case study provides more detail about this multi-institutional collaboration. (See page 28.)

Another form of collaboration highlighted in the survey was between libraries and iSchools, some of which involved more than one institution. Purdue Libraries and the iSchool at the University of Illinois at Urbana-Champaign were awarded an IMLS National Leadership Grant for the Investigating Data Curation Profiles Across Multiple Research Disciplines project. The Purdue case study provides more information about this project and other research projects. (See page 24.) Many survey respondents made reference to the summer institute on data curation offered by UIUC, including reporting partnerships to develop the institute or further develop internships related to the institute.

Other respondents described cooperation outside of grant proposals that provided some e-science connection across institutions. The Texas Digital Library and the California Digital Library were cited as multi-institutional collaborations that provided infrastructure in support of e-science. Other efforts outside of grant proposals include organized discussions about e-science issues among several institutions.

Data Support and Services in Institutions

Survey respondents with (or planning) institution-wide, hybrid, and unit specific e-science support structures were also asked questions about the development of data support and services at their institutions, such as data curation, data needs and resource assessments, data center(s), and digital lab notebook applications.

The majority of respondents (23 of 42) indicated that there were no designated units to provide data curation and support for scientific research data on their campus. However, 19 institutions (45%) did identify the presence of designated unit(s) for this purpose. The units included data centers, disciplinary informatics centers and institutes, statistical analysis and academic computing centers and services, library data and institutional repositories, digital research and curation centers, campus information technology units, and high-performance computing and cyberinfrastructure centers and institutes.

Twenty-two respondents indicated that they had both central and distributed data centers for research data on their campuses, while 18 institutions described a distributed data centers system. This reflects the distributed nature and culture of decentralization in large research institutions. A single respondent indicated that they had a central institutional data center. Survey comments indicated that some institutions are creating task forces to examine data center consolidation. For instance, although the Massachusetts Institute of Technology is a highly decentralized institution, in recent years several committees have been charged to investigate the creation of a centralized infrastructure for research data storage and processing with distributed funding. One outcome of the groups' work is the Holyoke High Performance Computing Center, which has a scope that goes well beyond campus. Although many institutions have established institutional data centers for e-scientists and data intensive researchers, and "encouraged" them to avail themselves of their services, survey responses did not clarify what incentives or policies were being used to provide such encouragement and enable transition from distributed to central research data center use. However, the Research Data Management and Publishing Support at Cornell Web site[6] provides an excellent window into both the complexity of data management on campuses and the resulting importance of identifying and designating units who support data management and curation for researchers.

The survey also explored institutional digital lab notebook application support. Only five institutions (12%) indicated that their institution supported them (Texas A&M University, Canada Institute for Scientific and Technical Information, Yale University, McGill University, and Purdue University). Support was provided by departmental IT, teaching support services, and most often, by individual researchers who self-support the use in their labs. Twenty-eight percent indicated that their campus did not support digital lab notebook applications. However, 60% of the respondents to this question were not sure whether such support existed at their institution, suggesting a low level of awareness of these activities amongst research libraries. None of the respondents knew of archival services for digital lab notebook applications.

More than a third of the respondents (16 of 42) reported conducting needs assessments relating to data services and more were planning assessment activities. These often proved to be sources of substantial information on variations in needs as well as gaps in existing e-science support services. The University of California, San Diego 2008 Research Cyberinfrastructure Survey, for example, revealed that research data management was their researchers' top need. The UCSD case study provides more information about the needs. (See page 28.) The complete survey results can be found in Appendix D of the report "Blueprint for the Digital University." The University of Illinois at Chicago Library conducted a survey of its researchers in 2009, and some of the primary recommendations were that the library should take a leadership role on campus for e-research/e-science and should also lead the development of a campus-level e-research program. The UIC case study provides additional detail about the survey findings. (See page 42.) Another needs assessment example can be found in the University of Wisconsin's "Summary Report of the Research Data Management Study Group."[7]

Approaches to E-Science in Research Libraries

Approximately 73% of the respondents (29 of 40) indicated the library was involved in e-science support at their institutions. Leadership of these efforts was primarily through a team effort (15 of 31 respondents or 48%) or some combination of individuals, units, and teams working together (13 or 42%).

The names of these groups/teams/committees ranged from the typical e-science team or e-research committee to labels focused on data curation to ones that were clearly a reflection of a particular project or department. The titles of the position designated to lead these teams were also wide ranging, although many were situated high in the library leadership structure. Almost half of the titles listed were at the associate dean level, with the next most popular title being a science librarian. Most groups were formed and began their work in 2008 or 2009, with one having begun as early as 2006.

When asked to describe the library organization for developing e-science plans and programs, only two respondents indicated that there was a specific library unit dedicated to e-science issues. Emory University has the Digital Programs and Services unit within their library, and Johns Hopkins University Library named the Digital Research and Curation Center (DRCC)[8] as their primary unit that provides e-science support. The Johns Hopkins interview clarified how, by creating a separate department in the library and naming a position at the associate dean level to focus on library digital programs, Johns Hopkins University Libraries provided a signal both within the library and externally to the campus of the importance of sustaining and supporting new ways of doing research. (See the case study on page 38 for more details.)

Beyond those two institutions, there were a wide range of survey responses with different levels of formality and a variety of stages of development. For example, on the side of formal and well-developed programs, a respondent from a medium-size, private university described their e-science efforts as follows:

> The Associate Director for Technology works on the long-term strategy for science data curation, including assessment of current needs and appropriate role for the library, as well as the technological infrastructure required; there is a public services committee who are developing expertise in the topic, talking to faculty, developing pilot archiving projects, and teaching one-hour courses on the subject to students. There are also a digital product manager and metadata expert assisting all of this. Overall, lots of people are involved in some way.

On the other side of the spectrum were statements such as, "We are very early in our planning…" and "Informal, evolving structure…" and "Planning in progress to develop data services positions." All of the libraries interviewed have clearly identified e-science and data management activities as strategic priorities and describe how their library's senior administration are actively involved in creating, resourcing, and fostering these activities.

Approximately 87% of the respondents (27 of 31) report that e-science services offered by libraries are provided in collaboration with other units on campus. Of the 26 respondents who provided more detailed information regarding campus units, half cited a working relationship with the campus or centralized IT organization. Six libraries indicated working with the vice provost for research (or similarly named unit) and eight reported working with a variety of individual departments. When asked about the subject disciplines of the individual campus departments, a set of "usual suspects" emerged: biochemistry/molecular biology, biomedical engineering, chemistry, computer science, environmental science, earth science, and health, etc. More unusual subject areas cited included management, education, Latin American studies, and biological anthropology.

A good example of these kinds of library-campus collaborations can be seen at the University of Oregon, where the library's Metadata Services and Digital Projects[9] unit is combining efforts with the Campus Information Services to make an "inventory of science data sets across these departments: biochemistry/molecular biology, biology, biological anthropology, chemistry, computer and information sciences, environmental science, geography, geology, human physiology, physics, and psychology." Another example, at the University of British Columbia, combines the efforts of the library's Institutional Repository and Scholarly Communications with the Office of Research Services. These two groups "are at the beginning stages of exploring how to handle data associated with research. In particular, they are considering the mandate of the Canadian Institutes of Health Research."

Reference and Consultation Services

Consultation services such as identifying data sets, providing access to data, and articulating current standards for organization of data in specific subject areas seem to be a natural fit for subject librarians who provide similar services for other types of information. This perception was confirmed in the survey results, where a clear majority of the 29 respondents provided some combination of the following services for researchers on campus:

Finding relevant data	83%
Developing data management plans	79%
Finding and using available technology infrastructure and tools	76%
Developing tools to assist researchers	76%

A few libraries described more advanced services, such as "archiving and curating relevant data and curating it for long-term preservation and integration across datasets" and "providing curatorial and data stewardship services" as part of data management plans.

Several libraries listed raising awareness as another key activity and have created Web sites dedicated to describing the e-science services they provide. See for example, the Massachusetts Institute of Technology's guide to Data Management and Publishing,[10] Cornell University Library's list of data management services on campus,[11] and the University of Oregon Libraries' description of data services.[12]

Only a few libraries (8 of 30 respondents) provide workshops for faculty regarding e-science issues and several of those were in the planning stages. A couple of the workshops focused on the use of data in geographical information systems while others cited data management, tools, and best practices. However, when the survey asked, "Does your library include policy issues associated with e-science (e.g., open data, compliance with federal agency policies) in its outreach program?" many more answers (17 of 31) revealed connections with faculty regarding open access/ open data, and NIH compliance as a part of their scholarly communication efforts. Most libraries are answering copyright and intellectual property rights questions and are even offering workshops on these topics. The libraries may not see these scholarly communication issues as being connected to e-science, when, in fact, the connection is closer than is realized.

In addition to consultation and reference services, many of the responding libraries (20 of 31) manage or participate in managing technology related to e-science, such as servers for data storage and tools/software for analysis. When asked to provide details, most libraries reported providing servers for data storage, often in the context of a specific project, such as DISCOVER,[13] VIVO,[14] DataStaR,[15] Chronopolis,[16] Harvest Choice,[17] EthicShare,[18] and DataONE.[19]

The case studies included in this report detail the wide variety of approaches, services, and projects in which libraries are actively involved, and in many cases the interviews highlighted how libraries are connecting their programs with faculty. Sayeed Choudhury, at Johns Hopkins University, emphasized designing services to support the stated needs of the faculty. Several of the case study institutions—including Johns Hopkins, Cornell, and Purdue—described interviewing individual faculty to discover those needs. As Gail Steinhart stated, "Gathering information directly from faculty will provide at least two major opportunities for the library: the interviews will allow the library to identify appropriate and significant ways the library can be involved in research, and the interviews with researchers will increase one-on-one conversations, which can lead to further research and grant partnerships."

Staffing E-Science Activities in Libraries

When asked, "Who provides…consultation services to researchers?" more than half of the survey respondents (17 of 29) indicated that they have both individual discipline librarians or staff and dedicated data librarians or specialists taking on these duties. Libraries reported using a mix of strategies to create a workforce with the skills and capacity to provide e-science services and programs. Considering the current economic climate, it is not surprising that most libraries (18 of 28 respondents) are reassigning existing staff or providing training to existing staff as part of an overall strategy to incorporate e-science responsibilities into their current portfolios. In addition to staff reassignments, libraries have hired or plan to hire staff specifically to provide e-science services as part of an overall strategy. Libraries have traditionally combined hiring new staff and re-training current staff to take on new areas of responsibility. This investment of resources for new hires even during widespread budget cuts[20] indicates the level of commitment to e-science services by many of the respondents to the survey.

Survey respondents were asked to provide details for up to three positions that have or will have data management or e-science-type duties as a major part of their portfolio. A total of 65 positions were described, most of which were permanent positions (only four were grant funded or temporary positions). The most popular titles for these positions included the word "data" (31%), while the next most popular title was a subject specialist (20%), closely followed by managers or directors of digital repositories (17%). The variety of approaches and types of positions created to resource e-science and data management activities across the libraries is demonstrated in the study cases.

The value of the MLIS degree in the context of e-science support has been debated, particularly as technology continues to rapidly change the library world. Among the 65 positions that were reported, 64 indicated degree credentials. Of those, 46 (72%) held degrees in library and information science at the master's or PhD level for current or planned positions. Six of these MLIS degrees were paired with a discipline master's, one was paired with a PhD, and one suggested some combination of an MLIS, discipline master's, and discipline PhD. Two library science degrees were at the PhD level or reflected enrollment in a library science PhD program.

Several of the institutions interviewed for case studies provided richer information on educational background, skills, and activities and how these contributed to individuals' success in working on data management. Case study interviews highlighted that valuable assets for this work include: an advanced science or engineering degree, with experience working as a researcher providing additional advantages; some useful IT skills such as programming; and some understanding of typical library concepts such as aspects of collection development including accessibility, preservation, and the application of metadata. A couple of institutions also noted the benefit of having staff with advanced degrees from the institutions in which they now work. Connections made with local faculty during the process of completing coursework for an advanced degree were utilized later when these individuals began working in the library on data management projects. Staff at Johns Hopkins summarized the observation common to the cases studied: "No matter the educational background, one important aspect of working with faculty is becoming a trusted member of their team, and this pressure will remain until enough evidence of successful faculty-library projects have been completed."

Several of the study interviews explored the question of the impact of faculty status of librarians working in data management. Librarians at Purdue cited faculty status as an advantage in that "such scholarship is a performance expectation. They are accustomed to applying their specialized knowledge to research endeavors and leading and collaborating in research projects. Librarians, particularly those in relevant roles, readily contribute their effort to data management and e-science research projects. This helps set the research agenda for the Libraries…" Yet, institutions without faculty status (MIT, Cornell, and Johns Hopkins) indicated that lack of faculty status for librarians largely went unnoticed by their faculty colleagues. Instead, they suggested that librarians with subject expertise, who were able to speak the language of the researcher along with bringing useful expertise to the research project, are key to establishing trust among their non-librarian faculty colleagues. At the University of Illinois at Chicago, librarians are granted faculty status, but there is debate as to the benefit of that status on the effectiveness of working with their researchers.

A vast majority of survey respondents (28 of 31) indicated that library staff were given opportunities to develop skills related to e-science. Of these 28 respondents, 26 indicated that support was provided for staff to attend e-science conferences and meetings. Another popular strategy was to provide in-house workshops and presentations (19 of 28) along with support to attend professional workshops elsewhere (14 of 28).

Pressure Points

Survey respondents most frequently identified three areas as common pressure points: a lack of resources, difficulty acquiring the appropriate staff and expertise to provide e-science and data management or curation services, and the lack of a unifying direction on campus. Although not mentioned as frequently, the lack of infrastructure to handle, preserve, and provide access to data was another area of stress as libraries consider their e-science roles on campus. All of these pressure points are based on the fact that e-science support is placing new demands on libraries that are developing services in this area. While some may argue that data curation can be seen as similar to many activities libraries already perform around collection development, the hardware, software, and expertise needed are often completely new. Thus, the generally difficult economic situation is just one part of the explanation for the frequency of respondents indicating lack of resources as a pressure point.[20]

Acquiring the appropriate staff and expertise to provide e-science services is complicated by the fact that many libraries are uncertain about the kind of expertise needed. There are questions about the appropriate educational background, previous work history, knowledge of computer science and programming, the ability to create good connections with researchers, and the weight that should be given to each of these areas when considering hiring an expert in the area of data management. Again, our current economic climate provides an added pressure to hiring new staff and therefore, resources are being spent on training existing staff to take on some level of service in the area of e-science.

The lack of a unifying direction on campus is often a reflection of the decentralized nature of research in many institutions. As long as faculty are getting their needs met to complete their research and apply for the next grant, a central, unifying focus regarding data management can seem unnecessary. Similarly, seven survey respondents indicated a lack of faculty interest in data issues as a major source of pressure. The University of California, San Diego case study provides a richer exploration of this challenge from their perspective. (See page 28.) Before libraries can play a credible role in e-science and provide data management, curation, and preservation services, there must be an identified need by the campus. Continued connection with faculty about other library services that they see as relevant will provide an avenue for discussions and education around issues regarding data curation, preservation, and access.

The relatively recent emergence of e-science support services from libraries and a desire for more expertise in this area begged the question of information exchange between ARL member libraries. A majority of respondents (48 of 53) indicated that they were willing to participate in an information exchange, but only 18 felt they had enough experience with e-science support to have something worthwhile to offer. Topics of interest were primarily around best practices such as staffing levels, descriptions of projects and services, policies, successes, grant opportunities, and how libraries established expertise on campus.

Observations

After careful review of the survey results and the detailed conversations reflected in the case studies, several common themes emerged that are worth noting.

Collaborations are essential to address even modest support of e-science. The survey revealed successful and frequent collaborations on all levels: between libraries of different institutions, between libraries and the departments they serve, between various departments to address interdisciplinary subject areas, and between institutions. Because the data sets created by modern scientific methods are often very large, the resources required to manage that data must also be extensive. The services to support research often require capabilities from different units. Since these services need to manage costs but also tap diverse expertise, collaborations will continue to be an important method to address the enormity of the challenges posed by e-science.

Faculty interest and institutional support at the administrative level are important for success of library services in this area. Without faculty and institutional engagement, libraries will find themselves preaching about the importance of data curation, preservation, and access without making an impact. Our institutional repositories and the lack of faculty contributions to them are good examples of this phenomenon. We can advocate the benefits of our IRs but without faculty or institutional acceptance, there will be limited use. Many libraries are actively working on enhancing understanding on campus regarding the need for services and infrastructure to manage data sets created from current research. Where campus leadership or significant faculty understanding exist, progress in developing a support system advances more rapidly, as can be seen in the Johns Hopkins case study, where substantial faculty acceptance of data management exists and projects and services are well developed. (See page 38.) At the University of California, San Diego, key campus leadership is working together to bring a level of support and understanding to data management services, allowing them to move forward with less hesitation. Consistently, building support is a process of making connections, understanding institutional culture, and finding success step by step rather than an "all or nothing" situation.

The Master of Library and Information Science degree has a place in this new area of librarianship. Of 64 positions described that provided degree requirements, 46 (72%) listed library and information science at the master's or PhD level that were either in place in current positions or were planned for future positions. Since reassigning staff is a major strategy for resourcing roles in e-science in libraries, it is not too surprising that the MLIS has shown up in the survey responses as a degree found in a majority of data positions.

There is some evidence in recent job postings that the focus on the MLIS is in flux. As can be seen in Appendix II, one recent posting required an "ALA accredited master's degree in library or information science," while another required "demonstrated expertise in data management or information science. This would preferably take the form of direct experience with data curation/management, but could include an MLS/MLIS degree with an emphasis on data management." Interviews with case study participants highlighted success of librarians in e-science services with advanced degrees in science or engineering. Connections between faculty and librarians appeared easier to begin and sustain if the advanced science or engineering degree had been obtained from the institution in which the librarian was now employed. At the same time, the cooperative and team-oriented values of most MLIS degree holders, along with the understanding of the role and importance of metadata and preservation, makes the combination of the MLIS and an advanced science degree a useful and successful combination.

