# The Institutional Challenges of
Cyberinfrastructure
and E-Research

By Clifford Lynch

Scholarly practices across an astoundingly wide range of disciplines have become profoundly and irrevocably changed by the application of advanced information technology. This collection of new and emergent scholarly practices was first widely recognized in the science and engineering disciplines. In the late 1990s, the term *e-science* (or occasionally, particularly in Asia, *cyber-science*) began to be used as a shorthand for these new methods and approaches. The United Kingdom launched its formal e-science program in 2001.[1] In the United States, a multi-year inquiry, having its roots in supercomputing support for the portfolio of science and engineering disciplines funded by the National Science Foundation, culminated in the production of the "Atkins Report" in 2003, though there was considerable delay before NSF began to act programmatically on the report.[2] The quantitative social sciences—which are largely part of NSF's funding purview and which have long traditions of data curation and sharing, as well as the use of high-end statistical computation—received more detailed examination in a 2005 NSF report.[3] Key leaders in the humanities and qualitative social sciences recognized that IT-driven innovation in those disciplines was also well advanced, though less uniformly adopted (and indeed sometimes controversial). In fact, the humanities continue to showcase some of the most creative and transformative examples of the use of information technology to create new scholarship.[4] Based on this recognition of the immense disciplinary scope of the impact of information technology, the more inclusive term *e-research* (occasionally, *e-scholarship*) has come into common use, at least in North America and Europe.

Clifford Lynch is Executive Director of the Coalition for Networked Information (CNI).

Illustration by Gemma Robbinson, © 2008
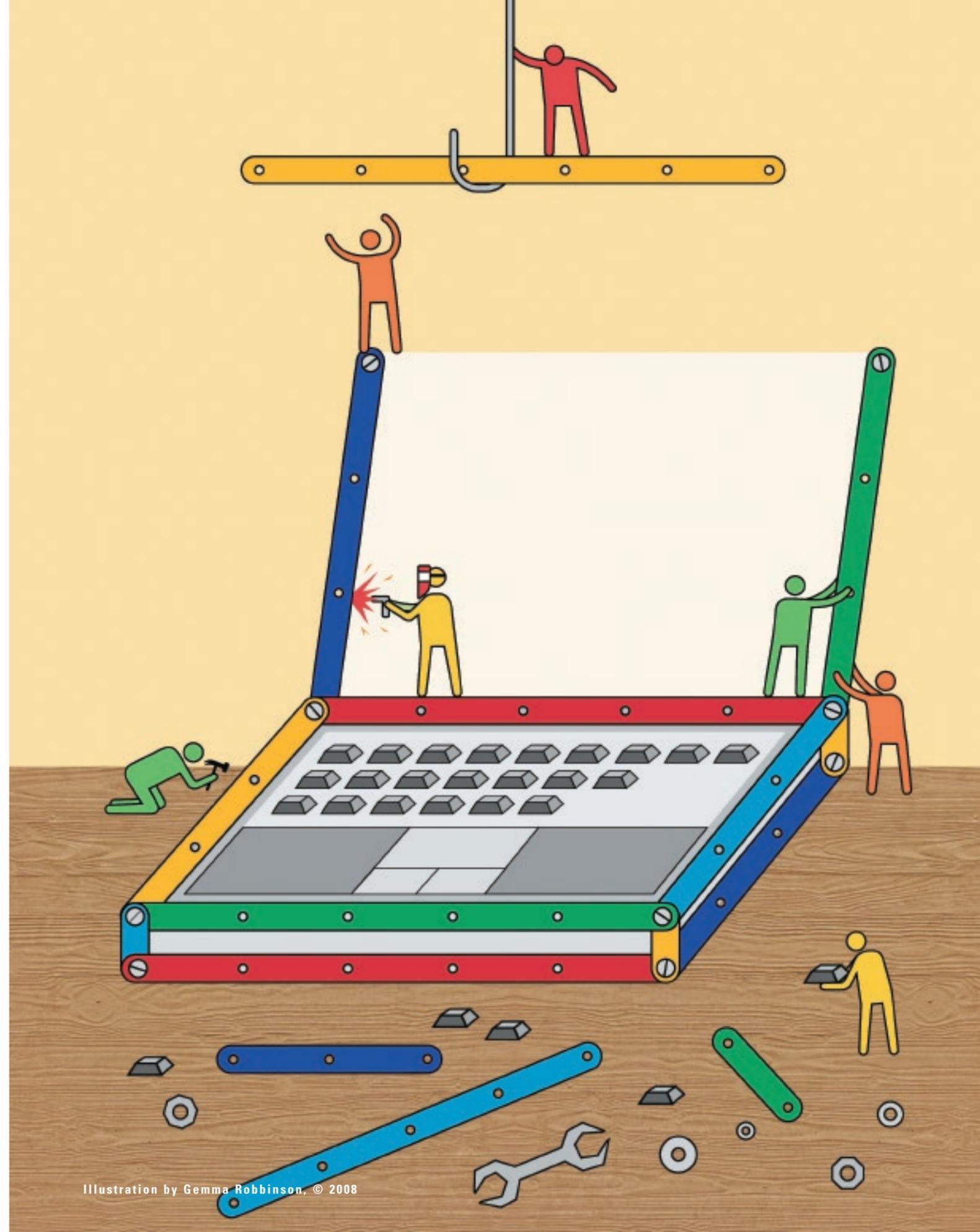
When we speak of the changes wrought by information technology, we consider information technology in its broadest sense: not only high-performance computing and advanced computer communication networks but also sophisticated observational and experimental devices and sensor arrays attached to the network, as well as software-driven technologies such as high-performance data management, data analysis, mining and visualization, collaboration tools and environments, and large-scale simulation and modeling systems. Content, in the form of reusable and often very large datasets and databases—numeric, textual, visual—is an integral part of advanced information technology also. Even the new collaborative social structures enabled by information technology might themselves be considered a part of the technology base in a broad sense.

