User-needs assessment of the research cyberinfrastructure for the 21st century

Lisa Johnston

University of Minnesota, ljohnsto@umn.edu

Follow this and additional works at: http://docs.lib.purdue.edu/iatul2010
User-Needs Assessment of the Research Cyberinfrastructure for the 21st Century

Lisa Johnston, Physical Sciences and Engineering Librarian, University of Minnesota Libraries

Abstract

In support of the University of Minnesota’s goal to become one of the top three public research universities, the Research Cyberinfrastructure Alliance (RCA) was developed with the vision of facilitating access to state-of-the-art research computing systems and services, enhancing interdisciplinary research, and allowing researchers to explore radically new concepts, approaches, and tools. Founding members of the University’s RCA included individuals from college-level research computing units, the Minnesota Supercomputing Institute, and key leaders from the University’s Office of Information Technology (OIT), the Office of the Vice President for Research (OVPR), and the University Libraries.

A key step for the RCA was to understand campus e-science needs and identify the challenges of engaging with relevant research computing resources and support. As a project of the 2009 President’s Emerging Leaders (PEL) program, our five member team of interdepartmental university staff was commissioned by the RCA to help lead this effort and recommend ways that the RCA university partners might respond.

In 2009 our team conducted an extensive user-needs assessment of 780 university faculty, research staff, and graduate students. The PEL survey assessed the current and future cyberinfrastructure needs in the following areas: data storage, data management, and networking infrastructure; collaboration with other researchers; tools and applications; high performance computing; and learning and workforce development.

The results of our PEL survey reflect a general need for e-science support and training that may affirm and further explain what other science and technology libraries are observing. Our formal recommendations and the resulting strategies toward implementing cyberinfrastructure for 21st century research will be described with emphasis on the opportunities and future roles that university libraries have in this campus-wide partnership.

INTRODUCTION

At the National level, several influential organizations are paving the way for institutions to develop cyberinfrastructure for the 21st century. Reports like the National Science Foundation’s Blue Ribbon Report on Cyberinfrastructure (Atkins, 2003), data sharing policies from the National Institutes of Health (NIH), and, in the corporate sector, Microsoft’s Towards 2020 Science (2006), each make it clear that research cyberinfrastructure, or the combined power of computing resources and expertly trained staff, has the potential to have an unprecedented impact on our world in the 21st Century.

For the library world, these sediments could not be summarized with a stronger call to action than the 2009 National Academies of Science’s report, "Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age," outlining the need for better practices in data management for the future sustainability of data in the digital age. Specifically it recommendation steps to ensure that important research data are managed, shared, and retained for future use. Cyberinfrastructure for the 21st century is a common goal of libraries and university campus leadership alike.

At the University of Minnesota, the Research Cyberinfrastructure Alliance (RCA) was formed in 2008 with the vision of facilitating access to state-of-the-art research computing systems and services, enhancing interdisciplinary research, and allowing researchers to explore radically new concepts, approaches, and
tools. Founding members of the University’s RCA included individuals from college-level research computing units, the Minnesota Supercomputing Institute, and key leaders from the University’s Office of Information Technology (OIT), the Office of the Vice President for Research, the Minnesota Super Computing Institute, and the University Libraries.

A key step for the RCA was to understand campus e-science needs and identify the challenges of engaging with relevant research computing resources and support. As a project of the 2009 President’s Emerging Leaders (PEL) program, our five member team of interdepartmental university staff was commissioned by the RCA to help lead this effort toward "Implementing Cyberinfrastructure for 21st Century Research" (RCA Project, 2008).

As a member of the PEL survey team, I was able to work with key stakeholders on campus (faculty, staff, students, collegiate IT directors, OIT staff, and the RCA group) to develop a strategic approach to cyberinfrastructure that addressed the growing need on campus and leveraged the partnerships across the University. Our report was designed to serve as the springboard for stakeholders to implement our plan to help the University become a leader in academic research, increase the University’s competitiveness for funding from federal agencies, and, finally, spark new types of research and collaborations. As a librarian, my take on the survey results and its implications on campus were also intended to foster an environment of data preservation, sharing, and retention for future use: in short, position the library to support e-science and research cyberinfrastructure at the University of Minnesota. The results demonstrated many opportunities for library engagement.