The fact that investments in e-science activities are being made even during difficult budget times demonstrates that this is a priority for libraries. Research libraries strive to increase their relevance to the institutions they serve. As research activities become more data intensive, and as faculty and institutions become increasingly concerned about the preservation and access to that data, libraries have an opportunity to demonstrate their expertise and relevance to their institutions and should take advantage of opportunities to move forward into data management activities. There is a parallel between information resources management (our traditional role) and data management. Collecting and preserving information for researchers is a recognized role of libraries and, therefore, libraries are obvious partners to provide the expertise to assist the university with a centralized plan for data management.

There is some concern that if libraries do not act quickly, others (publishers or vendors) will collect the data and then charge universities high fees to get it back. Other concerns are that a lack of quick action by libraries will lead to a rapid loss of relevance. Research libraries are positioned to step up to the opportunities to move forward as e-science becomes a usual part of research practices and as faculty and universities recognize the need to create structures to curate, preserve, and provide access to the results of that research.

Strategies for data curation, management, and preservation are still young and evolving. The survey revealed a wide variety of services and level of involvement with campus data management. Responses ranged from, "This work is in its early stages" to "These units are collaborating on the digital-curation project planning team." Several indicated that surveys were being conducted on campus to gain a better understanding of the data needs of their researchers. Again, the diversity of response is a reflection of the diversity of research, organizational culture and campus acceptance of the importance of data management. DataNet grants through the National Science Foundation, as well as other externally and internally funded programs, will provide substantial models and information that will help guide decisions over the next few years.

Conclusion

The results of this survey indicate that engagement by research libraries in e-science has been developing rapidly in the past few years. This has ranged from answering basic questions about metadata and open access standards to providing infrastructure for curating and managing large datasets. Institutions are quickly rising to meet the challenge of managing data, especially in light of the anticipated federal government requirements for data plans as part of grant proposals. There is great diversity in the strategies employed by institutions to address the needs of their researchers. Current strategies range from a decentralized series of data support services in a variety of departments or units to the creation of committees to discuss campus data needs and services along with the creation of centralized data centers to provide that support. The diversity of response reflects the needs and culture of the institutions, which is to be expected. The trend toward centralization of these services through campus committees and the development of central data centers reflects the growing understanding of researcher needs and an appreciation for benefits that can only be gained through a centralized service.

Collectively among the respondents, there is a more sophisticated view of data management skills, services, and resources, and promising strategies for engagement in these activities are becoming clear and are often highly collaborative. The Cornell and Johns Hopkins University case studies provide some models of the development of collaborations on campus and between institutions along with the development of services to support the data needs of researchers. (See page 32 and page 38.) For example, Cornell University Library provides their researchers with a compilation of data services in the library and on campus, including the policies of other organizations that collect and maintain data sets. This is an important and relatively easy first point of service, which utilizes the obvious strengths librarians bring to the data problem, namely organizing what types of services and information are available to their researchers just as we do currently for all kinds of other collections. It is immediately useful to researchers on campus and highlights the library in data management activities. Many respondents to the survey seem to have made substantial progress toward achieving e-science support systems. However, there are still substantial gaps in their support capabilities. Institution-wide activities are often policy or planning-focused, with service delivery still somewhat fragmented and underutilized.

The key challenges facing research libraries as they consider this new area of service to their communities are both tangible and social in nature. Many survey respondents cited lack of money and resources as the most obvious limitations to quick mobilization around these issues. However, just as compelling were pressures from lack of faculty interest and common direction on campus. Within libraries, the pressure of a work force not originally trained to manage data and the need to prepare them to take on these duties is a hurdle that many libraries have already begun to address. Collaborations between libraries at different institutions to initiate grant funding might alleviate the budgetary issues for a short time, but further discussions on sustainable budgetary models will be important for the future.

The investment of resources in e-science even during difficult budget times indicates a strong priority among libraries and institutions. However, the size of the issues involved demands collaboration between institutions and libraries to solve the collective data problems. As Susan Parham of Georgia Tech stated, "This area is very important, but is much larger than a single institution. We need a national framework for addressing the management, re-use, and preservation of scientific data."

While libraries identified a significant list of pressure points, an overall enthusiasm for new roles in the academic research process was evident throughout the survey responses and in the case studies. This enthusiasm stems from a desire to remain relevant to our institutions, which often depends upon our capacity to move quickly to support the needs of our researchers. Continued sharing of information, successes, and service and policy models will enhance the ability of libraries to rise to this current challenge.

Endnotes

[1] University of Washington, eScience Institute
http://escience.washington.edu/

[2] University of Minnesota, Research Cyberinfrastructure Alliance
http://sspu-test.oit.umn.edu/rca/

[3] Johns Hopkins University, Institute for Data Intensive Engineering and Science (IDIES)
http://idies.jhu.edu

[4] University of Utah, Cyberinfrastructure Council
http://www.it.utah.edu/leadership/committees/Cyber/index.html

[5] Johns Hopkins University, Data Conservancy press release
http://www.library.jhu.edu/about/news/releases/pressrel09/nsfgrant.html

[6] Cornell University, Research Data Management and Publishing Support at Cornell
https://confluence.cornell.edu/display/datasupp/Home

[7] University of Wisconsin, Summary Report of the Research Data Management Study Group
http://minds.wisconsin.edu/handle/1793/34859

[8] Johns Hopkins University, Digital Research and Curation Center
http://ldp.library.jhu.edu/dkc

[9] University of Oregon, Metadata Services and Digital Projects
http://libweb.uoregon.edu/catdept/home/

[10] Massachusetts Institute of Technology, Data Management and Publishing
http://libraries.mit.edu/guides/subjects/data-management/

[11] Cornell University, Research Data Management and Publishing Support at Cornell
https://confluence.cornell.edu/display/datasupp/

[12] University of Oregon, Science Data Services
http://libweb.uoregon.edu/faculty/SciDataInfo.html

[13] Cornell University, DISCOVER Research Service Group
http://arecibo.tc.cornell.edu/DRSG/Default.aspx

[14] VIVO, Research & Expertise Across Cornell
http://vivo.cornell.edu/

[15] Cornell University, DataStaR, Data Staging Repository
http://datastar.mannlib.cornell.edu/

[16] San Diego Supercomputer Center, Chronopolis
http://chronopolis.sdsc.edu/

[17] HarvestChoice
www.harvestchoice.org

[18] EthicShare
www.ethicshare.org

[19] DataONE
http://dataone.org/

[20] Association of Research Libraries. Tracking the Economic Crisis.
http://www.arl.org/rtl/plan/econ/index.shtml

All URLs accessed July 30, 2010

Detailed Case Studies

After analyzing the survey results, the authors contacted Purdue University, the University of California, San Diego, Cornell University, Johns Hopkins University, the University of Illinois at Chicago, and the Massachusetts Institute of Technology for interviews to further elaborate on their responses to the survey. These were selected in an attempt to provide a variety of types of institutions using such factors as size of student population, public vs. private, and level of e-science activity. Interviews were conducted over the phone with follow-up questions answered via e-mail. Due to time constraints only six institutions were selected for this portion of the report, thus leaving out many others who would have qualified if time had allowed.

Following the interviews, the authors developed six cases studies that synthesize the interviews with the corresponding institutions' responses to the survey. The case studies further illuminate programs and services mentioned only briefly in the survey and allowed some interesting patterns to emerge as individuals had the opportunity to speak about faculty connections, staffing levels, and organizational structure and culture. One pattern included the importance of faculty-librarian connections as a big part of the process of identifying research needs and ways the library can provide solutions. It is imperative to place an emphasis on face-to-face connections with our faculty, as it was clear from the case studies that what might appear to be a minor connection might result in further collaborations. Another theme that emerged was that organizational culture and the implementation of computing on campus contributed to the variety of approaches to successful partnerships with faculty research teams. For example, the University of California, San Diego Libraries have benefitted from the implementation and existence of the San Diego Supercomputing Center in that the libraries have strong connections with campus IT efforts. The campus itself has a very decentralized culture, so there is now an effort to bring the benefits of centralized computing to the attention of campus administrators.

Case Study: Purdue University

Background

Purdue University, located in West Lafayette, Indiana, is a public, land-grant institution founded in 1869, with a current student population of about 40,000. Purdue had a total research expenditure of $502.8 million in 2008–2009 including $49.5 million sponsored by NSF.

The content of this case study was derived from the institution's response to the Fall 2009 ARL Survey on E-science and Data Support plus a subsequent telephone interview and e-mail correspondence with James L. Mullins, Dean of Libraries, and D. Scott Brandt, Associate Dean for Research at Purdue Libraries.

Structure for Institutional Planning and Policy Development for E-Science

At the time of the survey, Purdue University was planning a hybrid structure with both institution-wide and unit-specific efforts to advance e-science planning and policy development. There was no one central group focusing on overall planning, but the provost had signaled the intention to convene a high-level task force to lead the university in its efforts to assess the challenges of data management in support of e-science. The task force is to include the Office of the Vice-President for Research, Information Technology at Purdue (ITaP), the libraries, the Provost's Office and the Colleges. Unfortunately, due to the departure of the provost, at the time of the follow-up interview the task force had not yet been convened. The libraries, ITaP, and the vice-provost for research are addressing the NSF Data Management Plan "requirement," with ITaP providing storage and management expertise, and the libraries providing expertise on curation and the application of discipline specific standards for data discovery, access, and archiving.

Purdue University has a highly decentralized culture, and currently e-science planning and policy development emerges from a collection of overlapping initiatives and activities. It has been essential for the Libraries to learn how to navigate this environment and to develop partnerships and relationships in order to integrate their research efforts and services into campus research initiatives. To advance the university's strategic priority of interdisciplinary research, almost a decade ago a major investment was made in the Discovery Park, which created facilities and space to foster interdisciplinary research and bring researchers out of their silos. The libraries seized the opportunity in 2004 and established an interdisciplinary research librarian position to integrate their activities with those of the Discovery Park, positioning them well as e-science evolved on campus.

Unit Specific

Three key centers of e-science research activity providing research into and support for e-science on campus were highlighted in the survey response:

Computing Research Institute (CRI)
Rosen Center for Advanced Computing (RCAC)
Distributed Data Curation Center (D2C2)

CRI's mission is to facilitate multidisciplinary research in high-performance computing at Purdue. It is a part of Purdue University's Cyber Center, whose vision is to engage in basic cyberinfrastructure (CI) research, and develop new CI tools and techniques for dissemination at Purdue and beyond. The libraries will be participating in one of their major NSF projects, the creation of the Network for Earthquake Engineering Simulation Community and Communications Center (NEEScomm Center) at Purdue, by exploring data curation. They are also collaborating with this group on another NSF proposal to investigate tools for embedding data curation into the research workflow.

The RCAC is the research branch of Purdue's ITaP (Information Technology at Purdue) organization and provides advanced computational and data storage resources and services to support Purdue faculty and staff researchers. It has evolved to provide computer science expertise for other disciplines to help resolve advanced computation needs. The RCAC also conducts its own research and development in order to continually improve the capacity and functionality of its resources. RCAC coordinates HUBzero development, described in further detail below.

The D2C2 is the data curation research arm of the libraries. Its primary role is to explore and conduct research into distributed data curation. It was established as a research center and receives input from an interdisciplinary board, and as such aligns itself well with the campus framework for research. This structure helps to dispel concerns that may arise from traditional perceptions of the library's role in research. The D2C2 collaborates with campus researchers from the grant writing stage, participating as partners in the development and exploration of innovative data curation solutions and methodologies. The D2C2's research provides input into R&D for repository services to deposit and disseminate research data. They actively bring together interested parties into project teams to apply for grants and promote collaboration on advancing solutions for data management. Please see the D2C2 Web site for a current list of projects, publications, and presentations.

It is important to note that both central and distributed data centers exist across Purdue University, and many departments, labs, and centers have established their own approaches to providing data access and retention services for specific communities.

Research Activities and Collaborations

Like several of the survey respondents, Purdue University Libraries were involved in the development of NSF DataNet proposals. For the first round, the libraries led a proposal addressing environmental data in collaboration with Purdue's Cyber Center and an engineering group working with data for hydrologic and hydraulic modeling. The preparation of this proposal was very significant for raising the awareness of the importance of data curation for e-science, for the recognition it established on campus of the library's expertise and potential for collaboration on data management issues, and for the relationships it fostered with researchers and disciplines across campus. It was also a driver for the creation of the Distributed Data Curation Center in the library. A team of researchers from across campus worked with the PI, James L. Mullins, Dean of Libraries, and the library was seen as the partner positioned to lead the project, due to their disciplinary perspective and ability to "see the big picture" that data curation requires. Unfortunately, although the submission received a good rating, this particular proposal did not proceed to the second round. The library did participate in the second NSF DataNet round in the capacity of co-PI with computer science proposing a proteomic data hub, but it too was unsuccessful. Purdue Libraries came to realize through the DataNet proposal efforts that they did not have the internal structure or capacity to manage such major research projects. While recognizing the important contribution the DataNet projects will make, Purdue Libraries has shifted its attention to working with the Purdue University community of researchers on projects that can have impact on broader research networks and collaborations within the research library community.

Other research and data curation activities of Purdue Libraries and its D2C2 center are numerous. Some of the key multidisciplinary and multi-institutional projects and collaborations will be described briefly below.

An IMLS National Leadership Grant was awarded to Purdue's D2C2 and the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign for a collaborative project entitled Investigating Data Curation Profiles Across Multiple Research Disciplines. The project addresses the question, "Which researchers are willing to share data, when, with whom, and under what conditions?" UIUC partners are adapting the profiles to work in the Data Conservancy, a DataNet award led by Johns Hopkins, and Purdue has been working with other universities who are interested in using the profiles to better understand data collections on their campuses. Purdue has been awarded a subsequent Laura Bush Twentieth Century Award for developing a series of workshops to teach librarians across the United States how to use the profiles.

Other data curation related projects in which the Purdue Libraries are co-PIs include collaborating with RCAC in a multidisciplinary NSF grant awarded for INTEROP: Developing Community-based DRought Information Network Protocols and Tools for Multidisciplinary Regional Scale Applications (DRInet), and with Earth & Atmospheric Sciences for Enabling End-to-End Geospatial Data Modeling Workflows via INPort: The Isotope Networks Portal. The Center for the Environment at Discovery Park provided a local seed grant for Ingest, Preservation and Access for Water Quality Datasets in an Institutional Repository.

Purdue Libraries is a founding member of the international consortium DataCite – International Initiative to Facilitate Access to Research Data that is establishing a registry of research datasets allowing for persistent identification through a digital object identifier (DOI) standard. Purdue is developing a service for persistence of data to support research dissemination and publishing.

Purdue University has an international program and partnership with Moi University in Kenya. The Purdue Libraries are collaborating on a project with Moi University that investigates water data curation and trains Moi University librarians in the use of the Data Curation Profile to collect information on water quality research. They are also helping to design a repository that can accommodate data collection and the infrastructure challenges they face.

Purdue's nanoHUB, the Hub Technology Group, and their HUBzero platform for scientific collaboration, represent another significant area of collaboration for Purdue Libraries. The Hub Consortium currently includes Purdue, Indiana University, Clemson University, and the University of Wisconsin-Madison, and the platform will be released as an open source project in the spring of 2010. Purdue Libraries have contributed to this project by assisting with the implementation of a protocol for metadata harvesting (OAI-PMH) and a Handle System to provide persistence of Hub objects and COinS for Hub resource citation. Currently, they are investigating the object reuse and exchange standard (OAI-ORE) to make HUBzero interoperable and aggregate collections of Hub objects, and are implementing DataCite digital object identifiers (DOI) for Hub resource citations with nanoHUB. They have initiated a linked data project to experiment with exposing hub data to semantic Web applications and are participating in a HUBzero project entitled Developing a Content Organization Framework for Healthcare Delivery Hub, which involves interviewing researchers about their needs related to depositing tools and developing an organizational framework that will facilitate the management, description, and discovery of healthcare delivery.

While Brandt served a fellowship with the Office of the VP for Research (OVPR), a survey was conducted to identify policy, guidelines, and practice for research record retention. The OVPR endeavoured to raise researchers' awareness of the university's obligation for data retention and to implement supportive practices. As a result, Purdue Libraries have been working with the College of Agriculture on research record retention and management issues. The libraries will provide expertise for data management planning, the establishment of data repositories, and to support data dissemination and preservation.

Evolving Liaison Roles

In their survey response, Purdue Libraries indicated that they participate in reference and consultation activities that assist scholars and researchers with the identification, access, and use of data, which includes assistance and provision of technological infrastructure and tools, finding relevant data, and development of data management plans and tools. This is seen as an evolution in the activities of disciplinary liaison librarians who provide these consultations. As an information literacy initiative, workshops are being developed for graduate students' development of research data management skills ("data curation literacy"). The success of D2C2 and consultation or outreach activities creates so many opportunities that the key to their continued success is in accurately assessing projects for fit with the center's vision and the ability to resource the resulting projects with external funding. Purdue Libraries have further contributed to the education of librarians by collaborating with UIUC GSLIS' Data Summer Institute, giving three presentations at the first two institutes.

Resourcing E-Science Activities in the Libraries

Five years ago, Purdue undertook an interdisciplinary research initiative that explored interest in research collaborations. The result was an overwhelming affirmation by faculty across campus that library science faculty are valuable partners in research. It also identified the need to create an associate dean for research position in the libraries as a peer to positions within the colleges that facilitate, support, and liaise with the university in the area of intra- and extramural research and funding. Library administration has made interdisciplinary research including e-science one of the libraries' key strategic goals. The primary sources of funding for these activities are research grants and cost sharing through the contribution of research effort or time to projects. In addition, it created a research council to oversee and develop policy related to research, as well as provide support and promotion of research. An example of the latter is an annual event titled Celebrating Research, in which libraries' faculty present an overview of recent and ongoing research to colleagues to demonstrate collaborations and further emphasize this important role for librarians.

Three years ago, Purdue Libraries created a full-time data research scientist position and based the job definition on recommendations contained in the 2006 ARL report to the NSF entitled "To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering." At the time this was unprecedented. This initial position was only partially funded with library funding, with a majority of the salary to come from grants. It was not defined as a librarian tenure track position to avoid being encumbered by responsibilities for existing library operations, and thus more closely resembled a research faculty appointment. The associate dean for research makes a significant and direct contribution as the acting director of the D2C2. Library leadership for e-science planning and programs is led by the libraries' administrative team and the D2C2 with input from its board.