In thinking about how best to support the changes in scholarly and scientific work and also to accelerate these changes as a way of advancing scientific progress,

digitizing the contents of cultural memory organizations such as libraries, archives, and museums worldwide. There are similar but less ambitious programs to import reference and evidentiary collections into the cyberinfrastructure for the sciences: biodiversity and taxonomic collections, materials from natural history museums, printed collections of historical scientific observations from archives, the historical corpus of scientific literature, and the like.[5]

National and Campus Perspectives on Cyberinfrastructure Deployment

Until recently, much of the articulation of the nature of cyberinfrastructure and the programs to implement it has come from visionaries such as Dan Atkins (who recently completed a two-year term as the Director of the Office of Cyberinfrastructure at the U.S. National Science Foundation) and Tony Hey (who formerly served as the head of the e-Science Programme in the United Kingdom) speaking

tors, very high-end supercomputers) into cyberinfrastructure components that can be shared by researchers around the world. Again, this is only natural, since the national support organizations fund these very expensive resources and are eager to see their value and utility maximized within the scientific community. And of course the funding organizations are very interested in advancing the development and deployment of services and tools that will be helpful to large numbers of investigators spread across many different institutions and disciplines. Of particular interest are systems that facilitate the sharing and reuse of data or that allow more active collaboration among geographically scattered scientists.

Characterizing the shape of the national strategies for the development and deployment of humanities cyberinfrastructure is more difficult. Humanities research, at least in the United States, is much less dependent on centralized funding from a few government agencies such

How does the campus cyberinfrastructure challenge differ from the national cyberinfrastructure challenge?

science funding agencies began speaking about the need to systematically invest in what they called *cyberinfrastructure*. This included not just the information technologies already mentioned but additionally the human and organizational resources needed to facilitate services and activities such as the training and retraining of scholars, the management and operation of the technical facilities that make up the IT environment and the scholarly tools that have been integrated with it, and the performance of data curation and preservation. As humanists subsequently explored how to adapt the idea of cyberinfrastructure to their own disciplinary needs, they also articulated the need for as much of the human record—expressed in text, images, sound and video recordings, and digital surrogates of cultural artifacts—as possible to be available in digital form, along with tools to facilitate the study and analysis of this corpus, thus implying a very large and open-ended program of

in the context of national-level science programs. Naturally, they have tended to focus on the need for cyberinfrastructure to support large-scale national and international scientific projects and programs.[6] Indeed, one characteristic of many of these large projects is that they are cross-institutional and have sufficient scale to include expertise on relevant information technology and data and information management as an organic part of the project team, rather than simply functioning as a client of some campus-based service. In many cases, these large projects have also been assisted by national-scale support organizations (such as Internet2), which have helped with intercampus technology coordination.

The national and international cyberinfrastructure implementation planning has also tended to focus on making unique or near-unique scientific resources (e.g., databases, telescopes, electron microscopes, undersea sensor arrays, particle accelera-

as the National Science Foundation or the National Institutes of Health. Funders such as the Andrew W. Mellon Foundation or the National Endowment for the Humanities have thus far financed mostly isolated exploratory projects, and the resources available for humanities cyberinfrastructure support are limited. In the United States, the Institute of Museum and Library Studies (IMLS) has funded some substantial digitization programs, and in the United Kingdom, the Joint Information Systems Committee (JISC), along with other government organizations, has made some substantial investments in digitization of key scholarly resources. However, with the growing interest in more systematic infrastructure building, the landscape here is likely to change substantially in the next few years. Project Bamboo (http://projectbamboo.org), supported by the Mellon Foundation and currently in the planning-grant phase, is particularly interesting, since it seems focused on building

campus cyberinfrastructure capability—including organizational support programs for scholars—for the humanities at participating institutions rather than on building pieces of *national* cyberinfrastructure resources for humanists everywhere or on simply funding exemplar projects by small groups of humanists.