PEL SURVEY METHODOLOGY AND DEMOGRAPHICS

The President's Emerging Leaders (PEL) is a 12-month development program for selected university staff to help prepare them for future leadership opportunities on campus. Each cohort, in addition to building individual leadership skills, addresses a major campus issue by implementing an experimental project and presenting the results to University leaders. As a project of the 2008-2009 President’s Emerging Leaders (PEL) program, our five member team of interdepartmental university staff was commissioned by the RCA to conduct an extensive user-needs assessment of university faculty, research staff, and graduate students on current state of cyberinfrastructure at the University and assess future needs.

During March 24th-April 8th, 2009, our team conducted an online survey of faculty, research staff and students across the University of Minnesota’s four campuses consisting of 130 questions (see full survey instrument online) on the following cyberinfrastructure topics: data storage, data management, and networking infrastructure; collaboration with other researchers; tools and applications; high performance computing; and learning and workforce development, as well as future trends within each of these areas. We had 780 successful responses (9.2% response rate) from a nice cross-section of affiliates representing various roles (Figure 1), research disciplines (Figure 2), research environments (Figure 3), and university campuses (Figure 4). The survey also generated over 300 individual comments and qualitative data.

*Figure 1: PEL Survey Response by Research Role*
Figure 2: PEL Survey Response by Research Discipline

Figure 3: PEL Survey Response by Research Environment
Figure 4: PEL Survey Response by Campus Affiliation

<table>
<thead>
<tr>
<th>UMN Campus</th>
<th>Researchers Solicited</th>
<th>Successful Responses</th>
<th>Response rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Twin Cities</td>
<td>7767</td>
<td>715</td>
<td>9.2%</td>
</tr>
<tr>
<td>Duluth</td>
<td>493</td>
<td>50</td>
<td>10.1%</td>
</tr>
<tr>
<td>Morris</td>
<td>102</td>
<td>9</td>
<td>8.8%</td>
</tr>
<tr>
<td>Crookston</td>
<td>41</td>
<td>6</td>
<td>14.6%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>8403</strong></td>
<td><strong>780</strong></td>
<td><strong>9.3%</strong></td>
</tr>
</tbody>
</table>

PEL SURVEY RESULTS AND RECOMMENDATIONS

The results of our PEL survey reflect a general need for e-science support and training that may affirm and further explain what other science and technology libraries are observing. The survey responses highlighted several trends in current practice, including: personal control of data storage and backup, strong preference for a local point of contact for research support, a variety of patterns in how data is generated and used, inadequate use of terminology and awareness, and a desire for more support around interdisciplinary practices. Finally, the opportunities and future roles that university libraries have in this campus-wide partnership are examined.

PEL Survey Trend: Personal control over data storage and backup

Researchers prefer personal or departmental-level data storage and back-up options over University- or cloud-level storage. As shown in Figure 5, only 14% of respondents use central data storage solutions, even though every researcher on campus has an online file storage account administered by OIT. Instead researchers relied more heavily on work computer desktops and laptops (63% and 39% respectively) and
departmental servers (40%). External hard drives and other media were also noted by 15% of respondents.

Back-up solutions also show a preference for local and "low-tech" storage (Figure 6). Nearly half of respondents (43%) reported using a secondary hard drive to back-up their data, and 29% use CD/DVD or other removable media. This is a serious security issue for many research's data. Interestingly, 72% of researcher reported never losing data due to the lack of a back-up; 27% admitted that they have.

Figure 5: PEL Survey Response to "Where do you store your research data electronically?"

Figure 6: PEL Survey Response to "How are your electronic data backed up?"
It was evident by the comments why this was occurring. A post-doc in the College of Science and Engineering summarized, "In my research group, research data is mainly managed by the respective owner. We evaluated the file sharing service from central IT but found it too cumbersome to use. It works best if research groups can set up their own servers to have control over their data storage platform."

However other comments reflect a desire for more central support. A PhD research assistant in the same college stated, "There needs to be a mechanism for centralized storage of data with high security for individual research groups. This at least partially removes the responsibility for secure backups from the research group alone."

Finally, storage solutions and backup were consistent across all disciplines but vary by user type: graduate student researchers are more likely to store data on personal computers and faculty are more likely to store data centrally.

**PEL Recommendation 1:** Develop enterprise-wide, integrated cyberinfrastructure by leveraging and better aligning existing services. This might be done by defining a system-wide University policy to streamline how data is stored and secured as well as more formal education data management practices at the university. Finally, it might be possible to leverage existing services at the University, and promote them as system-wide solutions while simultaneously discouraging the development of "as needed" solutions to research computing. For example, a PI might purchase large storage for a project which requires ongoing maintenance not covered by the grant, rather than relying on central services already in place (i.e. the University’s central Active Directory service that provides departmental and individual file storage at no cost). Training and awareness of existing services has the potential to lead to significant cost-savings for the University.