In 2007, the libraries created the interdisciplinary research librarian position to provide liaison in Discovery Park, and filled it with a former systems librarian, Michael Witt, who brought technology skills with the application of library science to other discipline's research. Witt was recently named a Fulbright Scholar. This highly regarded Department of State/CIES scholarship will allow Witt to participate in an international education program to do research and instruction related to data and resource management at the Bibliotheca Alexandria in Egypt.

Disciplinary liaison librarians also participate in projects, support e-science initiatives and programs, and provide consultations or make appropriate referrals. Purdue librarians are tenure track faculty, and as such scholarship is a performance expectation. They are accustomed to applying their specialized knowledge to research endeavors and leading and collaborating in research projects. Librarians, particularly those in relevant roles, readily contribute their effort to data management and e-science research projects. This helps set the research agenda for the Libraries and supports the efforts of the D2C2.

The programs and research activities of the D2C2 are producing results and tools that enable the capacity of librarians to evolve their roles in relation to research. The Data Curation Profiles are an example of a tool that can help bring disciplinary understandings of data management into focus. Several projects have highlighted the importance of controlled vocabularies and broadened their appeal and application beyond the traditionally limited terrain of cataloguers or metadata experts. Librarians can now see that organizing and retrieving data is really not all that different from more traditional information resources; this does not require a paradigm shift. They also realize that researchers are looking to librarians for expertise in this area. This translation and integration into Purdue Libraries' service is being reflected in the work of an internal task force that is reviewing and rewriting liaison librarian position descriptions. Taking their concept of librarianship well beyond a primary identification with reference and bibliographer roles into the new areas involved in data curation, among other emerging roles, reflects the changing context within research institutions. Liaison responsibilities will include support for e-science, and professional development activities are being planned for skill enhancement in this area.

As further evidence of the recognition of Purdue Libraries' participation in this area, Brandt was invited to serve on the Advisory Board for the IMLS-funded University of North Carolina initiative, Closing the Digital Curation Gap, "designed to serve as a locus of interaction between those doing leading edge digital curation research, development, teaching, and training in academic and practitioner communities and those with a professional interest in applying viable innovations within particular organizational contexts." Brandt was also asked at the 2010 Bloomsbury Post-Conference Workshop to co-lead, with Mike Furlough (Pennsylvania State), an IMLS collaborative planning grant to support an international workshop on digital curation and publishing.

Highlighted Resources

Discovery Park: http://www.purdue.edu/dp/
Computing Research Institute (CRI): http://www.purdue.edu/dp/cri
Rosen Center for Advanced Computing (RCAC): http://www.rcac.purdue.edu/
Distributed Data Curation Center (D2C2): http://d2c2.lib.purdue.edu
Cyber Center: http://www.purdue.edu/discoverypark/cyber/about/index.php
Investigating Data Curation Profiles Across Multiple Research Disciplines: http://www.datacurationprofiles.org
DataCite: http://www.datacite.org/

Case Study: The University of California, San Diego

Background

The University of California, San Diego (UCSD), located in La Jolla, California, is a public institution founded in 1960 with a current student population of about 30,000. UCSD had a total research expenditure of $881.6 million in 2008–2009 including $90.6 million sponsored by NSF.

The content of this case study is derived from the institution's response to the Fall 2009 ARL Survey on E-science and Data Support plus a subsequent telephone interview and e-mail correspondence with Luc Declerck, Associate University Librarian, Technology Services, and Ardys Kozbial, Technology Outreach Librarian, both in the University of California, San Diego Libraries; and David Minor, Head of Curation Services, San Diego Supercomputer Center.

Structure for Institutional Planning and Policy Development for E-Science

At the time of the survey, UCSD indicated that its institution had or was planning an institution-wide structure (such as a group or task force) to advance e-science planning and policy development. In April of 2009, the Research Cyberinfrastructure Design Team (RCIDT) released a report entitled "Blueprint for the Digital University: A Report of the UCSD Research Cyberinfrastructure Design Team." The UCSD Libraries figured prominently in the design team charge issued by the vice-chancellor for research and participated as core members alongside Administrative Computing and Telecommunications (ACT), the San Diego Supercomputer Center (SDSC), Academic Computing Services, the California Institute for Telecommunications and Information Technology (Calit2), and individual researchers, as well as representatives from laboratories, departments, or schools across campus.

The Blueprint proposes the establishment of UCSD research cyberinfrastructure (RCI) to be implemented through a partnership among SDSC, the UCSD Libraries, ACT, and Calit2 with five core components: co-location facilities, centralized data storage, digital curation and data services, a RCI network for high performance computing, and condo clusters. A business plan for the proposed UCSD RCI model is currently being developed for the chancellor and vice-chancellors.

Data Support and Services

"The San Diego Supercomputer Center (SDSC) enables international science and engineering discoveries through advances in computational science and data-intensive, high-performance computing." While UCSD has both central and distributed data centers, SDSC's new state-of-the-art central regional data center is, in and of itself, a strong incentive for researchers to take advantage of this key central service. The center's infrastructure and its business models have been designed to reflect UC researchers' expressed needs and to scale to accommodate new customers as they come on board. SDSC provides expertise in, and access to, high-performance computing (HPC), data management, storage, and curation. The UCSD Libraries provide expertise in long term preservation, use, and dissemination of data, building upon the work of the Metadata Analysis and Specification Unit, the Information Technology Department, and the Digital Library Program (DLP), which focus on the curation of digital assets and collections. The UCSD Libraries interact with faculty to gather and assess their data sets for long-term preservation and use, while the SDSC provides HPC and data storage and curation services.

The libraries and SDSC have been pursuing joint goals for campus infrastructure and data management for a number of years under the leadership of the Audrey Geisel University Librarian Brian E. C. Schottlaender and former SDSC Director Fran Berman. As a result, the two organizations have developed a strong collaborative relationship over time. They work together in a very integrated fashion to advise researchers from data development and creation right through the data lifecycle. This integration is embodied in their liaisons, the SDSC head of curation services and the libraries' technology outreach librarian, who work closely on data curation and lifecycle management as well as the Chronopolis program. They collaborate in response to researchers' immediate needs and requests for assistance with data creation, storage, curation, discovery, reuse, and long-term preservation, and advise them on policies and procedures. These activities are leading to the development of new data services but are currently primarily project driven.

The California Institute for Telecommunications and Information Technology (CalIt2) is playing two roles with respect to data initiatives on campus. First, they will be contributing work toward the creation of a portal where access to researcher expertise and content will be exposed. This will be an important part of the data lifecycle, helping to make data available to others. Second, they are leading the campus efforts in creating a "green cyberinfrastructure," which will help to keep the costs associated with data infrastructure low.

The cyberinfrastructure needs assessment conducted by the UCSD Research Cyberinfrastructure Design Team in 2008 revealed that data management was the researchers' top need. The entire stack of data management resources and services were requested, from data back up to collection management, and long-term preservation. Over 80% of the respondents cited data backup as their principal data need; 70% cited the need to store and analyze large quantities of research data; 64% expressed a need for long-term data preservation. Roughly 50% of the respondents expressed concern about the ability to move their research data from where it is generated to their desktops where it is analyzed, and the ability to share research data with others. (page 13 of Blueprint.)

If the proposed UCSD RCI model is implemented, the UCSD Libraries will be the lead partner responsible for operating campus digital curation and data services, the Research Data Depot, in collaboration with SDSC. The Research Data Depot services will include: data curation, data discovery and integration, and data analysis and visualization.

Research Activities and Collaborations

SDSC and the UCSD Libraries, with their partners at the National Center for Atmospheric Research (NCAR) and University of Maryland's Institute for Advanced Computer Studies (UMIACS), currently collaborate on the Chronopolis program, building a national center for the management and long-term preservation of digital assets. UCSD Audrey Geisel University Librarian Brian Schottlaender serves as Chronopolis' principal investigator. The program is funded by the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP). Chronopolis has created a trusted preservation environment that spans academic institutions and research projects with the goal of long-term collection management, preservation, and knowledge generation. Chronopolis is based at UCSD and is managed by the UCSD Libraries and the San Diego Supercomputer Center. Both NCAR and UMIACS contribute to the technical foundations of Chronopolis with their expertise in grid computing. The California Digital Library (CDL) contributes to Chronopolis as a data provider. The EDUCAUSE Center for Applied Research (ECAR) has recently published a case study by Judith A. Pirani and Donald Z. Spicer on the Chronopolis project, which is a great resource for more information about this collaborative program.

Like several of the survey respondents, the UCSD Libraries were involved in the development of a NSF DataNet proposal. The multi-institutional proposal was led by SDSC and proposed the creation of a "Datapedia of Science," an innovative platform for the long-term scholarly publication and preservation of scientific data. The goal was to provide standardized handling of scientific data to collaboratively engage researchers and promote scientific progress. Although this proposal was not funded, the organizational relationships and community of interest developed through this work have been maintained and new collaborative projects continue to arise resulting from this collaborative effort.

The UCSD Libraries are also involved in University of California system-wide data curation activities in collaboration with other UC campuses and the University of California Curation Center (UC3) which is based at the California Digital Library (CDL). UC3, established in 2009, is a collaborative partnership that brings together the expertise and resources of the University of California to provide a central preservation and curation service. UC3's stated mission is to

> "... provide high quality and cost-effective solutions that enable campus constituencies — museums, libraries, archives, academic departments, research units, and individual researchers — to have direct control over the management, curation, and preservation of the information resources underpinning their scholarly activities."

UC3 leverages UC and CDL community resources and projects such as Chronopolis, DataCite, and NSF DataNet DataONE to help provide data curation, management, and preservation services to researchers across UC and to further the digital data curation research agenda.

The UCSD Libraries are learning from and building upon the e-science and data curation activities at non-UC research libraries. They have been heavily informed by the Purdue-University of Illinois at Urbana-Champaign Data Curation Profiles

Project, reviewing the first draft of that group's data curation profiles and planning to participate as a beta partner in the investigation of methods of operationalizing the profiles. The UCSD Libraries charged its own Data Curation Task Force in 2009 with investigating possibilities for data curation services on the UCSD campus. In addition to information gathering, the task force created a UCSD version of the Purdue-UIUC profile which it is currently sharing with Purdue-UIUC. The UCSD task force disbanded in early 2010 in order to combine efforts with UC3. USCD's Chronopolis is collaborating with the MetaArchive Cooperative on the technical side, building a solid foundation for a national preservation environment.

Consultation Services

In their survey response, the UCSD Libraries indicated that they are planning to or are participating in reference and consultation activities that assist scholars and researchers with the identification, access, and use of data, including assistance and provision of technological infrastructure and tools, finding relevant data, and the development of data management plans and tools. These activities are conducted by a team of metadata specialists and discipline librarians, lead by UCSD's technology outreach librarian and SDSC's head of curation services.

Resourcing E-Science Activities in the Libraries

Library leadership responsibility for plans and programs for e-science support is provided by the Library's Administrative Team. Primary strategic leadership is provided by University Librarian Brian Schottlaender and Luc Declerck, the Associate University Librarian, Technology Services. The UCSD Libraries see e-science and data management and curation as a core part of the mandate of research libraries now and in the future. The primary sources of funding for these activities are research grants and cost-sharing through the contribution of research effort or time to projects.

The UCSD Libraries' data curation initiatives operate in a matrix management environment. Staff key to the development and implementation of e-science and data curation activities report through multiple library units (DLP, Metadata Analysis and Specification Unit, and disciplinary departments). The UCSD technology outreach librarian is the only full-time employee dedicated to these efforts. She reports to the AUL, Technology Services. The head of curation services is also dedicated to this work and is a librarian, and reports through SDSC. The data services librarian, reporting to the Social Science and Humanities Library department head, has been partially reassigned to support these efforts. The head of the DLP and two metadata specialists also support these activities. It is important to note that librarians from across the system are seconded to task-groups and projects and that the UCSD Libraries are working to integrate data into the mainstream of librarian responsibilities and involve them in the future Research Data Depot.

Pressure Points

Near its conclusion, the survey asked an open-ended question regarding the pressure points experienced by the library. UCSD's response was very straightforward and informative:

"At present, there are three primary pressure points related to e-science/e-research support at UCSD: turf, money, and interest. In reverse order, with the exception of a few very high-end data generators amongst the faculty, e-science/research lifecycle management is not high on the list of faculty concerns. The NSF's best efforts notwithstanding, most researchers, at least locally, have been slow to wake to the data challenge. They seem to think, to the extent they think about it at all, that they've already got it covered or that they lack the funds to cover it and, therefore, it should be somebody else's problem. As a consequence, this campus at least, has committed funds for providing the infrastructure and services necessary to curate data for the long-term, in the hope, frankly, that sufficient faculty (and students, but mostly faculty) will avail themselves of both to make the enterprise self-sustaining. That's the good news; the bad news is that it has committed only those funds and only with the understanding that the enterprise will become self-sustaining. Whether that proves to be the case remains to be seen of course. Finally, there is an awfully large number of parties interested in what remains a still-ill-defined problem space. The associated 'jostling' makes calculating the right mix of those parties in the solution space doubly challenging."

In the follow-up interview, Luc Declerck elaborated on the three pressure points identified: faculty interest, finances and sustainability, and competing interests. Declerck recognized that although the need for data preservation and curation is not at the top of many researchers' priorities, once it is introduced they do indeed see the significance, and data loss is of great concern to them.

Sustainability and funding are of concern not only because we live in times of fiscal constraint and downturn but also because, on the whole, preservation and curation are not "sexy." It is difficult to make the value-add proposition readily apparent and appealing, and it is not clear who is responsible for data preservation. Is it the university's responsibility, the responsibility of the funding agencies, or, is it a disciplinary responsibility? A follow-up committee to the UCSD research cyberinfrastructure design team (RCIDT) has been asked to prepare a business-case with the requirement of self-sustainability within 2–3 years.

Finally, competing interests and questions of turf present challenges and take time to overcome. Although it has been noted that the libraries and SDSC have had a long-term relationship of collaboration, there are many players in the e-science and research cyberinfrastructure arena. UCSD does not have a Chief Information Officer or any other central coordinating body to set central IT policy. The participants in the RCIDT worked together for over a year to develop the Blueprint, their comprehensive RCI plan, and there will be new relationships to forge as it is implemented considering the distributed nature of e-science and data infrastructure, expertise, and services currently.

Highlighted Resource

Blueprint for the Digital University: A Report of the UCSD Research Cyberinfrastructure Design*:* http://research.ucsd.edu/documents/rcidt/RCIDTReportFinal2009.pdf

San Diego Supercomputer Center (SDSC): http://www.sdsc.edu/about/About.html

California Institute for Telecommunications and Information Technology (CalIt2): http://www.calit2.net/index.php

Chronopolis Digital Preservation Demonstration Project: http://chronopolis.sdsc.edu/

University of California Curation Center: http://www.cdlib.org/services/uc3/

Office of Contract and Grant Administration, University of California, San Diego: http://ocga.ucsd.edu/OCGA/Annual_Reports/2009/5_Awards_by_Major_Agency.pdf

Case Study: Cornell University

Background

Cornell University, located in Ithaca, New York, is the federal land-grant institution of New York State, a private endowed university, a member of the Ivy League/Ancient Eight, and a partner of the State University of New York. Cornell was established in 1865 and has a current student population of approximately 20,000.

The content of this case study was derived from the institution's response to the Fall 2009 ARL Survey on E-science and Data Support plus a subsequent telephone interview and e-mail correspondence with Gail Steinhart, Research Data & Environmental Sciences Librarian. Additional e-mail correspondence with Medha Devare, Bioinformatics and Life Sciences Librarian, Dianne Dietrich, Research Data & Metadata Librarian, and Dean Krafft, Chief Technology Strategist, provided context to several of the programs and services offered by Cornell University Library.

Structure of Response to E-Science

Cornell's responses to questions on the survey indicated they organize their overall response to e-science through a hybrid structure that includes both institution-wide and unit-specific efforts.

Institution-wide

One of the main efforts on campus is the DISCOVER project, which is, according to its Web site, "a partnership between domain scientists, the Center for Advanced Computing (CAC), the Cornell University Library, and Fedora Commons." DISCOVER is supported by the Office of the Vice Provost for Research and is developing the technology infrastructure and the creation of tools to develop a single point of contact for the analysis of data across disciplines. The DISCOVER project team has begun an assessment of campus data needs through interviews with researchers. The results of that assessment are being compiled and a white paper summarizing the interviews will be forthcoming.

Another institutional-level committee is the Data Executive Group, consisting of 10 members from a variety of campus organizations including the library, who leads the group, the CAC, and the Cornell Institute for Social and Economic Research (CISER). According to the survey response, the purpose of this group is to "to track and initiate activities in [e-science]." A precursor to this group was the library's Data Working Group, which wrote an excellent white paper on e-science issues within the library and on the campus as a whole: "Digital Research Data Curation: Overview of Issues, Current Activities, and Opportunities for the Cornell University Library."

While these committees promote and establish institution-wide activities, they are not policy-making bodies. Rather, they are providing the structure to support data-driven research by bringing together people with expertise in the technology, software, and best practices for the management of the data created through research. There are not campus-wide policies or procedures for dealing with data resulting from research.

Unit Specific

As for individual unit-specific efforts, Cornell University Library (CUL) has begun a list of campus units providing assistance for researchers with data. The services range from data management to intellectual property rights and publishing the data to be used by others. In each of the categories listed, CUL provides some level of involvement and service to the campus. This is an excellent marketing tool and an opportunity for the library to define both what they have to offer and areas where they do not provide service. There are a wide variety of individual units providing some kind of data services to their users and include, but are not limited to:

Center for Advanced Computing (CAC)
Cornell Institute for Social and Economic Research (CISER)
Cornell University Library's DataStaR project
Cornell University Survey Research Institute
Lab of Ornithology Information Science
Weill Cornell Medical College

CUL is connected more intensely with the CAC and CISER in activities related to e-science. The collaboration with CAC is primarily through DISCOVER, as described above. Collaboration with CISER involves joint review of e-science needs of social science researchers to assess which unit is better suited to support those needs. According to Cornell's survey response, "CISER is represented on the Data Executive Group (described above), provides services in library space, and is jointly exploring with the library service provision to support social science researchers."

