ARL Bimonthly Report 237
December 2004

Libraries and Changing Research Practices: A Report of the ARL/CNI Forum on E-Research and Cyberinfrastructure

by Diane Goldenberg-Hart, Communications Coordinator, Coalition for Networked Information

Over 100 leaders from higher education, libraries, and information technology gathered in Washington, DC, on October 15, 2004, at a forum titled "E-Research and Supporting Cyberinfrastructure: A Forum to Consider the Implications for Research Libraries & Research Institutions," which was cosponsored by the Coalition for Networked Information (CNI) and by the Association of Research Libraries (ARL). The forum, spurred by developments in e-science and e-research, including a recent National Science Foundation report and the launch of a related study on cyberinfrastructure in the humanities, brought together representatives from over 75 different organizations.

Setting the stage for the daylong event were Sarah Thomas, University Librarian, Cornell University, and CNI Executive Director Clifford Lynch. Thomas discussed how the issues surrounding e-science and cyberinfrastructure fit well with ARL's three strategic directions: scholarly communication, information policies and other public policies, and the roles of libraries in transformations of research and education. Her comments introduced a major theme of the day: the need for (and the challenges involved in) collaboration within and among educational institutions and research organizations. Thomas emphasized that without truly effective internal collaboration, external inter-institutional cooperation could not be fully successful.

Following on with the topic of transformations in higher education, Lynch laid the groundwork for the imperative to plan and act strategically. He emphasized the massive changes occurring in the practices of scholarship—changes that are occurring across all disciplines. He argued that new practices, products, and modes of documenting and communicating research will have far-reaching implications for all organizations involved in managing the scholarly record and supporting the ongoing enterprise of scholarship, and that libraries in particular are in a central role due to their perspectives of managing the record

across time and across disciplines. These changes in scholarly practice will create profound changes throughout the entire system of scholarly communication, and a failure to put into place effective new support structures in response to these changes would pose tremendous risk to the enterprise of research and scholarship. "This is what is at stake when we consider how to lead our institutions in addressing these new needs," Lynch said. The role of libraries, he argued, will shift from primarily acquiring published scholarship to a broader role of managing scholarship in collaboration with the researchers that develop and draw upon it.

Revolutionizing Science and Engineering

Professor Daniel Atkins of the University of Michigan School of Information and Chair of the National Science Foundation (NSF) Blue Ribbon Advisory Panel on Cyberinfrastructure served as the day's keynote speaker. In February 2003, the Akins panel issued its report, *Revolutionizing Science and Engineering through Cyberinfrastructure* (http://www.cise.nsf.gov/sci/reports/toc.cfm). As part of its overall finding, the panel stated that, "a new age has dawned in scientific and engineering research, pushed by continuing progress in computing, information and communication technology…. The capacity of this new technology has crossed thresholds that now make possible a comprehensive 'cyberinfrastructure' on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research in new ways and with increased efficacy."

In his remarks at the forum, Atkins described what he referred to as cyberinfrastructure-enabled knowledge communities (CKCs), also sometimes called collaboratories, or grid communities, in which synchronous, face-to-face collaboration would not be replaced, but augmented, by a range of asynchronous and geographically dispersed modes of collaboration, promoting learning, research, and engagement. These collaboration environments would include widely distributed and accessible high-performance computation and data services, including software, which is now considered to be part of infrastructure, and, additionally, they would include services, personnel, and the participation and collaboration of various organizations. Atkins spoke with passion of the current window of opportunity for research institutions to take a systemic view of what is needed to develop and sustain support for e-research. Reactive and opportunistic use and deployment of information technology will not produce the same innovations.

"We are no longer in an era when even a well-endowed institution can proceed alone," Atkins said. What are needed are alliances among institutions, middleware, and federated libraries. He spoke of the need for the stakeholders to communicate with one another, understanding that the effort will have to be global in scope. Because collaboration is a critical element, the challenge, according to Atkins, is to find commonalities among interested groups.