**PEL Survey Trend: Local point of contact**
The Local IT point of contact, such as the departmental or collegiate IT admin, is the primary end-user resource for most cyberinfrastructure needs. Figure 7 shows this trend was consistent for research needs and services including Data storage (68%), electronic security (65%), space/environment (56%), and systems administration (64%).

However researchers are more evenly split between local and university-level support when it comes to Compliance & privacy issues (local support is contacted by 41% vs. 23% who contact university-level support), bandwidth (local 46% vs. university-level 34%), and obtaining outside data (local 26% vs. university-level 16%).

Outside data services are not needed for many researchers, particularly the Health Sciences. They are more likely to contact university-level service for support (15% vs~5%). Could be an artifact of how researchers in the Academic Health Center, the largest unit on campus, view "university level" support.

Figure 7: PEL Survey Response to "Who is your primary contact for providing each of these services?"

Researchers also indicated a preference (very large to moderate) for contacting their local point of support for obtaining access to:

- data storage (83% very large reliance and 55% moderate reliance),
- virtual collaboration (64% and 41% respectfully),
- software and tools (85% and 55% respectfully),
- and, to a lesser extent, centrally located high performance computing services (47% and 23% respectfully).

Although researchers rely on their local IT point of contact for support, they may not be satisfied with the quality of service and understanding they receive. "It's hard to know who to contact in the tech area for some research tech support. The basic tech person assigned to the department is fine for routine computer and computer related software things, but beyond the basics it's not clear who to contact for what" says one Faculty Member in the College of Liberal Arts. A post-doc in the School of Density was
also unsatisfied with the support in the department saying, "As far as I know, we don't have a contact person for data storage. We just save our data in multiple places and on an external hard drive. There is no oversight for data storage." Comparing the responses to the current level of IT support (Figure 8) versus the support necessary (Figure 9) reveals some interesting expectations.

*Figure 8: PEL Survey Response to Rate the overall level of support you receive from your primary technology point of contact in the following areas.*

*Figure 9: PEL Survey Response to What overall level of support should your primary technology point of contact offer in the following areas.*

**PEL Recommendation 2:** Make use of the local point of contact by better positioning cyberinfrastructure services around the end-user. This can be accomplished by providing local IT staff with the tools and knowledge to share and leverage existing resources across University service providers.
This respondent, a faculty member in the College of Liberal Arts (CLA) sums up this idea nicely: "Instead of looking at individual requirements for storage, management, software and security take a look at how CLA now manages its research services. There is one single point of contact, they provide space, security, management, software -- all you do is apply and you get an account WITH phone or email support!"

Other strategies include building incentives to incorporate the local IT staff early on in the grant planning process. This way cyberinfrastructure needs can be identified with trained staff and existing resources in mind. Perhaps IT staff can take on more of a liaison role, like librarians, and with a broad knowledge connect end-users with resources and support around campus. This is in contrast to the current model where researchers must contact individual services on campus looking for support.

**PEL Survey Trend: Researcher needs related to data production, access, and retention**

The amount of data produced by the majority of researchers (80%) is less than 1 gigabyte (GB) of data per week. This is contrary to the wide-spread belief that all researchers need unlimited amounts of storage. In fact, because central-level storage options only allow 5GB-10GB of storage space per user (on par with the average Web email account), researchers are dissatisfied in the perceived lack of space. "Once or twice a year I have to negotiate for storage space; I always feel as if my research is at odds with IT policy. I feel as if I'm living on borrowed time, no confidence in having access to adequate data storage for research in the future," says a Faculty Member in the College of Education & Human Development.

Figure 10 shows that of the 19% of researchers who do generate over 1GB of data per week, only a small fraction of these generally produce terabytes of data (or 1024 GB) per week. Interestingly, these large data produces occur in nearly all disciplines. Of those who generate over 1 GB of data per week, Biological sciences represent 21% of responses, Physical sciences15%, Engineering14%, Health sciences 19%, Interdisciplinary fields11%, and Social Sciences 15%. Only arts and humanities, each representing 2%, where less evenly distributed.

*Figure 10: PEL Survey Response of the participants who generate over 1GB of data on average per week (19% of respondents) how much data do you generate.*

All researchers may not be producing terabytes of data, but they do want constant access to and indefinite retention of their data. Most researchers (64%) access their data every day (Figure 11). Daily access is most consistent in the physical sciences (89%) and by many in social sciences (75%), biological
(75%), engineering (71%), humanities (75%), and interdisciplinary (62%) fields. Arts (56%) and Heath (59%) researchers reported slightly lower daily access to their data.