Nature of Multi-institutional Collaborations

As with many of the respondents to the survey, Cornell University's involvement with other institutions revolves around grants, specifically an NSF DataNet proposal, in which Cornell is a subcontractor on the Johns Hopkins Data Conservancy DataNet proposal. The Data Conservancy (DC) will present an in-depth review of current data practices and recommend a model for data management that allows for the preservation, access, and easy reuse over time, both within and across disciplines. According to a recent presentation by Sayeed Choudhury at the Spring CNI meeting in April 2009, "The overarching goal of DC is to support new forms of inquiry and learning to meet [the grand research] challenges through the creation, implementation, and sustained management of an integrated and comprehensive data curation strategy." The lead PIs on this grant proposal include Sayeed Choudhury (Johns Hopkins, Library) and Carl Lagoze (Cornell, Information Science). The initial disciplinary focus of the Data Conservancy will include astronomers, computer and information scientists, health professionals, climatologists, earth scientists, ecologists, biologists, and social scientists.

The collaboration between Cornell and Johns Hopkins on the Data Conservancy proposal was facilitated by several existing connections. Dean Krafft, Chief Technology Strategist at CUL, named the Fedora repository project (now Duraspace) as a pivotal component that brought both Sandy Payette (Cornell, Computer Science) and Carl Lagoze to play roles in the development of the Conservancy. Carl Lagoze further recommended Steve Kelling from Cornell's Lab of Ornithology. Furthermore, Jim Cordes (Cornell, Astronomy) had worked with astronomers at Johns Hopkins. Dean Krafft worked closely with Sandy Payette and Carl Lagoze as a result of working on the National Science Digital Library (NSDL). Dean Krafft's initial work with the Data Conservancy involved evaluating the use of DataStaR as a potential "small science" front-end for deposit into the Data Conservancy. These connections demonstrate the importance of faculty being a major force for the direction of the library and the significance of working relationships across units and between institutions.

CUL is collaborating with faculty from the School of Information at Syracuse University to develop a curriculum for master's students with a background in the sciences to prepare them to work in libraries with data management as a major component of their responsibilities. This connection was the result of an IMLS award to Syracuse with CUL as a partner in completing the work of the grant. Beyond just curriculum planning, the grant will also provide student scholarships and incorporate a mentorship program at Cornell so that students can obtain direct experience working on e-science projects.

The IMLS grant to Cornell and Syracuse began as the result of close collaborations between the two libraries. For example, some Cornell librarians guest lecture or teach at the School of Information, and faculty from the Syracuse School of Information have been invited to speak at Cornell. Several librarians at Cornell have MS-LIS degrees from Syracuse and remain in touch with the iSchool. This simple networking encouraged faculty at the School of Information at Syracuse to consider collaborating with the librarians they knew from Cornell. Since the two universities are relatively close (an hour drive), face-to-face meetings were used to get the grant started, but video conferencing through Skype and conference calls have increased in order to make meetings more efficient. When asked what has worked well regarding multi-institutional grant opportunities, it was noted that if the responsibilities of each partner are concrete with easy to understand limits, working together is easier and more successful. Collaborating across distances is challenging because travel takes time, but technology can bridge that gap.

Another project created at Cornell that became a multi-institutional endeavor is VIVO, a tool designed to assist Cornell researchers in finding others on campus who have expertise in a certain discipline or who are working on a particular project or grant. VIVO searches across disciplines and retrieves public information about faculty and staff. In addition, VIVO presents the user with featured research, a master list of campus events, seminars and exhibits, and allows browsing by department, research centers, institutes, and programs as well as research facilities. Initially, this database was populated manually by students and librarians. Now there are automated methods for adding data; for example, Human Resources supplies a load set containing information on new faculty hires. As a result of their successful implementation of VIVO at Cornell, it was picked up by the University of Florida and implemented there. Further collaborations and an NIH grant were initiated by the University of Florida, Cornell University, University of Indiana, and four pilot universities to expand VIVO to include information about biomedical researchers across the country. The National Science Library, Chinese Academy of Sciences, Beijing has three instances of VIVO running and sent a visiting scholar with a particular interest in VIVO to work in CUL for six months. Also, the University of Melbourne is exploring the possibility of using VIVO as a data registry as part of the Australian National Data Service. Further collaborations and a grant were initiated by the University of Florida to expand VIVO to include information about biomedical researchers across the country, and the University of Melbourne is exploring the possibility of using VIVO as a data registry as part of the Australian National Data Service.

While not a collaboration between institutions, there is another grant and resulting product that deserves mention. Gail Steinhart is the primary investigator on another NSF grant through the Directorate for Computer & Information Science & Engineering (CISE), III-CXT: Promoting the curation of research data through library-laboratory collaboration. This grant lead to the creation of DataStaR, which is a service designed to facilitate publication of data to appropriate repositories, including Cornell's institutional repository, eCommons. The Web site for this service has interesting policy statements and guidelines for data authors. Although many are in draft format, they make an excellent beginning to define policies for data management around such areas as metadata standards and preparing tabular data for publication.

Reference and Consultation Services

CUL, through many of the efforts listed above, provides a high level of consultation and support services, including advising faculty on the use of available technology infrastructure and tools, discovering data sets, developing plans for data management, and the creation of tools to assist researchers. These services are provided by a combination of librarian liaisons to departments and also through librarians who have specific data responsibilities as part of their portfolio. Through its Web site, CUL assembles and organizes information about data management for the campus, including the policies of other organizations that collect and maintain data sets. CUL has been providing data services through geographic information systems (GIS) services, including a GIS data repository. Since these services have been available longer, they tend to be the most heavily used.

After speaking with Gail Steinhart, it was clear that many informal, one-on-one conversations between librarian and researcher brought about a variety of interesting opportunities, including participation as a member of grant proposals. These informal connections led to questions a librarian was able to answer either directly or through referral. As confidence in the informal connection grew, faculty began to think about the library in a different context and the invitations to participate as a member of a research team followed. Jeremy R. Garritano and Jake R. Carlson from Purdue University provide an excellent overview of this process in their article, "A Subject Librarian's Guide to Collaborating on e-Science Projects." Dean Krafft also suggested hiring people with previous research experience and connections into library positions and then utilize those connections to leverage opportunities for the library.

In addition to the efforts of the DISCOVER project team regarding campus data assessments, the library is considering following up with interviews of their own, asking slightly different questions than those covered in the DISCOVER interviews. Gathering information directly from faculty will provide at least two major opportunities for the library: the interviews will allow the library to identify appropriate and significant ways the library can be involved in research, and the interviews with researchers will increase one-on-one conversations, which can lead to further research and grant partnerships.

On occasion, CUL provides workshops on e-science topics for researchers. According to the survey response, "Past workshops include acquisition and use of remote sensing data and creation of Ecological Metadata Language records to describe ecological data sets. Geographic information systems workshops are standard fare. We expect to offer workshops on the use of DataStaR in the coming months."

Marketing these services is a bit challenging, both in terms of reaching the correct audience, but also in identifying willing partners that have projects within the scope of CUL's capabilities. In other words, if marketing efforts are too successful, CUL may be faced with more projects than they are able to handle well. The directory is featured on the Cornell University Library's home page (mouse over Library Services) and as a news item on the Mann Library Web site. It is also permanently listed on Mann Library's Research Tools and Services page.

Resourcing E-Science Activities in the Libraries

Cornell University Library has three permanent positions that make up their suite of librarians dedicated to e-science issues. Two of them, research data & environmental sciences librarian and research data & metadata librarian, have MS-LIS degrees. The former works as a librarian liaison to the Environmental Sciences Department and reports to the head, services for academic programs (Mann Library). The latter is currently not a liaison to a specific department, reports to the chief metadata librarian (Olin Library), and operates under the Metadata and Batch Processing Services unit in Central Library Operations. Both of the people in these positions have science backgrounds. The research data & environmental sciences librarian has a master's in ecology and evolutionary biology, and the research data & metadata librarian has a bachelor's in mathematics.

The third position, chief technology strategist, has an MS and PhD in computer science and a BA in mathematics and reports directly to the university librarian. A variety of programming staff are in temporary positions linked to grant money. The funding of certain grants, like DataStaR, has allowed the continued employment of these individuals. Maintaining appropriate staffing levels and expertise during a time of economic downturn will be a challenge over the next 3–5 years for many libraries.

In attempting to acquire more expertise in data management activities, CUL has evaluated every open librarian position and has restructured several of them to incorporate data management responsibilities as a significant portion of their duties. For example, when two vacancies occurred at approximately the same time, several existing positions were redefined, with one of the results being the creation of the research data and environmental sciences librarian position.

The development of formal and informal committee-type structures also contributed to a transformation of Cornell librarians. The Data Discussion Group, hosted by Keith Jenkins, GIS/Geospatial Applications Librarian, and Dianne Dietrich, Research Data &Metadata Librarian, provides a space for informal dialogue around a number of data issues. Recent journal articles in both librarianship and research areas are read and discussed, guests are invited to speak to a particular topic, or a new dataset is explored. This group meets once a month for 90 minutes each time. Initially, the conversations focused on reference services alone, but have gradually moved to take on data management types of services as well. The "Data Posse" is a very informal group that brings together everyone throughout the Cornell University Library where "data" informs some or all of their day-to-day duties, and it is this group that shares information and resources pertinent to accomplishing job duties. The Data Executive Group is the most formal committee, and, as described above, is comprised of representatives from the library, Center for Advanced Computing (CAC), and the Cornell Institute for Social and Economic Research (CISER). It is this group that discusses the overall direction of data services and makes sure that there is not duplicative effort.

As a result of this more informal structure, a distinct "data unit" has not been established in CUL. The current structure allows for connections around projects, such as DataStaR, as well as more formal opportunities through the Data Executive Group. The number of people involved in data activities is currently on the smaller side, but as it grows, the need for a distinct unit or more formal methods of communication may arise. The success of this informal structure, as with so many organizational structures, relies upon a group of people willing to work together and to be more collaborative than competitive.

Librarians at Cornell do not have tenured faculty status. When asked about the advantages or disadvantages, Gail Steinhart stated that, generally speaking, faculty do not care about the tenure status of librarians, but they do want to know that you can speak their language and that you understand how they do research. Gail's master's in ecology and evolutionary biology, which she received from Cornell University, and her years working as a scientist prior to attending library school, have been instrumental in her success in working with faculty at Cornell. Additionally, having liaison responsibilities to a particular department has allowed her access to faculty to make the informal connections that have created faculty-librarian partnerships on projects or grants.

The need for the master's in library science as a requirement for a position whose major responsibilities reside in data management is a question that deserves some attention. The IMLS grant to Cornell and Syracuse mentioned above speaks to the need for iSchools to develop curriculum to meet the demand for skills in data management. Individual courses on database design and management as a part of an overall master's degree in library and information science would appear to be an important aspect of becoming ready to tackle data-oriented responsibilities. Dianne Dietrich, a recent graduate from the iSchool at the University of Michigan, recommends an emphasis on "tying together courses that already exist…to show how students from different specializations [in an information school] bring unique, complementary perspectives to the material. [For example], data curation folks can learn a lot from archivists and information policy specialists; interface specialists can learn from librarians and vice versa; and so on." Considering the information gathered in this report, three areas of specialty seem important in producing a candidate likely to succeed in the realm of research data: an advanced science degree and/or experience as a research scientist, some useful IT skills such as programming, and some understanding of typical library concepts such as accessibility, preservation, searching, and the application of metadata.

A wide variety of opportunities have been provided to CUL librarians and staff for continuing education around e-science. The obvious in-house workshops and presentations have been offered along with the opportunity for librarians to attend e-science conferences and meetings, as well as intensive programs such as the Inter-University Consortium for Political and Social Research four-week summer program. Perhaps unique to Cornell is a reciprocal agreement with Syracuse University, where CUL receives a one-course tuition credit in exchange for hosting an SU iSchool intern. These credits are then available for use by CUL staff. While not an option that many librarians have taken advantage of, courses offered through the iSchool at Syracuse, such as IST 659 - Data Administration Concepts and Database Management, would be especially valuable. All Cornell staff are eligible to take one course at Cornell per semester, which presents CUL librarians with additional professional development opportunities. Dianne Dietrich, for example, took advantage of this opportunity to audit a course in scientific computing.

Motivation

There are many reasons, particularly cost, that would encourage any campus unit to determine that data curation activities were somebody else's problem. With that in mind, it is interesting to note the motivations for CUL's involvement in data management on the Cornell campus and the catalyst for those activities.

In their report, "Digital Research Data Curation: Overview of Issues, Current Activities, and Opportunities for the Cornell University Library," the CUL Data Working Group details the impetus for moving forward with the many activities detailed in this case study.

> "There are three primary (and related) motivations for developing a robust data curation infrastructure: enabling new discoveries by exposing data for use in data-driven research, ensuring access to and preservation of scholarly output, and meeting existing or forthcoming requirements of funding agencies or institutions regarding data management, retention, and access. Libraries have demonstrated expertise in several areas that could be productively applied to the practice of data curation, and in some cases, cyberinfrastructure development." (Page 4 of the report)

In addition to these well-articulated reasons for CUL's participation in data curation, there was strong visionary leadership, both on campus through former provost Biddy Martin and in the library through university librarian Anne Kenney. As always, the library must support the direction of the institution it supports. Martin provided two major goals for the university that explicitly spoke to the creation of infrastructure and expertise in the areas of data management, preservation, and access. The CUL Data Working Group summarizes these goals in their final report:

> "Two goals from [former] Provost Martin's address, which focused on the goals articulated in Cornell's consolidated planning document, are supported by undertaking data curation efforts within CUL. The first of these is Goal III: 'Enable and encourage the faculty, their students and staff to lead in the preservation, discovery, transmission, and application of knowledge, creativity, and critical thought.' Exercising responsible stewardship of the outputs of research, including digital research data, would support that priority. Goal IV – 'Extend our leadership in the use of research and education to serve the public good in fulfillment of Cornell's land-grant mission and its long-standing commitment to capacity building in

communities in the United States and around the world' – is also supported by developing data curation services and infrastructure, by enabling the sharing and reuse of research data by members of other institutions and the public at large."

Pressure Points

The survey asked an open-ended question regarding the pressure points experienced by the library. Cornell responded: "The rapidly changing landscape and uncertainty regarding the roles of different units and institutions is significant, as we attempt to identify our responsibilities and capabilities in this area. Finding and attracting creative programmers/developers, and librarians with an appropriate mix of subject and technology expertise can also be a challenge. Resource limitations and a lack of sustainable business models is a very significant challenge." In her interview, Gail Steinhart elaborated more on this topic, "Now that data has been identified as a problem, it's a very hard problem to determine how to deal with all of it at once. For that reason, DataStaR is meant to be a funnel to move data to more appropriate permanent repositories. Dianne Dietrich and I created the wiki (Research Data Management and Publishing Support) to help direct researchers to appropriate services. This is just a beginning."

Despite these identified pressure points, there is a positive attitude overall that focusing on data management is a significant and important role for academic libraries and that libraries can contribute considerably to campus goals in this area. Gail's words of advice capture this entrepreneurial spirit: Be willing to try out new ideas and new ways of doing things. Worry less about failure and more about the resulting problem if we don't try anything at all.

Highlighted Resources

DISCOVER: http://arecibo.tc.cornell.edu/DRSG/Default.aspx,
Center for Advanced Computing (CAC): http://www.cac.cornell.edu/
Cornell Institute for Social and Economic Research (CISER): http://ciser.cornell.edu/
Cornell University Library's DataStaR project: http://datastar.mannlib.cornell.edu/
Cornell University Survey Research Institute: http://sri.cornell.edu/
Lab of Ornithology Information Science: http://www.birds.cornell.edu/is/
Weill Cornell Medical College: http://www.med.cornell.edu/research/rea_sup/
VIVO: http://vivo.cornell.edu/
"Syracuse University currently recruiting for eScience Librarianship Fellows." http://ischool.syr.edu/newsroom/news.aspx?recid=802
DataStaR: http://datastar.mannlib.cornell.edu/
DataStaR, For Data Authors: http://datastar.mannlib.cornell.edu/index.jsp?primary=386684264
Cornell University Library: http://www.library.cornell.edu
"It's all about the data (management)." http://www.mannlib.cornell.edu/news/its-all-about-data-management
Albert R. Mann Library. Research Tools and Services: http://www.mannlib.cornell.edu/research-help/research-tools
Data Discussion Group: https://confluence.cornell.edu/display/culddg/Home
Research Data Management and Publishing Support at Cornell: https://confluence.cornell.edu/display/datasupp/

Case Study: Johns Hopkins University

Background

Johns Hopkins University (JHU), located in Baltimore, Maryland, is a private institution established in 1876 with a current student population of approximately 7,000.

The content of this case study was derived from the institution's response to the Fall 2009 ARL Survey on E-science and Data Support plus a subsequent telephone interview and e-mail correspondence with Sayeed Choudhury, Associate Dean for Library Digital Programs.

Structure of Response to E-Science

JHU's responses to questions on the survey indicated they organize their overall response to e-science through a hybrid structure that includes both institution-wide and unit-specific efforts.

Institution-wide

Johns Hopkins instituted an e-science task force a few years ago that has since disbanded. The task force was comprised of members from the faculty in a variety of departments, including astronomy, biostatistics, computer science, as well as representatives from the library and campus IT. Provost Lloyd Minor began his position in September of 2009 and has not yet reassembled this task force. However, response to the survey indicated, "The members of this task force continue to work with units throughout the university to advance our eScience initiatives."

The Institute for Data Intensive Engineering and Science (IDIES) has become one of the most visible university-wide organizations for e-science activities. IDIES is run by a steering committee of faculty from astronomy, physics, computer science, and the library, and has 42 affiliated faculty from the same departments plus 12 others, most notably several departments focusing on language and language processing. As stated on their home page, "The IDIES mission is to coalesce data-intensive science efforts at Johns Hopkins into a well-focused center of activity, and to propel various fields towards new discoveries and breakthroughs. By bringing together scholars from the Krieger School of Arts and Sciences, the Whiting School of Engineering, and the Sheridan Libraries to form interdisciplinary teams, IDIES aims to facilitate the development of tools and methods to derive knowledge from data in an exponentially expanding world."

The Digital Research and Curation Center (DRCC), a unit within the Sheridan Library, provides data curation and management support for the campus as well as very specific library projects. Sayeed Choudhury, Associate Dean for Library Digital Programs, is also the Hodson Director of the DRCC. The staff of ten began their work in 2004. The DRCC is open to supporting campus data curation needs and as a result, projects must be prioritized so they are manageable. One major DRCC project is supporting data management for the Sloan Digital Sky Survey (SDSS). The SDSS is governed by the Astrophysical Research Consortium, which provides legal oversight, making data curation for this multi-institution project possible.