Atkins explained that the digital library technologies and programs form a critical enabling component of the cyberinfrastructure movement, and he pointed out that both undertakings require an alignment of mutual self-interest among collaborators. Fundamentally, Atkins argued that knowledge communities are not just a way to do old things better, but that they could represent an opportunity for innovation, to do new things and to explore new

methods—CKCs provide economic advantages, but they also provide a space for new synergy.

In the discussion following Atkins's presentation, one key question addressed how meeting participants could return to their institutions and convince their campus leadership to invest themselves and their resources in projects of this kind. Atkins responded, "Don't wait until you are asked. Step forward and convene symposia, share visions of the possibilities and look for opportunities for collateral learning and mutual self-interests." Paul Courant, Provost, University of Michigan, added, "don't just drive it up" to institutional leadership, "drive it down" to faculty—a theme that was invoked repeatedly throughout the day.

The National Virtual Observatory

As an example of the kind of groundbreaking innovation cyberinfrastructure has enabled in the sciences, George Djorgovski, Professor of Astronomy at the California Institute of Technology, presented his experiences and impressions in a discussion entitled "Virtual Observatory, Cyber-Science, and the Rebirth of Libraries." The National Virtual Observatory (NVO) was described by Djorgovski as "a complete, dynamic, distributed, open research environment for the new astronomy with massive and complex data sets."

The concept for the NVO grew out of the astronomy community's need to cope with massive and ever-increasing data sets. Djorgovski explained that astronomers have been gathering and generating enormous amounts of data, through analyses and theoretical simulations, and, thus, there has grown an ever-increasing need for high-quality storage, management, and access methods: the provision of data-curation services has become a driving and critical issue in the field. The NVO offers a means to access stored data, but it also supplies researchers with computational services and tools with which to mine and analyze the data. Additionally, the NVO is linked to other virtual observatories around the world (the International Virtual Observatory Alliance), providing opportunities for international collaboration.

The NVO illustrates an innovation that is driven by the science but enabled by the technology; it represents a new type of scientific research environment in the field of astronomy, but the flood of increasingly complex data sets is presenting challenges across all sciences, and Djorgovski asserts that it is the advances in information technology combined with quantitative changes in data volumes that will produce qualitative changes in the way science is conducted.

The role of libraries in the midst of these sweeping transformations was, of course, on the minds of many forum participants, and the speaker expressed his own uncertainty. That libraries would be required to reconsider traditional roles was understood, and he emphasized that there would be a growing need for domain expertise. In this same vein the issue of printed material was introduced: in Djorgovski's view, and for his purposes, journals, and perhaps books, are "obsolete formats"—much of his information gathering is done through constantly changing electronic sources, including blogs and data sets, which may or may not be subject to quality control. Therefore, he said, there is an increasing need for experts to determine and communicate what is worth having or keeping. He speculated that libraries

could offer federated services much like those of the NVO. Libraries might affiliate with one another to provide two different types of services, for example, tracking on specific scientific domains and/or archiving and preserving data. These are inherently interdisciplinary tasks requiring the involvement of multiple institutions.

But many questions remain: Should there be quality control, and if so of what sort and through what mechanisms? How much domain expertise should be provided, and how should this be distributed between the work of the scholars and the work of libraries in supporting those scholars? One thing is certain, however, according to Djorgovski: the data must not be consolidated within any single institution—it must remain distributed, and cooperation is essential.

Djorgovski's presentation generated considerable discussion, notably concerning authentication, refereeing, and curation of data and of scientific results. He articulated very effectively how and why knowledge communities could be instrumental, indeed fundamental, in his discipline's ability to grow and develop, and he demonstrated how this premise is likely a universal truth at least across the science disciplines.