Finally, and not surprising to librarians, 70% of researchers would like to keep their data forever (Figure 12). Luckily, many researchers do not require their data to be securely stored (with 40% respondents not requiring password authentication and 38% of respondents not needing their data physical secure onsite) therefore many can take advantage of remote storage options such as data repositories.

**Figure 11: PEL Survey Response to How often do you access your data?**

![Pie chart showing access frequency]

**Figure 12: PEL Survey Response to How long do you typically keep your electronic research data?**

![Pie chart showing data retention]

**PEL Recommendation 3:** In order to meet the current and future growing needs of researchers, IT service providers should improve and expand the University's ability to handle data at the central level by providing more flexible data storage and access tools. When possible, researchers should utilize external cloud storage for sharing and long-term archiving from the private sector (i.e., Amazon) taking into account the benefits of scalability, high-availability, low-latency (fast) at commodity costs.

**PEL Survey Trend: Terminology and Awareness of Existing Services**
The PEL survey questions were modeled after a cyberinfrastructure study of collegiate IT directors (EDUCAUSE, 2006). Even though our audience was primarily composed of end-user researchers, and theirs high-level IT service providers, surprisingly, we uncovered a similar trend in dissatisfaction with cyberinfrastructure terminology. The EDUCAUSE report explains:

“The very term cyberinfrastructure is part of the problem; it means everything and it means nothing. It’s not that the [National Science Foundation] was unclear when it coined the term in 2001, but its subsequent usage by the higher education community has been inconsistent, sometimes broadening and sometimes restricting its original scope. As we conducted this study, we heard many times from respondents to our quantitative survey and in our qualitative interviews that the term has become a buzzword and lacks enduring meaning. Several respondents preferred the equivalent European terms e-science and the somewhat less restrictive e-research. Although various questions in our survey asked about research outside science and engineering and inquired about CI technologies used in creative activities and teaching and learning, the bulk of our findings did reflect science and engineering research use.”

In the PEL study our results uncovered similar sediments, from the humanities disciplines in particular. A faculty member in the College of Liberal Arts stated "This instrument seems to have been constructed using the sciences as its template. I find it hard to translate the terms into my research in the Humanities." Another faculty colleague concurs, "This survey is largely geared to researchers outside the Humanities. Consequently, many if not most of the questions are stretch -- in terms of relevance. Please find a way to incorporate a genuine commitment to research in the humanities (without which no university with which we aspire to compete, either public or private, is ranked highly)."

Another potential problem associated with terminology is inconsistent awareness of existing resources. There was no central understanding of what research resources were available on campus. Also, there was a misunderstanding of existing resources. For example, many researchers incorrectly stated that they had no automatic backup of their files, yet they were using central storage solution that contained regular automated backup. This problem lies in the users' perception, as shown in this comment, “There is no automated backup of office or lab computers to central University server provided, as I have had at other Big 10 universities” as well as from the lack of training of our IT professionals, as described by this researcher, “When I started working at this lab (2.5 years ago) there was no tech support or data backup system in place. I asked my department administrators how I can backup my lab data and they did not know.”

PEL Recommendation 4: In order to bridge the jargon gap among the research community, the IT service providers and University leaders should re-brand cyberinfrastructure to meet the research for all disciplines (ie. Research Computing or E-Science). Also, needs may be temporarily met by more effective promotion and awareness of current technology and resources on campus through an online resource directory or glossary of resources with descriptions appropriate for grant application use. Finally, educate and engage end-users consistently through appropriate venues such as New Faculty Orientation, graduate/professional student orientation, and Responsible Conduct in Research Seminars for PI’s.

PEL Survey Trend: Interdisciplinary research involves collaborating and sharing data

Even though most researchers do not believe that effective incentive exist to encourage collaboration on campus (Figure 13), most share their data anyway. Data is shared by 92% of all researchers (Figure 14), and primarily with researchers in their own unit or researchers on campus (51%, 18% respectively). Only 8% of researchers do not share their data at some level. Of those, 52% would share their data, 30% would not and 19% did not know.
Compared by discipline, the majority share their data in statistically balanced ways with Biological and Health Sciences leading the way with sharing within their unit. Of those 8% of respondents who are not sharing their data, the arts and humanities disciplines represent the majority of responses with 44% and 50% respectively.