In each of these major campus efforts, the IDIES and the DRCC, the library is a pivotal player in directing the programs and contributing to the success of data curation and management at JHU.

Unit Specific

JHU has several data centers within different departments with some e-science focused centers residing in the departments of physics and astronomy along with the Sheridan Libraries. Overall, JHU is much more decentralized in its approach to data curation and management issues. There are some signs that a central and systematic approach is desired, for example, IDIES provides services across different schools. Once the data is in a centralized location, campus administrators will want to know basic statistics such as the amount of data being created and cited, how data is used in teaching, whether data is being reused to further science at JHU or elsewhere.

Even though there are researchers on campus with concerns and interests in data, there are still many who are unwilling to consider the longer-term issues once the project ends or the research takes them in another direction. The JHU Library is casting themselves as partners to assist with data management by taking on several retrospective inventorying projects to get experience with the details of data curation, preservation, and access. The ideal is to be involved with the entire research process, encouraging faculty to think about and plan for keeping and accessing the data at the beginning of a project as opposed to the end.

Nature of Multi-Institutional Collaborations

JHU collaborates with so many institutions it would be difficult to describe each one in this document. Instead, several were selected as particularly notable. For a full list of projects conducted by the DRCC, some of which are collaborative efforts, see http://ldp.library.jhu.edu/dkc/projects.

Data Conservancy

The Data Conservancy (DC) will present an in-depth review of current data practices and recommend a model for data management that allows for the preservation, access, and easy reuse over time, both within and across disciplines. According to a recent presentation by Sayeed at the Spring CNI meeting in April 2009, "The overarching goal of DC is to support new forms of inquiry and learning to meet [the grand research] challenges through the creation, implementation, and sustained management of an integrated and comprehensive data curation strategy." The lead PIs on this grant proposal include Sayeed (JHU, Library) and Carl Lagoze (Cornell, Information Science). The initial disciplinary focus of the Data Conservancy will include astronomers, computer and information scientists, climatologists, earth scientists, ecologists, biologists, and social scientists.

The Cornell case study presented earlier provides some insight on the development of this collaborative Data Conservancy proposal effort between Cornell and JHU. A similar viewpoint was echoed by Sayeed when describing the collaboration from JHU's perspective. When determining the participation of various institutions, the initial factor for inclusion was an intellectual one and focused on the ability of the institution to contribute to the major objectives of the project and the problems being addressed. For example, Cornell was asked to participate based on the contributions of the data modeling expertise, including the Open Archives Initiative Object Reuse and Exchange (OAI-ORE). After that, inclusion was based on existing partnerships or the work of researchers that had become nationally recognized. For example, one researcher at Cornell participating in the project would recommend another researcher at a different institution who is well known for their contributions in a particular discipline. This would happen along the lines of, "If you're thinking of including earth sciences, you really ought to talk to person X at institution Y who is doing some really great work in that area." Again, the recommendation is still an intellectual one, but based on personal knowledge or connections.

Summer Data Curation Institute

JHU Libraries work with the Graduate School of Library and Information Science (GSLIS) at the University of Illinois at Urbana-Champaign to develop their summer institute program that furthers the education of "practicing academic librarians and other information professionals who want to learn more about data curation services in academic libraries." Participation in creating this program, and sending JHU librarians to attend, allows for the dissemination of that information throughout library staff. In addition, this connection led to the creation of an internship with one of the GSLIS students. The internship provided access to scientists and an opportunity to experiment with data curation in a real setting.

Digital Data Curation

As stated on the Library Digital Programs Web site, "The Library Digital Programs is collaborating with the Virtual Observatory to develop strategies for data curation of large-scale, digital astronomy datasets. These data curation activities will result in repository-based processes, tools, and systems that will provide long-term archiving of datasets to support research, learning, and dissemination."

Digital Audio Archive Project (DAAP)

A team of librarians at Indiana University and Johns Hopkins University through the DRCC will combine their efforts to create a system of digitizing and preserving audio tapes and serving the digitized products through a Web-based audio library. By focusing on best practices, open source software and standards, the team aims to provide a financial model for audio digitization that is economical. This effort was funded through an IMLS grant.

Services for a Customizable Authority Linking Environment (SCALE)

This combined effort between JHU and Tufts digital library researchers will focus on providing metadata linking for common terms and phrases for users of the National Science Digital Library to supplementary information found in dictionaries, encyclopedias, thesauri, and subject hierarchies. This service is designed to allow scientists reviewing information outside of their expertise, less experienced researchers, or the general public to retrieve information that will be useful in interpreting unfamiliar technical terminology.

A Technology Analysis of Repositories and Services

Three institutions, JHU, the University of Virginia, and the Massachusetts Institute of Technology, are working together to review and evaluate technology and architecture that support the delivery of electronic publishing, digital repositories, and e-learning, and provide a report of best practices and recommendations. This project was funded by The Andrew W. Mellon Foundation.

Reference and Consultation Services

Through the efforts of the Library Digital Programs, the Digital Research and Curation Center, and the Institute for Data Intensive Engineering and Science, the libraries at Johns Hopkins University provide an extensive array of services and products directly related to data management. According to the JHU survey responses, individual librarian liaisons, dedicated data librarians, and software developers assist researchers in finding and using available technology infrastructure and tools, finding relevant data, developing data management plans, and developing tools to assist researchers.

Sayeed emphasized the need to focus on faculty when considering moving the library in the direction of e-science. An interview survey of researcher needs is one of the best places to start, even if the survey is not comprehensive. Asking faculty about the data they care about, why one data set is more important than another, how they might imagine others or themselves needing access to the data in the future, is important in developing services and infrastructure that best meet their needs. If the faculty are engaged and have a sense of urgency, there is a better chance of success.

Resourcing E-Science Activities in the Libraries

JHU Libraries have hired staff specifically to provide e-science services, and the survey response indicated that there are plans to hire three additional staff through the NSF DataNet award. While the current associate dean for library digital programs also holds the position of Hodson Director, there may be plans to hire another person to take on the role of directing the DRCC and then to report to the associate dean for library digital programs. JHU has hired an executive director for the Data Conservancy. The other two positions include a software developer and project manager, both with bachelor of science degrees. The project manager will report to the DC executive director and the software developer will report to the Hodson Director. For those librarians and staff who have added data services to their portfolio, there is support available to attend conferences and meetings with e-science themes.

Sayeed has a master's of science in engineering and has completed coursework toward a PhD, also in engineering. His graduate degree was obtained at JHU, and, as described in the Cornell case study, his having been educated by the faculty at JHU has strengthened the library's connections to faculty and has proved very beneficial in moving a program of data curation forward on campus. Sayeed cited advantages to the MLIS degree in regards to understanding how libraries operate and their professional connections and networking with faculty. The advantage of an advanced degree in a discipline is the ability to view the world in a particular way with an understanding of how individual pieces need to be brought together. No matter the educational background, one important aspect of working with faculty is becoming a trusted member of their team, and this pressure will remain until enough evidence of successful faculty-library projects have been completed. As for other staffing in this area, the DC grant is a major opportunity to expand staffing in e-science for JHU Libraries. Subject librarians' involvement will increase as a result of the grant because their connections with faculty will be leveraged to discover more information about current data practices.

Sayeed is consistent in his message as he discussed the future of the MLIS degree. He recommends following how science research and teaching are changing and then adjust the education and training of future information professionals accordingly; for example, leverage collection development courses to focus on data curation as one aspect of building collections in libraries. Libraries will be defined by the services they build that are useful to researchers.

Pressure Points

JHU's response to the open-ended question regarding pressure points in the library related to e-science support or e-research more broadly ranged from staffing to financial support for basic infrastructure:

- The need to balance "traditional" services and needs with emerging eScience priorities.
- The difficulty in hiring and retaining the appropriate talent for such work.
- Embracing the somewhat unique culture that is required to support a unit with innovative goals and programs.
- Financial support from base library budget for core infrastructure needs (e.g., storage systems, servers, software management tools).

Motivation

Considering the pressure points revealed by many institutions in the ARL survey, it is interesting to note the reasons JHU Libraries are making significant investments in data curation. Sayeed emphasized that JHU Libraries were following the lead of their faculty, who are using new methodologies in their research and who are engaged and interested in managing data sets created by their research for future use. By creating a separate department in the library, the Digital Research and Curation Center, and by naming a position at the associate dean level to focus on Library Digital Programs, JHU Libraries provided a signal both internally and externally to the campus of the importance of sustaining and supporting new ways of doing research. Just as importantly, if libraries and academia are not invested in developing solutions to the data problem, someone else will. And it will likely be a company/vendor/publisher who is interested in making a large profit off the work of our faculty, much like we have experienced in the publishing world. It is possible that libraries will have projects that are grand success stories and others that fail miserably. Each of these possibilities will bring new knowledge to the table and allow us to move forward. Sayeed encourages his library colleagues to "act locally and participate globally." Follow the lead of your faculty on your local campuses and become engaged in multi-institutional projects to address the future of data management.

Highlighted Resources

Institute for Data Intensive Engineering and Science (IDIES): http://idies.jhu.edu
Digital Research and Curation Center (DRCC): http://ldp.library.jhu.edu/vhost-base/dkc
Data Conservancy: http://www.library.jhu.edu/about/news/releases/pressrel09/nsfgrant.html
"GSLIS to Hold Summer Institute for Humanities Data Curation." http://www.lis.illinois.edu/articles/2009/01/gslis-hold-summer-institute-humanities-data-curation
Library Digital Programs: http://ldp.library.jhu.edu/scp
Virtual Observatory: http://www.us-vo.org/
Services for a Customizable Authority Linking Environment (SCALE): http://dca.lib.tufts.edu/scale/

Case Study: University of Illinois at Chicago

Background

The University of Illinois at Chicago (UIC) was formed in 1982 by the consolidation of two University of Illinois campuses: the Medical Center campus, and the comprehensive Chicago Circle campus, and has a current student enrollment of about 26,000 students.

The content of this case study is derived from the institution's response to the Fall 2009 ARL Survey on E-science and Data Support, a subsequent telephone interview, and e-mail correspondence with Dr. Robert J. Sandusky, Assistant University Librarian for Information Technology at the University of Illinois at Chicago.

Response to Survey

The UIC responses to questions on the survey indicated that individual units (i.e., departments, colleges, schools, etc.) develop infrastructure and policies related to their own e-science needs. Within the library, e-science infrastructure or support services are in the planning stages.

Activities within the Library

The UIC Library has been involved in different areas of importance to scholarly communications, including support of online journal publishing (Open Journal System) and an institutional repository (DSpace), and has recently become more active in areas of e-science.

The library's leadership role in the area of online journal publishing began with the publication of First Monday, which has been hosted by the UIC Library since the journal left Denmark in 1998. Since that time, the library has implemented the Open Journal System (OJS) software from the Public Knowledge Project and currently hosts four journals on its site, Journals@UIC. The goal of Journals@UIC is to make journals openly available to the scholarly community worldwide. It also aims to assist UIC faculty and others with the management and editorial work associated with the journals they edit.

The UIC Library's ventures into institutional repositories mirrors the experiences of other libraries. UIC uses the DSpace software package to support its institutional repository, named Indigo. This service is an "online collection of the research and scholarship of faculty, students, and staff at the University of Illinois at Chicago." The bulk of the items publicly available through Indigo belong to the University Archives, with very little use of the IR by campus faculty. Working with faculty to upload their research materials into Indigo has not been an area of focus for the library, and technical staff support for the system is limited. Opportunities exist for collaboration with other schools in the University of Illinois system (Springfield and Urbana-Champaign) to share costs/operations of OJS and DSpace.

Realizing that the library needed to know more about campus researchers and their needs, an e-research team was charged in 2009 with developing plans and programs to support e-research. Specifically:

> The purpose of the E-Research Team at UIC is to develop a plan for the library on how it will meet the needs of those disciplines in the sciences and social sciences beginning to work with the cyberinfrastructure and heavily dependent on data. [Humanities may be addressed at a later date.] The Team should conduct an environmental scan to determine the issues in this emerging field and the needs and opportunities at UIC. Keeping in mind that no single institution can do it all and that collaboration with other institutions will be essential, the Team should make recommendations for what UIC can and should do to meet the identified needs. The Team should address what expertise may need to be developed, what resources are needed, what collaborations should be pursued, and how the library might be better organized to respond to faculty needs.

The first major activity for this group, which is comprised of nine staff plus the committee chair (the science librarian), was a survey of data resources and needs across campus. While a report from that group is forthcoming, a preliminary analysis of the data yielded clear direction for the library in the realm of e-science and e-research.

Assessment: Methodology and Findings

In the fall of 2009, a Web-based survey was sent to researchers across campus, where the population being queried was mostly faculty, staff, and graduate students. For the purposes of this survey, the library was most interested in medical sciences (the campus has a health sciences facility), hard sciences, and social sciences. As noted in the charge, humanities and the arts are not a current topic of focus. Preliminary recommendations for the library include the following:

The library should take a leadership role on campus for e-research/e-science and cyberinfrastructure, and specifically help facilitate cross-disciplinary and interdisciplinary research.

A near-term topic of investigation is whether there is a tool that the library can install that will help represent faculty and their research interests and foster collaboration. Some applications that might help address this need include BibApp, Collexis, and VIVO. Alternatively, the medical campus has a rapid-moving, funded research project, one aspect of which is to select a tool that will do something similar. One question is whether the library can collaborate with the medical campus project. Because most of these applications require data feeds from campus human resources systems and harvesting of citation information to populate the publication records, this partnership could benefit all involved and reduce redundant efforts.

In the area of education and engagement, the library should help researchers and students develop improved practices around data management.

This initiative's goal is to help them with personal data management in a way that leads to data that is well-managed, collected, and ready for an archive. Another aspect to this issue is helping researchers identify appropriate existing archives of data for use in their research and educating them about data reuse and best practices for data management.

There is also a need to highlight existing collaboration tools that support educational and research efforts. Among the frequently cited difficulties on campus are the challenges of finding people to collaborate with, scheduling time to meet with them, and having easy-to-use and effective tools for online collaboration. A range of collaborative tools are available, but there is apparently low awareness and use of these, including RefShare (part of RefWorks), Skype, and Google Docs.

From the engagement perspective, the library needs to conduct additional analysis of the survey results, then share that information with other departments and university administration so everyone is aware of what is going on around campus. This work is currently in process.

The library should lead the development of a campus-level e-research program.

Certainly this will take more time and effort to develop, and the general goal is that campus IT should play a role in this project. Tools should be brought in that are useful to a particular community, yet which also can be used more broadly (and thus collaboratively across campus).

Hire an e-research librarian.

This person would perform a coordinating role within the library for e-research efforts.

Resourcing E-Science Activities in the Library

To develop staff capacity for e-science and data support, some staff at UIC were specifically hired to provide e-science service, while others have been reassigned to e-science. In cases where staff have been reassigned to e-science, most are being brought up to speed through a mix of external and internal training and travel support.

As the Assistant University Librarian for Information Technology, Robert Sandusky has an MS in computer science and a PhD in library and information science. Prior to his arrival at UIC in 2007, Robert was Assistant Professor, School of Information Sciences, University of Tennessee, Knoxville. It was this association which led to his involvement in the DataONE project sponsored by NSF, and he has successfully written grants to support special collections and metadata development. In addition, he has collaborated with Health Sciences and the National Library of Medicine to study outreach and evidence-based medicine.

The scholarly communications librarian also has some responsibility for e-science, with particular focus on journal publishing, the institutional repository, and engagement with the broader university community. This is a permanent position that reports to the university librarian.

Three librarians who provide support for e-science services have been hired within the last two years: the maps and data services librarian, the science librarian, and the metadata librarian. Within the UIC Library, the maps and data services librarian has specific responsibilities for GIS and social sciences data. This librarian has created two Web pages that provide service and other information related to e-science and data services. The science librarian plays an important role as the liaison librarian for the hard sciences and serves as chair of the e-research team. A new metadata librarian, reporting to the head of the resource acquisition and management department, came to the library with data preservation and curation experience. Both the maps and data services librarian and metadata librarian are involved with the e-research team and both have received additional funding and support from the library for training and events surrounding data curation issues.

To compensate for constraints on hiring new staff, UIC is providing opportunities for current staff to develop skills related to e-science by providing in-house workshops and presentations, support for staff to attend e-science conferences and meetings, and support for staff to take coursework related to e-science or data management in a discipline. At UIC, working with faculty is very much relationship-based and relies more on ties that individual subject librarians have created with their departments than any general marketing efforts. Although librarians at UIC have faculty status, Sandusky believed it is difficult to determine whether that has made it easier for librarians to work collaboratively with researchers across campus.

Activities of staff on e-science/e-research issues are currently project-driven and focused around the workings of the E-Research Team. Perhaps this will change when the library hires an e-research librarian and has developed additional knowledge and expertise. A current focus for all staff is looking for initial and early contacts with potential project partners, although, based on the survey results, e-science activity levels on campus are low.

Research Activity and Opportunities for Collaboration

While Sandusky wasn't aware of any NSF DataNet proposals with UIC as the lead institution, he is a co-investigator on DataONE, one of the first two DataNet proposals selected for funding by NSF:

> DataONE will ensure the preservation and access to multi-scale, multi-discipline, and multi-national science data. DataONE will transcend domain boundaries and make biological data available from the genome to the ecosystem; make environmental data available from atmospheric, ecological, hydrological, and oceanographic sources; provide secure and long-term preservation and access; and engage scientists, land-managers, policy makers, students, educators, and the public through logical access and intuitive visualizations. Most importantly, DataONE is not an end but a means to serve a broader range of science domains both directly and through the interoperability with the DataONE distributed network.

Part of the technology proposal is to develop a distributed set of nodes that hold data, and the role envisioned for academic libraries is to participate as storage nodes in the network. If the library has data collections, that data could be stored in that node and benefit from DataONE's preservation, discovery, and access services. Alternatively, the library could participate by providing a node to form a part of the storage and preservation fabric.

From a services perspective, one of the goals is to develop materials that will promote and develop informatics literacy on campus. These could take the form of classroom-based instruction or journal and report publication through repositories/open journal systems. An especially interesting idea is the development and support of a virtual reference service around DataONE that includes domain experts and librarians. UIC would be the first academic library to participate that way, and they would help develop library-oriented/friendly interfaces to these datasets that are held in the DataONE network. Work in this area is envisioned to include integrating the disparate data into faceted browsing systems.