Humanities and Social Sciences

But how do the technology-driven changes in the nature and practice of research relate to other areas of scholarship outside the sciences and engineering? Paul Courant addressed this question in his conversation at the forum. He is Provost at the University of Michigan and an economist by training who currently serves as a member of a commission established by the American Council of Learned Societies (ACLS) to investigate cyberinfrastructure needs in the humanities and social sciences (see sidebar on page 5).

Unlike the sciences, which already share many common elements across disciplines, including a long history and a strong tradition of collaboration, the humanities face a more difficult challenge regarding the prospect of developing a common cyberinfrastructure. According to Courant, complete knowledge communities in the humanities and social sciences are more difficult to foster than in the sciences, due, in part, to a core difference between the sciences and the humanities and humanistic social sciences: in the humanities, data are not developed by scholars the way they are in the sciences. Scientists create instruments to record and develop data according to protocols of their own design. For humanists, "data" come out of the human experience. Anything in the cultural record can be "data." For social scientists, the sources and uses of data run the gamut from scientific traditions (e.g., in experimental psychology, demography, and empirical economics) to humanistic traditions (e.g., the "linguistic turn" employed by many anthropologists and sociologists). The majority of social science work falls somewhere in-between.

Information technology can have profound effects on the ways in which we do our work, and, indeed, can make possible projects and inquiries that could not be conducted without the new technologies, but, in the humanities and social sciences, the fundamental research project, or purpose, remains unchanged. Courant emphasized the importance of this latter message as one for forum participants to take back to their home institutions. The value of widespread

collaborative infrastructure is a tougher sell to humanists (than to scientists and quantitative social scientists) in particular because its universal benefits are not as immediately recognizable across the community, and while individuals may accept the overall value of such a foundation, they do not necessarily see a direct benefit to themselves. So, the challenge is to demonstrate that the payoff is high for the humanists through, for example, having ubiquitous access to data already in digital form and the potential of reformatting other data.

To cope with this challenge, the ACLS commissioners have discussed the idea of asking funders to require recipients to follow certain protocols for the dissemination of their research that would promote advancement in making that research accessible in digital form and to do so in ways that support inter-institutional collaboration and collaboration among scholars generally. Courant observed that it will soon be true to say that, for many people, "if scholarship is not online, it does not exist. If scholarship is hidden, it will not be used. A robust cyberinfrastructure will enable people to find quality resources on the Internet. The absence of such an infrastructure will lead much good work and valuable materially to be essentially invisible and forgotten." In a world where everyone uses the Web, that which cannot be found on the Web is in danger of not being found at all.

Additionally, users other than traditional scholars and students, e.g., members of the broader public whom Courant dubbed the "Google users and museum goers," would benefit from the availability of a humanities-oriented cyberinfrastructure that would increase their access to high-quality, authoritative information; this population may well increase demand and represent a significant driving force for development of the cyberinfrastructure. It is also essential to providing political support for the humanities more broadly.

Drawing on his academic roots as an economist and his experience as a provost, Courant spoke of the pressures that influence choices of institutional competition vs. cooperation. The scale of the cyberinfrastructure vision, he concluded, is so great that it can only be achieved through global cooperation. Acknowledging that academic institutions will always compete in many ways, "when we reach the point of asking ourselves the question of going it alone or pursing a goal with partners, the right answer is always to cooperate. The goal of the cyberinfrastructure, of libraries, is to advance human thought. If the academy, if the library, does not press that view, who will? Competition cannot successfully lead us to address such a fundamental issue as access to information." In support of his point that collaboration among institutions and their libraries will do no harm, Courant observed "None of our institutions will lose their 'brand' or be diminished in any way by contributing their scholarly content to a larger 'pie' of federated content."

Among the many questions and comments addressed to Courant, audience members expressed interest in his perspective as a provost with respect to copyright issues and the use of standards in making materials accessible. Regarding copyright, Courant emphasized that in the realm of teaching, an aggressive stance has to be taken. He observed that parents of college students could be engaged politically on this issue, and that librarians should make a business case to the entertainment industry that they can be allies and not threats by, for example, offering the industry services for the long-term preservation of their materials in exchange for their use for educational purposes. With respect to standards, here, he stated, there is a need

for leadership from deans, department chairs, and provosts in promoting their use among faculty. The way he views it, money is wasted if standards that lead to interoperability are not applied when faculty create and post digital scholarship.