Still, some researchers wisely recognize the concerns around data sharing, like this faculty member in the College of Education and Human Development suggests: "International collaboration will increase; how do we encourage & assure compliance with security standards from our collaborators?"

Finally, perhaps not surprisingly, researchers feel that department data assets and resources are not well documented. The majority of all disciplines do not know of any documented inventory of data sets within their department.

*Figure 13: PEL Survey Responses to* Within my department/unit, effective incentives exist to encourage researchers to share technology with other researchers on campus.*

![Pie chart showing survey responses to share technology]

*Figure 14: PEL Survey Responses to With whom would you share your data?*
Figure 15: PEL Survey Responses to What is the primary source of your research funding?

Note: “Domestic Other” includes specific internal, private, and government grant funding bodies.

Figure 16: PEL Survey Responses to Do you comply with your primary funding agency’s suggested data management plan?
**PEL Recommendation 5:** The university RCA partners should work with departments to create incentives for collaboration and provide the necessary tools to facilitate local and external data sharing and build virtual collaborative relationships. This involves working with researchers on using virtual collaboration tools such as VPN and video conferencing. Also, it is necessary to create a universal platform for data sharing within the University, and one that interfaces well with external audiences. This platform may need to be modified slightly within each major area of research (i.e., social sciences, humanities, science).

**PEL Survey Trend: Library Opportunities for Involvement**

Library resources and services are mentioned several times in the PEL survey comments, though not specifically addressed in the questions. For example comments from faculty members question the omission of library cyberinfrastructure: “These questions do not address information management resources that are or should be provided by central library services,” and “What about access to remote library databases?” state separate faculty members in the humanities. Also a faculty member in the social sciences suggests, “Better centralized archiving and metadata for large data sets. [Digital] Conservancy and other library initiatives are promising.” Also,

Also, in response to the question Who do you contact when obtaining outside data? Several answers references library resources such as: “library and indexes,” “UMN Libraries,” “Interlibrary Loan,” and “Libraries have some good resource contracts.”

There were also responses suggestion the library take on more responsibility in providing technology and collaboration support. For example, a post-doc in the College of Liberal Arts suggests, “I don't mean to under appreciate our tech staff in this (they have their hands full), but as younger faculty are hired with higher expectations of online collaborative tools, we'll need people with more knowledge of software/database skills -- and I would tend to look toward the library system as a central portal for this.”

Finally, in response to the Technology needed to be more competitive for funding (Figure 17), several library resources were included in comments that suggest the library an identified provider of cyberinfrastructure, even in our more traditional roles such as collection development. For example, researchers comment:

- “I need better training regarding how to do effective searches through the internet and library on-line services”
“I don’t need any additional technology, but would appreciate greater access to article indexes and databases.”

“The library indexes (with links directly to papers) are invaluable for keeping up with whose doing what and for writing proposals and papers.”

**Figure 17: PEL Word Cloud of Survey Comments on Technology Needed to be Competitive?**

**CONCLUSIONS AND RECOMMENDATIONS FOR LIBRARY ACTION**

In the quest toward understanding the needs and challenges of researchers, it was clear that this campus-wide partnership toward implementing cyberinfrastructure for 21st century research will have future implications on the role that the library has on campus. Our formal recommendations and the resulting strategies are, briefly, that university RCA partners to:

1. Develop enterprise-wide, integrated cyberinfrastructure by leveraging and better aligning existing services.
2. Better position services around the end-user by utilizing one local-point of contact.
3. Improve and expand the University’s ability to handle research data with more flexible storage, access and retention services.
4. Re-brand and promote research cyberinfrastructure at the University to improve accessibility for all disciplines.
5. Create incentives for collaborative by enabling virtual partnerships and data sharing.

Enabling this 21st research requires large-scale investments in high-performance computing, storage and networking, as well as the development of research infrastructure to integrate these components into a meaningful whole. But our survey found that the most important component in implementing cyberinfrastructure is the human factor. Enabling the right connections of knowledgeable local IT staff with highly trained central service-providers will be key to any successful research cybinfrastructure program.

**ACKNOWLEDGMENTS:**
This work was completed by the University of Minnesota 2009 President's Emerging Leaders team: Tracy Anderson, Craig Gjerdingen, Bryan Herrmann, Katherine Himes, and Lisa Johnston.

Thanks to our sponsors Ann Hill Duin and Bernard Gulachek, our project leads Kemal Badur and John Sonnack, our advisors from the University of Minnesota Research Cyberinfrastructure Alliance, and Dave Dorman from the PEL Program.

REFERENCES