Other opportunities for collaboration exist in the context of consortia such as the Committee on Institutional Cooperation or the Consortium of Academic and Research Libraries in Illinois, and with other campuses in the University of Illinois system.

Because the demands of operation can exceed local capacity, the Chicago and Urbana-Champaign campuses are beginning to explore using shared instances of OJS and DSpace.

Campus Activities

In the survey results, UIC responded that there was a distributed data center for research data on campus, the National Center for Data Mining (NCDM). NCDM was founded in 1998 as a national resource for high-performance and distributed data mining. The center performs research, coordinates standards, operates testbeds, and engages in outreach. It "has received support from numerous funding agencies, including the National Science Foundation, the Department of Energy, the Department of Defense, and the National Aeronautics and Space Administration, as well as from other universities and other private agencies and industrial partners." At this time there is no direct tie between this group and the library.

Pressure Points

UIC's response to the open-ended question at the end of the survey elicited the following response:

> Scarcity of resources in an era of flat/declining budgets limits the library's ability to engage in all the opportunities that exist.
> Library still lacks sufficient expertise, but expertise has improved significantly in the past two years, and resources are being allocated to staff development to increase expertise.
> The campus does not have an institution-wide structure advancing e-science/e-research, which would help the library engage with other units across campus.
> Library is working on several collaborative projects within the Committee for Institutional Cooperation, the three-campus university system, and nationally/globally through DataONE to overcome some of the resource and expertise challenges.

As noted above, it is difficult for any library to engage with its faculty and researchers in the areas of e-research and e-science without expertise and/or an institutional support for such issues. Despite this, the UIC Library is developing strategies to raise the visibility of e-science/e-research on campus by engaging when opportunities present themselves, an approach Robert Sandusky advises to those wanting to get more involved.

Highlighted Resources

Public Knowledge Project: http://pkp.sfu.ca/?q=ojs
Journals@UIC: http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php
Indigo: http://indigo.lib.uic.edu:8080/dspace/
BibApp: http://bibapp.org/
Collexis: http://www.collexis.com/
VIVO: http://vivo.cornell.edu/
Finding Data and Data Services: http://researchguides.uic.edu/data
Maps and Geographic Information Systems: http://library.uic.edu/home/collections/maps-and-geographic-information-systems
DataONE: https://dataone.org/
Committee on Institutional Cooperation: http://www.cic.net/
Consortium of Academic and Research Libraries in Illinois: http://www.carli.illinois.edu/
National Center for Data Mining (NCDM): http://www.ncdm.uic.edu/

Case Study: Massachusetts Institute of Technology

Background

The Massachusetts Institute of Technology (MIT), located in Cambridge, Massachusetts, is a private land-grant university founded in 1861 with a current student population of about 11,000. In FY 2009, MIT had a total research expenditure of $718.2 million, including $61.4 million sponsored by NSF.

The content of this case study was derived from the institution's response to the Fall 2009 ARL Survey on E-science and Data Support plus subsequent telephone interviews and e-mail correspondence with MacKenzie Smith, Associate Director for Technology, Technology Research and Development, and Katherine McNeill, Data Services and Economics Librarian.

Structure of Response to E-Science

MIT's responses to the survey questions indicated that individual units (i.e., departments, colleges, schools, etc.) at the institution develop infrastructure and policies related to their own e-science need.

Background and Unit Response

MIT has a highly decentralized culture, and as such e-science planning and development has been ad hoc. At MIT, the cultural norm is that research is paid for and supported by the funders out of direct grant charges, not the institution. Indirect costs can cover campus infrastructure, including part of the libraries, but each PI is responsible for determining and supporting their own IT infrastructure. Most science and engineering departments, labs, and centers (DLCs) at MIT have some infrastructure to support High Performance Computing, or provide software tools to process/visualize research data. None of these efforts are clearly documented on a single Web page or other place where researchers can easily locate it, nor are there policies specific to MIT as opposed to the scientific field (e.g., the tiered research storage standards in the high energy physics community). The MIT-affiliated Lincoln Lab does offer centrally supported infrastructure via its next-generation Lincoln Laboratory Grid (LLGrid), an interactive, on-demand parallel computing system.

Even though there is a prevailing culture of decentralized efforts, the Institute has charged several committees and task forces over the past five years to investigate options surrounding the creation of a centralized, but decentrally funded, infrastructure for research data storage and processing. One outcome of the groups' work is the Holyoke High Performance Computing Center (HPCC), which has a scope that reaches well beyond the MIT campus boundaries. The HPCC is a partnership between MIT, Boston University, the University of Massachusetts, the Commonwealth of Massachusetts, and industry partners Cisco and EMC. The Center aims to be a "green" data and computing center, using nearby waterways for hydroelectric power and cooling, and existing fiber optic networks to connect to major research universities around the state. In the Governor's press release about the center, it was noted that this is a "collaboration that will lead to the development of a world-class, high-performance computing center in Holyoke, and a statewide collaborative research program."

Research Activities and Collaborations

MIT has a successful track record of developing software for the library community. In 2001–2002, the libraries partnered with Hewlett-Packard (HP) Labs to create DSpace, an institutional repository system that captures, stores, indexes, preserves, and redistributes an organization's research material in digital formats. It was released as open source software in 2002, and as of mid-2010, almost 900 institutions have self-registered their site in the DSpace Registry, giving DSpace the largest market share of any digital repository system in use by the library community. In 2009, the DSpace Foundation and Fedora Commons joined to become DuraSpace, an organization that "is committed to providing technologies and services that help ensure that our digital heritage is accessible over the long term."

Like other institutions highlighted in these case studies, MIT has partnered with other institutions in the development of an NSF DataNet proposal in rounds one and two. The second round MIT-led proposal is entitled DataSpace. Both proposals were submitted by a lead PI from the Sloan School of Management and the School of Engineering, with the Libraries as co-PI and main contributor. Other co-PIs from MIT are from the central IT department, the computer science department, civil and environmental engineering, and brain and cognitive sciences. For this particular proposal, MIT worked with a variety of institutions that have a strong focus on science and technology. Geographic and institutional diversity were important, as were existing connections with the other project partners. The proposal's summary states:

> Web technology brought tremendous efficiency gains for commerce, yet the world of scientific research has failed to fully leverage all its capabilities. As a result, scientists duplicate research and miss opportunities for discovery, collaboration, and translation of research into public goods. The DataSpace Project will bring these gains to science by providing a dramatically new approach to data management and long-term curation that accommodates multiple, heterogeneous data from a variety of distributed locations, and supports research across diverse disciplines and modalities, enabling investigators to easily access and aggregate data of known quality and provenance. It will build on proven technology and business models while bringing to bear the best research capabilities of MIT and nine partner organizations. To encourage sustainability and collaboration, organizations producing research data will be able to take responsibility for long-term stewardship of their data as part of a global network with only modest investment and expertise.

Three other universities are partnering with MIT to be "nodes" in a distributed model of data services, where each institution manages its own research data. Data will be easily deposited and integrated by faculty, and thematic Web standards will help with interoperability. As with any technology, when the infrastructure can be standardized, it is easier to layer services on top, and it is at the service-provision layer where research libraries can play a lead role. Libraries have an intimate knowledge of what is going on at the faculty level and have a proven track record of providing customized services and assistance to researchers that other parts of the organization do not. Regardless of the outcome of the NSF proposal, there is a demonstrable campus need for data management and curation services that the Libraries are now trying to provide.

Working with Researchers

A major challenge in providing e-science services and support to researchers is the very first step of finding the right approach. What works for one institution/department/researcher may or may not work with another, so subject librarians at MIT are using their existing relationships to talk informally with researchers about their data needs. Questions for discussion include: how much data is the researcher producing, who is in charge of it, is it being backed up? Some researchers are skeptical that the Libraries can help with their expensive/large/hard data needs and question why the Libraries would want to engage in that arena in the first place. Many researchers are eventually convinced that the library is an organization that can be trusted and is the right provider for their data needs. Librarians at MIT are actively engaging faculty who currently manage their own data, particularly large datasets. Other opportunities for faculty engagement have come in the face of a data management challenge such as the departure of a graduate student or an unexpected computer crash.

Librarians at MIT do not hold faculty status — theirs is an academic appointment with no tenure. In MacKenzie Smith's opinion, the lack of faculty status helps librarians when they meet with faculty to discuss their data management needs. Researchers are more likely to talk with librarians who come to the table with expertise they do not have and who understand and are able to talk to them in their subject-specific language. There is currently some debate in e-science circles about whether we need domain experts who can be taught information management skills or librarians who can learn the subject area of their researchers. At MIT, the latter approach currently works well, although that may shift over time. It is important that libraries experiment with different service models and share experiences so the library community can learn collectively.

Through many different modes, the MIT Libraries offers specific services that include helping researchers find relevant data, developing data management plans, archiving relevant data, and curating it for long-term preservation and integration across datasets. These services have been provided for some time, and the need for them is stronger than ever. Examples include successful statistical and geospatial data management services, and a bioinformatics training program. Through conversations and other assessment efforts, the Libraries have learned that many graduate students are tasked with these duties and are grateful for the help. Classes on data management are constantly re-tooled and attendance patterns cannot be tracked to a specific department or group, although graduate students are the most frequent attendees. Most of the need is for practical, just-in-time tools: "What software do I need to get X done?" with little interest in the theoretical framework. In January of each year, MIT has "Independent Activities Period," a month-long break from regular classes designed to give students time to learn new things. In 2010, the Libraries offered dozens of classes during this time, some of which focused on tools and software for data analysis, finding data sets, and using data. In addition to offering classes during IAP, other sessions are taught two or three other times during the year. Much of the information on Data Management and Publishing is online and accessible to users at the time of need.

In their survey response, it was noted that "MIT is committed to developing a major e-science curation program with the next few years. We are already supporting e-science data in limited ways (e.g., in our IR, with our Metadata Services group, by education and consulting) but we plan to do much more. As we explore the concept with the MIT administration, faculty, and students, the need for more services in this area becomes clearer, and our possible role in providing those services becomes more accepted."

Resourcing E-Science within the Library

The associate director for technology works on the long-term strategy for science data curation, including the assessment of current needs and appropriate role for the library, as well as the technological infrastructure required. There is a public services committee who is developing expertise in the topic, talking to faculty, developing pilot archiving projects, and teaching one-hour courses on the subject to students. There are also a DSpace product manager and a metadata expert assisting in these efforts. Overall, many people are involved with e-science plans and programs in some way. A recent reorganization created additional capacity by consolidating in certain areas and moving others to new areas of work.

Many of the MIT staff working with e-science were not hired into that role, nor did they necessarily come in with the skills targeted to working in e-science. In order to better meet the Institute's needs, select liaisons have added e-science and data management expertise to their current portfolio. Katherine McNeill noted that many librarians have added that expertise through reading, being active in relevant associations and professional opportunities, and, in many cases, just jumping in and doing the work. A broad-based data services interest group was formed a few years ago with staff who had been thinking about data initiatives from various format- or discipline-based perspectives. The group provides a venue for exchanging information, learning from one another, and has been a starting point for several collaborations. In order to keep their colleagues up-to-date with data services, some group members send out periodic e-mail reminders about their work to all public service librarians. Regular formal training is not part of their purview, but some members have done training in the past, including meeting with new public service librarians to let them know about the available services and that group members are willing to collaborate as opportunities arise. On-the-job e-science training will likely reduce in intensity as more librarians have an e-science background (e.g., data curation, GIS, data services), thus making it possible to recruit specifically for those positions.

An interesting difference between MIT and other research institutions is that the campus as a whole is very entrepreneurial, which means the Libraries looks for liaisons who are both interested in the department and willing to tailor services in whatever way will meet departmental needs. Experience at MIT has shown that it is crucial to get the right librarians involved in e-science issues — they get excited about it, and that spreads to other staff. Having data curation and similar issues on the horizon has helped staff see the future and that there is something different — albeit still tied to preserving the scholarly record — to look forward to.

Pressure Points

MIT's response to an open-ended survey question regarding pressure points is that "space is at a premium at MIT, so the campus is considering options for building a new high performance computing data facility off-campus. What to do about data curation in that scenario is unclear. The Libraries are involved in that discussion. More specifically on data curation, two MIT professors recently co-chaired an NAS committee to examine data curation called "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age" and the libraries are discussing with them how to implement the recommendations at MIT."

Highlighted Resources

Lincoln Laboratory Grid (LLGrid): http://www.ll.mit.edu/news/llgrid.html

Holyoke High Performance Computing Center (HPCC): http://web.mit.edu/newsoffice/2009/hpcc-0611.html

"Patrick Administration Announces Collaborative Plan To Build High-Performance Computing Center & Research Program In Holyoke." http://www.mass.gov/?pageID=gov3pressrelease&L=1&L0=Home&sid=Agov3&b=pressrelease&f=090611_holyoke&csid=Agov3

DSpace Registry: http://www.dspace.org/whos-using-dspace/Repository-List.html

DuraSpace: http://duraspace.org/about.php

DataNet Full Proposal: DataSpace Project Summary: http://web.mit.edu/smadnick/www/DataSpace/2009FullProposal-abbrev.pdf

Data Management and Publishing: http://libraries.mit.edu/guides/subjects/data-management/index.html

"Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age." http://www.nap.edu/catalog.php?record_id=12615

Selected Bibliography

Association of Research Libraries. "Agenda for Developing E-Science in Research Libraries: ARL Joint Task Force on Library Support for E-Science Final Report & Recommendations." (2007). http://www.arl.org/bm~doc/ARL_EScience_final.pdf

Association of Research Libraries. E-Science Survey Resource Page. http://www.arl.org/rtl/eresearch/escien/esciensurvey/surveyresearch.shtml

Association of Research Libraries. "To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering." (2006). http://www.arl.org/bm~doc/digdatarpt.pdf

Association of Research Libraries Joint Task Force on Library Support for E-Science. "Agenda for Developing E-Science in Research Libraries: Final Report and Recommendations to the Scholarly Communication Steering Committee, the Public Policies Affecting Research Libraries Steering Committee, and the Research, Teaching, and Learning Steering Committee." (2007). http://www.arl.org/bm~doc/ARL_EScience_final.pdf

Atkins, Daniel E., Kelvin K. Droegemeier, Stuart I. Feldman, Hector Garcia-Molina, Michael L. Klein, David G. Messerschmitt, Paul Messina, Jeremiah P. Ostriker, and Margaret H. Wright. "Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure." (2003). http://www.nsf.gov/publications/pub_summ.jsp?ods_key=cise051203

Choudhury, Sayeed. "Data Curation. An Ecological Perspective." *College and Research Libraries News* (2010): 194–96.

Gastinger, Almuth. "9th International Bielefeld Conference 2009: Upgrading the eLibrary: Enhanced Information Services Driven by Technology and Economics." *Library Hi Tech News* 26, no. 1/2 (2009): 1–5.

Garritano, Jeremy R., and Jake R. Carlson. "A Subject Librarian's Guide to Collaborating on e-Science Projects." *Science and Technology Librarianship* 57 (2009). http://www.istl.org/09-spring/refereed2.html

Gold, Anna. "Data Curation and Libraries: Short-Term Developments, Long-Term Prospects." (april 4, 2010) http://digitalcommons.calpoly.edu/lib_dean/27/

Hey, Tony, Stewart Tansley., and Kristen Tolle, eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft, 2009. http://research.microsoft.com/en-us/collaboration/fourthparadigm/contents.aspx

Hey, Tony, and Jessie Hey. "E-Science and Its Implications for the Library Community." *Library Hi Tech News* 24, no. 4 (2006): 515–28.

Lynch, Clifford. A. "Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age." *ARL: A Bimonthly Report* 226 (2003): 1–7. http://www.arl.org/resources/pubs/br/br226/br226ir.shtml

Pirani, Judith A., and Donald Z. Spicer. *The Chronopolis Project: A Grid-Based Archival Digital Preservation Solution*. ECAR Case Study 1,2010. http://www.cs.rpi.edu/~bermaf/ECS1001.pdf

Rambo, Neil. "E-Science and Biomedical Libraries." *Journal of the Library Medical Association* 97, no. 3 (2009) doi: 10.3163/1536-5050.97.3.001

Steinhart, Gail, John Saylor, Paul Albert, Kristine Alpi, Pam Baxter, Eli Brown, Kathy Chiang, Jon Corson-Rikert, Peter Hirtle, Keith Jenkins, Brian Lowe, Janet McCue, David Ruddy, Rick Silterra, Leah Solla, Zoe Stewart-Marshall, and Elaine L. Westbrooks. "Digital Research Data Curation: Overview of Issues, Current Activities, and Opportunities for the Cornell University Library." (2008). http://ecommons.library.cornell.edu/handle/1813/10903

University of California Curation Center. http://www.cdlib.org/services/uc3/

University of California, San Diego Research Cyberinfrastructure Design Team. "Blueprint for the Digital University." (2009). http://research.ucsd.edu/documents/rcidt/RCIDTReportFinal2009.pdf

Witt, Michael, and Jake R. Carlson. "Conducting a Data Interview." (December 2007) http://docs.lib.purdue.edu/lib_research/81/

Appendix I: The Survey Instrument

ARL Survey on E-science and Data Support

1. Introduction

When the ARL E-science Task Force began its work in 2006, general information was collected on members' activities through an exploratory survey posing a few basic questions. As understanding of the issues involved in developing e-science services has advanced, the time is ripe to develop a more current and focused picture of ARL member engagement. The E-science Working Group has developed a new, more detailed survey that we believe will assist the membership in understanding the community's involvement with e-science support and help the Working Group act to support members in this crucial work. We also hope to surface models for e-science programs that would be of interest to our broader community.

Please complete this survey on e-science (or refer to relevant staff for completion) by September 14, 2009. For purposes of this survey, e-science is defined broadly not only as big computational science, but also team science and networked science. It includes all scientific domains, as well as biomedicine and social sciences that share research approaches with the sciences.

NOTE: If you are not able to complete the survey in one sitting, you may return to the survey and resume where you left off. You will need to use the same computer and browser each time you access the survey and have cookies enabled.

An * indicates a required response.

The E-science Working Group plans to share survey findings with the membership and to make a general report public. If there is information included in your survey responses that you wish to have treated confidentially, please indicate that when you submit the document, URL, or description.