The Roles of Federal Funding

Both Atkins and Courant discussed the need to carefully shape research funding to provide incentives and play a major role in forming an environment conducive to the establishment of broad-based, far-reaching cyberinfrastructures, regardless of the field or discipline.

Dr. Sangtae "Sang" Kim, the next speaker, indicated that his organization, the Computer and Information Science and Engineering (CISE) Directorate within the National Science Foundation (NSF), created the Division of Shared Cyberinfrastructure because it recognized the need for long-term, continued support for, and management of, cyberinfrastructures. He discussed what he described as a culture change at NSF, and explained that funding from the new division is now more equally divided between new development and sustenance of operational systems and services. Cyberinfrastructure is not seen as an end in and of itself, but, rather, as being important in that it can enable new research in science and engineering. Further, as with many earlier investments in information technology and computer-communications networking infrastructure pioneered by NSF, there may be enormous leveraged payoffs in terms of the way our society as a whole functions, not just in the scientific enterprise. According to Kim, "a billion dollars invested in cyberinfrastructure may well result in ten trillion dollars in economic growth."

Reactions to the Speakers

The reaction panel provided an opportunity for the participants to take stock of the day's presentations as they were framed by reflections from panelists with a variety of different roles within the higher education enterprise. It also allowed for the discussion of some additional audience questions about advancing cyberinfrastructure adoption and deployment.

Marjory Blumenthal, Associate Provost, Academic, at Georgetown University, put the day's discussions in perspective and reminded participants of important public policy issues raised when federal agencies approach funding for development of networks and supporting infrastructure. For example, the Department of Homeland Security (DHS) is charged to protect "critical infrastructure" for the nation; the way DHS approaches this challenge is very different from the way an agency with a research mission might approach building infrastructure, and this has already raised considerable privacy issues. She also highlighted the rising technology expectations of students and scholars and the challenges faced by institutions in determining how to pay for common technological enhancements without exacerbating communities of digital haves and have-nots. She called for more analysis of both the economics and the policy issues associated with achieving the cyberinfrastructure vision.

Carol Mandel, New York University's Dean of Libraries, identified three perspectives for the path to achieve the transformations called for by the speakers: institutional collaboration, human resources development, and inter-institutional partnerships. Within the institution she

identified the need for libraries and information technology departments to extend current collaborations on digital libraries to develop the set of services that Dan Atkins described as "middleware".

Perhaps one of the most striking, attention-catching moments of the panel presentations came when Mandel observed that scientists are asking libraries to be "collectionless, stateless, egoless…" in this new design. Referencing a recent essay by Wendy Lougee of the University of Minnesota on diffuse libraries taking on a range of new roles as they become more deeply engaged in the creation and dissemination of knowledge (http://www.clir.org/pubs/abstract/pub108abst.html), Mandel pointed out a cycle of science libraries on campus. The cycle began with a small library in the department run by scientists, then run by librarians, to a virtual library, then back to a science-owned "library," and now to calls for librarians to manage the content.

In this context Mandel raised the challenge of developing human resources to achieve the cyberinfrastructure vision. This will require people with both domain and digital expertise and now is the time to start identifying how to develop this talent. As an initial step, she proposed encouraging partnerships between graduate students and subject librarians. Inter-institutional partnerships could be encouraged by requirements in grants such as was done in the recent Library of Congress grants from the National Digital Information Infrastructure and Preservation Program. To encourage a "de-branded, altruistic future" as envisioned throughout the day, she concluded that grant incentives are critical.

Indiana's Dean of Information Technology, Brad Wheeler, began by describing the current environment as one rich with vertical, self-contained collaboration aimed at harnessing disciplinary knowledge. To be successful, the cyberinfrastructure needs to be built horizontally and to scale across disciplines and institutions. Wheeler spoke of the importance and difficulty of building organizational capabilities to collaborate, including stepping on local incentives that defeat inter-institutional collaboration and linking strategic goals to budgets so that progress toward the goals are assessed and stays focused. Finally he pointed out that all speakers confirmed that data curation is essential to the success of science but that, at present, this task is taken on by and within the disciplines, not by libraries, for example, the National Virtual Observatory. Wheeler concluded by saying that the time for leadership from libraries is now or the library will fade just as the family farm has faded.

Jane Bortnick Griffith, Assistant Director for Policy Development, National Library of Medicine, wrapped up the panel by observing that in the world of medicine, especially as digital information resources have expanded, users are looking for connections and integration of content across the range of formats. In her experience, it is not an "either digital or nothing" situation but a desire to have a seamless interface to integrate access to all these resources.

Echoing themes from Sangtae Kim's presentation on the NSF cyberinfrastructure initiatives, she also pointed out that the same theme of recognizing value in the support of a common information infrastructure appears throughout the recently produced "NIH Roadmap" describing the future directions of the National Institutes of Health (http://nihroadmap.nih.gov/).

The closing plenary, led by University of California, San Diego, University Librarian Brian Schottlaender, provided some synthesis of the ideas introduced throughout the forum, as well as opening the door for continued discussion. Schottlaender noted that the profound change in scholarship will have a similar impact on the academy, and he echoed the themes that dominated the day: collaboration and cooperation, within institutions, between them, and those of cooperation even in favor of competition. He and attendees summarized some of the numerous challenges ahead, including copyright, funding, achieving trusted federations to reduce redundancy, content curation and archiving, and personnel issues (e.g., domain expertise), among other things. Despite the difficulties, Schottlaender restated the imperative to advance the vision of innovations made possible by a comprehensive cyberinfrastructure in support of e-research which originated early in the day: "drive it up, drive it down, but drive it!"

The author acknowledges the significant contributions of Jaia Barrett, Joan Lippincott, and Clifford Lynch to early drafts of this report.

Presentations from the forum are available on the Web, http://www.arl.org/forum04/.

ACLS Convenes Hearings on Cyberinfrastructure

The American Council of Learned Societies (ACLS) Commission on Cyberinfrastructure for the Humanities and Social Sciences has convened a series of regional, public, information-gathering sessions to hear from those interested in contributing to the work of the commission prior to issuing a report in early 2005. At these sessions, the commission heard from experts in a variety of fields—from those who are actively engaged in digital scholarship and teaching to leaders in libraries and archives, publishing and distribution, academic administration, information technology, and industry development. The intended audience for the commission's report includes the scholarly community and the societies that represent it, university provosts, federal funding agencies (including but not limited to the National Science Foundation), and private foundations. The commission is chaired by John Unsworth, Dean of the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign.

On October 26, ARL was represented before the ACLS commission by Fred Heath, Vice Provost and Director of General Libraries, University of Texas at Austin, and Past President of ARL. ARL's statement addresses examples of how research libraries are now contributing to advance the emerging cyberinfrastructure and the challenges and barriers that are faced. The statement will be available on the ACLS Web site with other presentations from the regional hearings at http://www.acls.org/cyberinfrastructure/cyber_public_sessions.htm.

Back to top

Goldenberg-Hart, Diane. "Libraries and Changing Research Practices: A Report of the ARL/CNI Forum on E-Research and Cyberinfrastructure." *ARL*, no. 237 (December 2004): 1-5.
http://www.arl.org/newsltr/237/cyberinfra.html.

Table of Contents for Issue 237 | Other Current Issues Articles
Other Networked Information Articles

ARL Bimonthly Report Home ARL Home

Maintained by: ARL Web Administrator Site Design Consultant: Chris Webster **Last Modified: December 13, 2004**