Questions can be directed to Karla Hahn.
* Select your institution:

* Please provide the following contact information for the person completing the survey.

    Name:
    Job Title:
    E-mail:

2. Background

Is your institution providing infrastructure or support services for e-science as defined above?

Yes (Please click the Next button below to continue the survey.)
E-science infrastructure or support services are in the planning stages (Please click the Next button below to continue the survey. Answer as many of the questions as possible based on current planning.)
No (Please click the Next button below to skip to section 23.)

3. E-science: Institutional Planning and Policy Development

Please indicate which of the following structures best describes how your institution has organized itself to advance e-science planning and policy development.

My institution has or is planning an institution-wide structure (such as a group or task force) to advance e-science planning and policy development (When you click the Next button below you will continue to section 4.)

At my institution individual units (i.e., departments, colleges, schools, etc.) develop infrastructure and policies related to their own e-science needs (When you click the Next button below you will skip to section 5.)

My institution has or is planning a hybrid structure that includes both institution-wide and unit-specific efforts. (When you click the Next button below you will skip to section 6.)

There is another organizational structure to advance e-science planning and policy development (When you click the Next button below you will skip to section 7.)

4. Institution-wide Structure

You indicated that your institution has a institution-wide structure to advance e-science planning and policy development. Who are the members of the group/task force? Check all that apply..

IT staff
Faculty/researchers
Office of Research staff
Library staff
Other, please describe

Please briefly describe the group's e-science planning and policy development responsibilities.

If there is a URL to a Web page that describes the group and/or its charge, please enter it here. If there is a document (not online) that describes the group and/or its charge, please e-mail it to tricia@arl.org.

When you click the Next button below you will skip to section 8.

5. Individual Units

You indicated that individual units (i.e., departments, colleges, schools, etc.) at your institution develop infrastructure and policies related to their own e-science need. If you are aware of any Web pages that describe the specific e-science efforts, please enter the URLs here. If there are documents (not online) that describe these efforts, please e-mail them to tricia@arl.org.

When you click the Next button below you will skip to section 8.

6. Hybrid Structure

You indicated that your institution has a hybrid structure that includes an institution-wide group(s) to advance e-science planning and policy development. Who are the members of the group/task force? Check all that apply.

IT staff
Faculty/researchers
Office of Research staff
Library staff
Other, please describe

Please briefly describe the group's e-science planning and policy development responsibilities.

If there is a URL to a Web page that describes the group and/or its charge, please enter it here. If there is a document (not online) that describes the group and/or its charge, please e-mail it to tricia@arl.org.

You indicated also that individual units (i.e., departments, colleges, schools, etc.) at your institution develop infrastructure and policies related to their own e-science needs. If you are aware of any Web pages that describe the specific e-science efforts, please enter the URLs here. If there are documents (not online) that describe these efforts, please e-mail them to tricia@arl.org.

When you click the Next button below you will skip to section 8.

7. Other Organizational Structure

You indicated that your institution has another organizational structure to advance e-science planning and policy development. Please briefly describe it here.

If there is a URL to a Web page that describes the structure and/or its charge, please enter it here. If there is a document (not online) that describes the structure and/or its charge, please e-mail it to tricia@arl.org.

When you click the Next button below you will continue to section 8.

8. Data Support and Services

Does your institution have a designated unit or units to provide data curation support (e.g., consultation services, policy interpretation, storage infrastructure, etc.) for scientific research data?

Yes
No

If yes, please identify the unit or units and briefly describe their role.

Has your institution conducted an assessment of data resources and needs?

Yes
No

If there is a URL for a publicly available report from the assessment, please enter it here. If there is a publicly available report (not online), please e-mail it to tricia@arl.org.

9. Data Support and Services, cont.

Does your institution have a central data center or distributed data centers for research data?

Central
Distributed
Both central and distributed

Comments

If there is a URL to a Web page that describes the data center(s), please enter it here. If there is a document (not online) that describes the data center(s), please e-mail it to tricia@arl.org.

Does your institution (or individual units in your institution) provide support for digital lab notebook applications?

Yes
No
Don’t know

If yes, what unit(s) provides support?

If yes, please briefly describe the support your institution provides for digital lab notebook applications.

If yes, does your institution have an archiving plan for digital lab notebook data?

Yes
No
Don’t know

If yes, please briefly describe the archiving plan.

10. Grant Proposals

Has your institution submitted an NSF DataNet proposal?

Yes
No
Don't know

If yes, please briefly describe all proposals submitted. Include which unit submitted the proposal and which discipline or research field the proposal addressed.

Is your library a participant in the proposal?

Yes
No

Additional comments about NSF DataNet proposals.

Is the library the lead or a participant in any other grant applications related to e-science?

Yes
No

If yes, please briefly describe the grant proposal(s) submitted. Include which unit(s) submitted the proposal(s) and which discipline or research field(s) the proposal(s) addressed.

11. E-science: Library Support

Is your library involved in e-science support at your institution?

Yes
No (Please click the Next button below to skip to section 22.)

If yes, who in the library has primary leadership responsibility for plans and programs for e-science support? (Questions about participants in service delivery come later in the survey.)

A single individual is responsible for developing plans and programs (When you click the Next button below you will continue to section 12.)
A group/committee(s)/team(s) is responsible for developing plans and programs (When you click the Next button below you will skip to section 13.)
A department/unit is charged with developing plans and programs (When you click the Next button below you will skip to section 14.)
A combination of the above or Other (When you click the Next button below you will skip to section 15.)

12. Individual

Please provide the following information about the individual in the library who has leadership responsibility for developing e-science plans and programs.

Position title:
Year position took on e-science support responsibility:
To whom the position reports:
Approximate percentage of time spent on e-science activities:

If there is a URL for the position description, please enter it here. If there is a position description (not online), please e-mail it to tricia@arl.org.
Additional comments:
When you click the Next button below you will skip to section 16.

13. Library Group, Committee, or Team

Please provide the following information about the library group/committee/team charged with leadership responsibility for developing e-science plans and programs.

Name of group/committee/team:
Position title of group/committee/team leader:
Year group/committee/team took on e-science support responsibility:
Number of group/committee/team members:

If there is a URL to a Web page that describes the standing committee or its charge, please enter it here. If there is a document (not online) that describes the standing committee or its charge, please e-mail it to tricia@arl.org.

When you click the Next button below you will skip to section 16.

14. Library Department/Unit

Please provide the following information about the library department/unit charged with leadership responsibility for developing e-science plans and programs.

Name of department/unit:
Position title of department head:
Year department/unit took on e-science support responsibility:
Number of staff in the department/unit who provide e-science support:

If there is a URL to a Web page that describes the department/unit or its charge, please enter it here. If there is a document (not online) that describes the department/unit or its charge, please e-mail it to tricia@arl.org.
Additional comments:
When you click the Next button below you will skip to section 16.

15. Combination or Other E-science Support Structure

Please briefly describe the library organization for developing e-science plans and programs.

When you click the Next button below you will continue to section 16.

16. Library Collaboration

Does your library collaborate with another unit(s) in your institution to provide e-science support?

Yes
No

If yes, please provide the following information.

Library unit(s) involved:
Institution unit(s) involved:
Please briefly describe of the project(s) and the discipline(s)/research field(s) involved:

If there are URLs to Web pages that describe collaborative efforts, please enter them here. If there are documents (not online) that describe these efforts, please e-mail them to tricia@arl.org.

17. Components of Library Activity

Which of the following reference/consultation services that assist scholars and researchers with the identification, access, and use of data does your library offer or plan to offer? Check all that apply.

Finding and using available technology infrastructure and tools
Finding relevant data
Developing data management plans
Developing tools to assist researchers
Other, please describe

Who provides reference/consultation services to researchers? Check all that apply.

Individual discipline librarians/staff
Dedicated data librarian(s)/specialists
Other, please describe

Additional comments:

18. Components of Library Activity, cont.

Does your library have a Web site to provide service and other information related to e-science and data services.

Yes
No

If yes, please provide the URL(s).

Does your library offer researchers workshops, classes, etc., related to e-science and data issues?

Yes
No

If yes, please briefly describe the content of the workshops, classes, etc..

19. Components of Library Activity, cont.

Does your library include policy issues associated with e-science (e.g., open data, compliance with federal agency policies) in its outreach program?

Yes
No

If yes, please briefly describe the issues covered or provide the URL(s) to a description of policy issues associated with e-science, or e-mail tricia@arl.org.

Does your library manage, or participate in managing, technology infrastructure (e.g., data storage, tools for data analysis, virtual community support) that supports e-science?

Yes
No

If yes, please briefly describe the infrastructure managed and the library's role or provide the URL(s) to a description of e-science infrastructure support, or e-mail tricia@arl.org.

20. Library Professional and Workforce Development

How has your library developed staff capacity for e-science and data support? Check all that apply.

Hired staff specifically to provide e-science services
Reassigned existing staff
Planning to hire staff
Other (please describe)

Please provide the following information for up to three positions: type of position (e.g., permanent, contract hires, graduate student, etc.), degree(s) the individual holds (e.g., MLS, discipline Masters, discipline PhD, etc.), to whom the position reports. If there is a URL for the job posting or position description, please enter it also. If there is a job posting or position description (not online), please e-mail it to tricia@arl.org.

Position 1
Type of position (e.g., permanent, contract hires, graduate student, etc.):
Title:
Degree(s) the individual holds (e.g., MLS, discipline Masters, discipline PhD, etc.):
To whom the position reports:
URL for the job posting or position description:

Position 2
Type of position (e.g., permanent, contract hires, graduate student, etc.):
Title:
Degree(s) the individual holds (e.g., MLS, discipline Masters, discipline PhD, etc.):
To whom the position reports:
URL for the job posting or position description:

Position 3
Type of position (e.g., permanent, contract hires, graduate student, etc.):
Title:
Degree(s) the individual holds (e.g., MLS, discipline Masters, discipline PhD, etc.):
To whom the position reports:
URL for the job posting or position description:

21. Library Professional and Workforce Development, cont.

Has your library provided opportunities for staff to develop skills related to e-science?

Yes
No

If yes, please indicate the type(s) of opportunities that have been provided. Check all that apply.

In-house staff workshops, presentations
Support for staff to attend e-science conferences, meetings
Support for staff to take coursework related to e-science or data management in a discipline
Support for professional workshops elsewhere (e.g., ICPSR summer program)
Other, please describe

Does your library collaborate with an "I School" or other academic program to develop professionals with skills relating to e-science or data management?

Yes
No

If yes, please briefly describe the program that is offered.

22. Multi-institutional, collaborative activity

Is your institution involved in a collaborative program with other institutions that support e-science?

Yes
No

If yes, does your library play a role?

Yes
No

If yes, please briefly describe each of the participants and their contribution to the collaborative effort.

23. Pressure Points

Please describe particular pressure points for your institution and your library related to e-science support or e-research more broadly.

24. E-science/Data Management Information Exchange

Would your library be interested in participating in an information exchange on e-science/ data management?

Yes, we're ready to share
Yes, but we might not be ready to contribute
Not at this time
No, not of interest

Please briefly describe the kind of information that could be exchanged in a community forum to advance e-science and data management services.

25. Additional Comments

Please provide any additional comments about your institution's or library's efforts related to e-science support or e-research more broadly that may assist the authors in accurately analyzing the results of this survey.

26. Requested Documents

If the documents requested throughout the survey are not available on the Web or if the URL is for a page that is accessible only by the library staff, mail or e-mail the document(s) by **September 14, 2009** to:

Tricia Donovan
ARL E-Science Survey
21 Dupont Circle NW
Suite 800
Washington, D.C. 20036

OR

tricia@arl.org

NB: Submitted documents may be chosen for inclusion in the published survey report. Please alert Tricia if a document should not be published.

27. Thank You

Thank you for your contribution to this survey!

Questions about the survey, or a request for a PDF of your survey response, can be directed to Tricia Donovan.

Appendix II: Recent Relevant Job Descriptions

Data Services Coordinator, University of Washington

April 13, 2010

Internal Applicants Only

Temporary 2-year appointment (renewable)

LOCATION: Suzzallo and Allen Library

Reference and Research Services Division

THE POSITION:

This position is an exciting opportunity for an energetic and entrepreneurial UW Seattle librarian holding a provisional/permanent appointment to assume a leadership position in defining, establishing, and implementing a data services program in the Libraries. The report of the Data Services Program Planning Committee will provide initial planning guidance.

The Data Services Librarian reports to the Head of Reference and Research Services. Located in the Suzzallo and Allen Libraries, the Reference and Research Services Division is comprised of the following units: Government Publications, Information Services, Maps & Cartographic Services, Microform and Newspaper Collection, Reference Services, and the Research Commons.

SPECIFIC RESPONSIBILITIES AND DUTIES:

· Acts as "data concierge" to refer clients to data from multiple sources and provides both in-person and virtual consultation services.

· Pursues and develops connections with other entities (UW and beyond) offering data services; promotes collaboration and referral as well as substantive partnerships.

· Coordinates data collection and consultation among other subject librarians, especially those in data-intensive disciplines.

· Coordinates and offers training (face-to-face and virtual) in areas of data discovery and data literacy.

· Serves as a member of the Data Services Team (cross-functional team of staff in Libraries and possibly other units on campus who work together to support data users and usage).

· Works closely with the Research Commons librarian on developing and delivering on-site and virtual products and services.

· Engages in outreach activities and marketing of data services.

· Maintains current awareness and understanding of developments and trends in data services.

· Assumes other responsibilities as assigned; performs other duties as required.

APPOINTMENT AND STIPEND:

The initial appointment is for two years, renewable at the discretion of the Dean of University Libraries. At the end of the appointment term, the Data Services Coordinator will return to his/her former position or a similar position at the University of Washington Libraries. Every effort will be made to accommodate the preferences of the librarian in determining the new assignment.

For the duration of the appointment, the Data Services Coordinator will receive a salary stipend for serving in this position.

The Data Services Coordinator will likely retain some aspects of his/her current position responsibilities. The successful candidate will be able to work with her/his current supervisor and Nancy Huling, Head, Reference and Research Services, to identify the right mix of new and existing duties.

APPLICATION PROCESS:

Candidates are to submit a statement of interest, of no more than two pages, focusing on their background and experience to serve in this position. Candidates should demonstrate evidence of an ability to work collaboratively with faculty, librarians, and other partners across the campus in meeting the data services needs of faculty, graduate, and undergraduate students.

Candidates should possess excellent oral and written communication skills and analytical skills, and demonstrate evidence of proficiency with data management and analysis. An entrepreneurial spirit and record of successful collaboration is required. A demonstrated commitment to diversity and an understanding of the contributions a diverse workforce brings to the workplace is also required.

Statements of interest, along with a current resume or CV, are to be e-mailed to Charles Chamberlin, Senior Associate Dean, at cecuwa@uw.edu by Wednesday, April 28, 2010.

Science and Emerging Technologies, Temple University

Date Posted: April 2010

The Temple University Libraries seek a user-oriented and innovative librarian to join the staff of its Science, Engineering and Architecture Library (SEAL). SEAL is located on the main campus of Temple, a vibrant, urban research university that is among the most diverse in the nation. For more information about Temple and Philadelphia, visit www.temple.edu/about/.

Description:
Reporting to the Head of the Science, Engineering & Architecture Library, the Science and Emerging Technologies Librarian will provide a full range of multi-disciplinary reference, research and consultative services for students and faculty. Responsibilities include: working collaboratively with faculty to assist them with information literacy instruction, including designing effective assignments and delivering classroom-based user education; maintaining awareness of emerging technologies and working both collaboratively and independently to adapt technologies that improve access to and use of information; promoting the use of good data management practices to support data preservation, access and re-use; consulting with University departments and offices to identify appropriate data management practices for each discipline; utilizing social networks and other technologies to create awareness about library services and resources; providing training and support to improve use of information technology by library staff; designing or collaborating with other librarians to produce web-based research guides and tutorials; serving as the liaison to three or more science departments; supervising student assistants or other staff, as required; performing collection development and managing the collection budgets in those assigned areas; and participating as scheduled in shared evening and weekend service hours. In addition, the incumbent will participate in library-wide activities and serve on library or university committees; and is expected to be active professionally.

Compensation:
Competitive salary and benefits package, including a relocation allowance. Librarian rank and compensation will be commensurate with qualifications and experience.

Required Education:
ALA accredited MLIS.

Required Skills and Abilities:
*Demonstrated knowledge of current technology trends as they apply to the design and delivery of instruction and information services.
*Demonstrated ability to provide science reference, consultation and liaison services in an academic setting.
*Demonstrated ability to provide instruction in an academic setting.
*Strong understanding of information literacy in an academic setting.
*Knowledge of trends in e-science and scholarly communications in relevant disciplines.
*Demonstrated ability to develop and manage science-related collections in all formats.
*Strong analytical, organizational, customer service, interpersonal, and communication skills.
*Demonstrated ability to apply existing and emerging technology to new projects/ventures.
*Ability to work both independently and collegially in a demanding and rapidly changing environment.

Preferred:
*Undergraduate degree in science or engineering, or relevant experience in one of the science disciplines. Coursework in Chemistry or biology strongly preferred.
*Knowledge of data sets in numeric or other formats.

E-Science Librarian, University of North Carolina at Chapel Hill

Date Posted: February 4, 2010

The University of North Carolina at Chapel Hill seeks an innovative, collaborative, and service-oriented individual for the position of E-Science Librarian. The E-Science Librarian will serve as the subject librarian for chemistry. The Librarian will work with the science research community and library colleagues to develop and sustain resources and services that assist faculty and students with preserving their own and accessing others' research data, with a focus on chemical informatics. The E-Science Librarian will develop and maintain close relationships with faculty, graduate students, and undergraduates in the assigned and related academic disciplines to ensure the highest and most effective level of library support for their research, teaching, and learning.

The E-Science Librarian participates in a team of subject librarians who share responsibility for developing high quality collections and delivering both general and specialized reference, research and instructional services. Within this context, they participate in long-term planning, conduct on-going assessment of collections and services, develop web-based guides and other research and learning products, collaborate on special projects, and serve on committees and task forces as needed.

The subject librarian has primary responsibility for selecting and managing collections in all formats for the assigned subjects. In addition, the person in this position works with colleagues in media, special collections, digital publishing, and curation units to develop and promote the library's unique resources and digital services whenever appropriate.

The E-Science Librarian will work closely with members of the Data Management Working Group to develop sustainable library services for campus researchers that support archiving and accessing their research data. The Librarian will also maintain awareness of tools and methodologies for computationally centered, data-driven research (data mining, visualization, text mining, etc.). The E-Science Librarian is also expected to participate actively in and contribute to the work of library and campus committees, professional organizations and initiatives dealing with data and metadata.

The E-Science Librarian will oversee the operation of the Kenan Chemistry Library, currently under construction, anticipated to open summer 2010. For more information, visit http://www.lib.unc.edu/science and http://sallisaw.chem.unc.edu/alumni/.

Qualifications Required:
ALA accredited master's degree in library or information science. Proven ability to effectively manage and deliver on multiple projects. Demonstrated subject knowledge and experience with relevant online resources. Ability to think creatively in developing and promoting the use of collections through services, such as workshops, course-integrated instruction, and other outreach efforts. Strong commitment to public service. Excellent oral and written communication skills. Excellent interpersonal skills and ability to work well with diverse population of faculty, students, and academic colleagues.

Preferred:
Significant study or a second advanced degree in chemistry or related science discipline. Three years professional experience as a librarian. Experience in managing a branch library. Supervisory experience. Experience with data sets in numeric or other formats (images, GIS, video, etc.). Experience with SciFinder Scholar, Beilstein or Reaxys, and other chemical databases.

Data Librarian for Business and Economics, University of New Mexico Libraries

The University of New Mexico Libraries (UL) has an opening for a Data Librarian for Business and Economics. This position is a full-time, 12-month tenure track faculty position with the rank of Assistant Professor. The desired start date is January 3, 2011. The annual salary is negotiable based on qualifications and includes full benefits.

Position Description

Working in a team-oriented and highly electronic environment, the Data Librarian will serve as the Library's subject and data specialist for business and economics, acting as liaison and data use expert in those disciplines. This Librarian will provide instruction in research and the use of library resources in a variety of settings. S/he will serve as a key library instructor. S/he will keep current with emerging information technologies especially Web 2.0 functions and trends in scholarly communication, anticipating and facilitating new uses of social science research especially in the various areas of business, e-scholarship and digital tools, and research data in response to evolving patterns of publishing and information dissemination. The Librarian will also participate in many types of outreach services, guiding and instructing library patrons in identifying, retrieving, evaluating, and curating data in all formats. S/he will also play an active role in the development, marketing and enhancement of outreach services in general. The Data Librarian will establish and maintain strong interpersonal communications and will employ organizational, analytic, and problem-solving skills. Working some evenings and weekends is required. The Librarian will participate in faculty governance as detailed in the UNM Faculty Handbook.

Education and Experience

Minimum Requirements:

Earned Master's degree by start date from an ALA-accredited Library/Information Science program or international equivalent.

Two years (24 months) of direct information service experience in a research library within the last 5 years.

Earned degree in social sciences such as business, economics, public administration, or related disciplines.

Preferred (Desired) Qualifications:

- One year experience as a research library business subject specialist.
- Library instruction or teaching experience using current and emerging technologies.
- Demonstrated knowledge and proficiency with contemporary and emerging information technologies such as web authoring tools, digital learning objects, LibGuides, Web 2.0, social software, informatics, etc.
- Demonstrated interest in Latin American economics.
- Experience in database and collection evaluation and development.
- Demonstrated problem-solving and analytical skills.
- Excellent oral, written, and interpersonal communication skills.

Primary Duties

Act as liaison and data expert in Business and Economics disciplines. Provide information and consultation at service points and by appointment on a schedule which includes evenings and weekends. Provide instruction in research, in the use of library resources, and in the collection and management of data in a variety of settings, with special emphasis on social sciences and use of data. Provide effective and timely supervision of any assigned employees including training, career development, and performance management. Participate in faculty governance meetings, as required, and in library management meetings as required. Contribute to Library initiatives that further UNM's commitment to diversity and inclusion.

Environment
UNM enrolls nearly 27,000 students and employs 2,800 faculty and 4,400 staff. UNM offers 102 baccalaureate degrees, 75 master's degrees, and 45 doctoral degrees/professional degrees. The University of New Mexico is a Tier I Research Institution and a Hispanic-Serving Institution. UNM attracts a culturally diverse student population and has strong academic and research programs concerned with the Southwestern United States, indigenous studies, and Latin America.

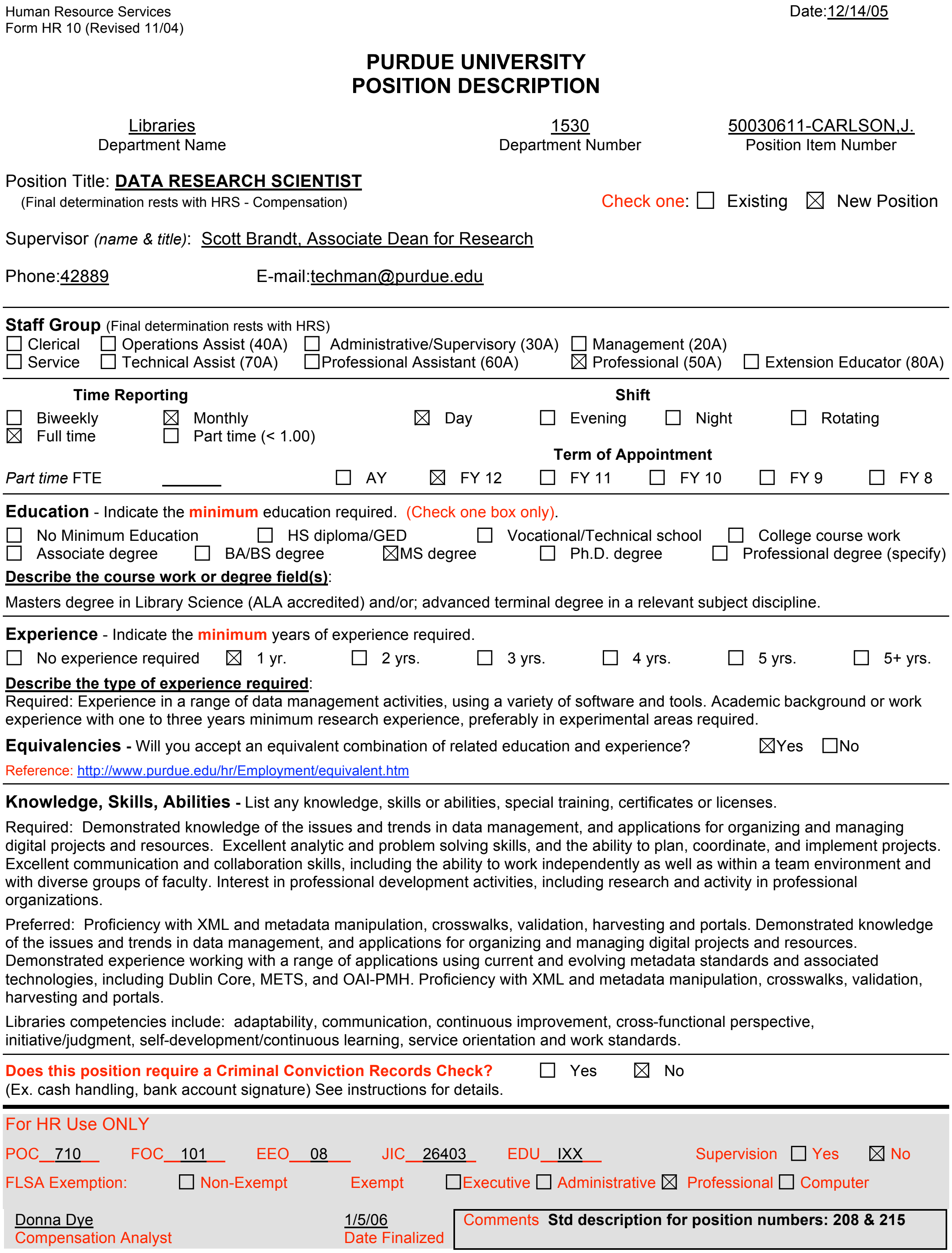
UNM is located in Albuquerque, New Mexico. Albuquerque is ranked number one in creativity among medium-sized cities in Richard Florida's book "Rise of the Creative Class." Albuquerque is an ethnically diverse city with a rich culture and history located within minutes of the Sandia and Manzano mountain ranges which provide opportunities for hiking, biking, rock-climbing, and skiing.

To apply: Please visit UNMJobs at http://unmjobs.unm.edu/

Deadline
The search will remain open until the position is filled. For best consideration, complete applications must be received through the UNMJobs website no later than August 16, 2010.

UNM's confidentiality policy ("Disclosure of Information about Candidates for Employment," UNM Board of Regents' Policy Manual 6.7), which includes information about public disclosure of documents submitted by applicants, is located at http;//www.unm.edu/~brpm/r67.htm

The University of New Mexico is an Equal Employment Opportunity/Affirmative Action Employer and Educator.



Human Resource Services
Form HR 10 (Revised 11/04)

Date:12/14/05

PURDUE UNIVERSITY
POSITION DESCRIPTION

Libraries	1530	50030611-CARLSON,J.
Department Name	Department Number	Position Item Number

Position Title: **DATA RESEARCH SCIENTIST**
(Final determination rests with HRS - Compensation)

Check one: □ Existing ☒ New Position

Supervisor *(name & title)*: Scott Brandt, Associate Dean for Research

Phone:42889 E-mail:techman@purdue.edu

Staff Group (Final determination rests with HRS)

□ Clerical □ Operations Assist (40A) □ Administrative/Supervisory (30A) □ Management (20A)
□ Service □ Technical Assist (70A) □ Professional Assistant (60A) ☒ Professional (50A) □ Extension Educator (80A)

Time Reporting

□ Biweekly ☒ Monthly
☒ Full time □ Part time (< 1.00)

Shift

☒ Day □ Evening □ Night □ Rotating

Term of Appointment

Part time FTE _______ □ AY ☒ FY 12 □ FY 11 □ FY 10 □ FY 9 □ FY 8

Education - Indicate the **minimum** education required. (Check one box only).

□ No Minimum Education □ HS diploma/GED □ Vocational/Technical school □ College course work
□ Associate degree □ BA/BS degree ☒MS degree □ Ph.D. degree □ Professional degree (specify)

Describe the course work or degree field(s):

Masters degree in Library Science (ALA accredited) and/or; advanced terminal degree in a relevant subject discipline.

Experience - Indicate the **minimum** years of experience required.

□ No experience required ☒ 1 yr. □ 2 yrs. □ 3 yrs. □ 4 yrs. □ 5 yrs. □ 5+ yrs.

Describe the type of experience required:
Required: Experience in a range of data management activities, using a variety of software and tools. Academic background or work experience with one to three years minimum research experience, preferably in experimental areas required.

Equivalencies - Will you accept an equivalent combination of related education and experience? ☒Yes □No

Reference: http://www.purdue.edu/hr/Employment/equivalent.htm

Knowledge, Skills, Abilities - List any knowledge, skills or abilities, special training, certificates or licenses.

Required: Demonstrated knowledge of the issues and trends in data management, and applications for organizing and managing digital projects and resources. Excellent analytic and problem solving skills, and the ability to plan, coordinate, and implement projects. Excellent communication and collaboration skills, including the ability to work independently as well as within a team environment and with diverse groups of faculty. Interest in professional development activities, including research and activity in professional organizations.

Preferred: Proficiency with XML and metadata manipulation, crosswalks, validation, harvesting and portals. Demonstrated knowledge of the issues and trends in data management, and applications for organizing and managing digital projects and resources. Demonstrated experience working with a range of applications using current and evolving metadata standards and associated technologies, including Dublin Core, METS, and OAI-PMH. Proficiency with XML and metadata manipulation, crosswalks, validation, harvesting and portals.

Libraries competencies include: adaptability, communication, continuous improvement, cross-functional perspective, initiative/judgment, self-development/continuous learning, service orientation and work standards.

Does this position require a Criminal Conviction Records Check? □ Yes ☒ No
(Ex. cash handling, bank account signature) See instructions for details.

For HR Use ONLY

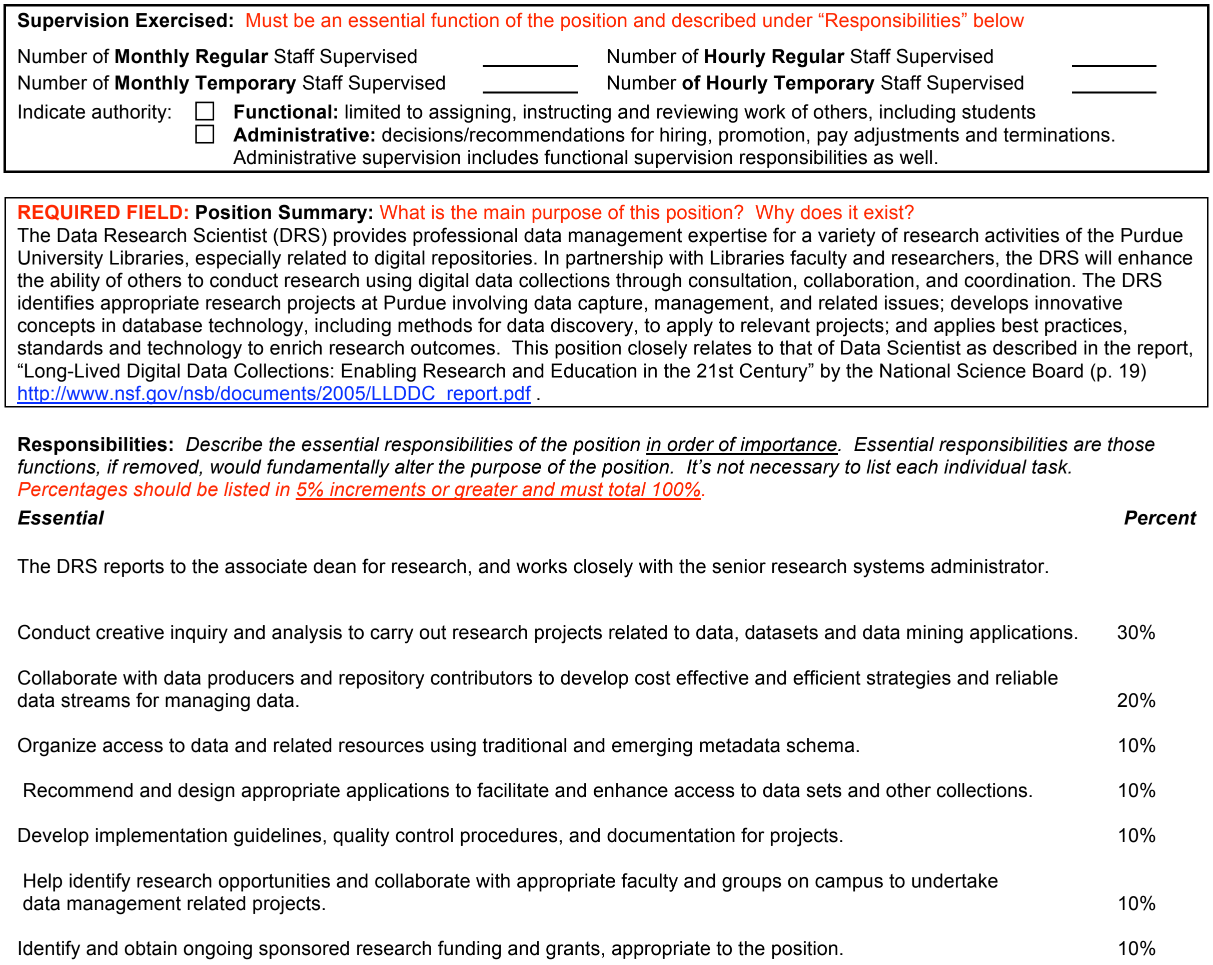
POC 710 FOC 101 EEO 08 JIC 26403 EDU IXX Supervision □ Yes ☒ No

FLSA Exemption: □ Non-Exempt Exempt □Executive □ Administrative ☒ Professional □ Computer

Donna Dye
Compensation Analyst

1/5/06
Date Finalized

Comments **Std description for position numbers: 208 & 215**



Supervision Exercised: Must be an essential function of the position and described under "Responsibilities" below

Number of **Monthly Regular** Staff Supervised ________ Number of **Hourly Regular** Staff Supervised ________
Number of **Monthly Temporary** Staff Supervised ________ Number **of Hourly Temporary** Staff Supervised ________

Indicate authority: □ **Functional:** limited to assigning, instructing and reviewing work of others, including students
□ **Administrative:** decisions/recommendations for hiring, promotion, pay adjustments and terminations. Administrative supervision includes functional supervision responsibilities as well.

REQUIRED FIELD: Position Summary: What is the main purpose of this position? Why does it exist?
The Data Research Scientist (DRS) provides professional data management expertise for a variety of research activities of the Purdue University Libraries, especially related to digital repositories. In partnership with Libraries faculty and researchers, the DRS will enhance the ability of others to conduct research using digital data collections through consultation, collaboration, and coordination. The DRS identifies appropriate research projects at Purdue involving data capture, management, and related issues; develops innovative concepts in database technology, including methods for data discovery, to apply to relevant projects; and applies best practices, standards and technology to enrich research outcomes. This position closely relates to that of Data Scientist as described in the report, "Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century" by the National Science Board (p. 19) http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf .

Responsibilities: *Describe the essential responsibilities of the position in order of importance. Essential responsibilities are those functions, if removed, would fundamentally alter the purpose of the position. It's not necessary to list each individual task.*
Percentages should be listed in 5% increments or greater and must total 100%.

Essential	***Percent***
The DRS reports to the associate dean for research, and works closely with the senior research systems administrator.	
Conduct creative inquiry and analysis to carry out research projects related to data, datasets and data mining applications.	30%
Collaborate with data producers and repository contributors to develop cost effective and efficient strategies and reliable data streams for managing data.	20%
Organize access to data and related resources using traditional and emerging metadata schema.	10%
Recommend and design appropriate applications to facilitate and enhance access to data sets and other collections.	10%
Develop implementation guidelines, quality control procedures, and documentation for projects.	10%
Help identify research opportunities and collaborate with appropriate faculty and groups on campus to undertake data management related projects.	10%
Identify and obtain ongoing sponsored research funding and grants, appropriate to the position.	10%

Association of Research Libraries