How does the campus cyberinfrastructure challenge differ from the national cyberinfrastructure challenge, recognizing that investments in these areas should be not just complementary but mutually reinforcing? First, there is a strong obligation and mandate for a base level of universal service across the campus: all scholars, in all disciplines—whether working individually or in small or large collaborative teams—need to be able to apply information technology in their research and to access and build on cyberinfrastructure services that include data management and data curation (which are probably the services in broadest demand); campus scholars also need to be able to get help in determining and learning *how* to do this (simply providing access to services isn't enough). This includes scholars who typically do not receive grants and whose disciplines get little or no national funding support and who cannot pay for or compete for resource allocations at the national level. And it particularly includes individual scholars or groups of scholars who cannot afford dedicated specialist IT collaborators as part of a research team.

Second, the campus perspective is concerned with the "average" rather than the "extreme" scholar, in terms of demands for cyberinfrastructure-based resources. Many researchers can do what they need to do by employing primarily local IT services and resources rather than national-level ones, and may need to consult or contribute to national or international shared-data resources at levels of intensity easily accommodated by basic campus-provided network connectivity. Thus there is a need to plan for the design, development, deployment, and support of cyberinfrastructure components that are intended primarily to support common local needs, including the needs of the campus community to reach and work with popular, widely used national and international cyberinfrastructure components and services. One of the key challenges—politically, financially, and technically—is defining the demarcation between free universal service and the more specialized package of support services offered to extreme users, a package that may be predicated on such users' ability to obtain funds or other resource allocations.

Finally, national-level cyberinfrastructure has been almost exclusively focused on supporting research and advanced graduate education closely associated with research. But cyberinfrastructure can also be adapted for and placed in the service of teaching and learning more broadly.

The opportunities here have received attention through a recent NSF report on what is now being called *cyberlearning*.[7] The nature of and the balance between national funding agencies and campus initiatives in pursuit of these goals have yet to be shaped but will be of great importance going forward; it seems likely that as with research cyberinfrastructure, a good deal of local investment will be called for if a campus is to be in a position to fully benefit from national investments.

Cyberinfrastructure Components from the Campus Perspective

Computational Resources and Data Storage
Traditionally, funding agencies have followed two parallel tracks in providing computational resources to support research projects. At the very high end, they have established and financed national-level computational centers (supercomputer centers) and have used a competitive peer-review process to allocate the majority of the time at these centers. For

other projects, investigators have used grant funds to purchase computers that are housed and managed locally (historically at a lab or departmental level). But a number of developments are causing a reshaping of these traditional practices.

As computing cycles have become less expensive and more plentiful, a larger number of researchers can get their computing needs filled on the campus level rather than going through the complexities of applying for time at a national-level center. Computing capacity is usually expanded by the addition of more commodity computers into dense parallel clusters; these are environmentally demanding and rapidly exceed the power and cooling that is available outside of carefully designed, centralized data centers. Cluster system administration, particularly in
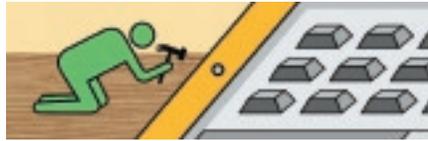
today's challenging security environment, is becoming a more complex and more professionalized activity best handled centrally. Growing in popularity are arrangements in which investigators can invest in campus-level shared computing clusters that are professionally managed (including redundant, properly backed-up data storage); often, the campus will contribute funds to help underwrite this centralized resource and may also offer programs to provide at least some access to computing cycles for those faculty who do not have grant funds to underwrite a contribution to the shared pool. The shared computational resources are attractive not only because researchers are relieved of the complexities of managing the systems but also because the usage level for available cycles tends to be substantially higher when they are pooled.

The primary campus policy challenges here are in providing the right incentives for researchers to contribute to a shared resource and in managing the allocation of

> The demands for data management and curation to facilitate data sharing and reuse form one of the fundamentally new aspects of e-research.

the shared resource among the contributors (and the broader campus community). Practically speaking, simply staying ahead of demand in making appropriate conditioned space available, installing machines, and administering the clusters is a significant challenge for the campus IT organization.

Researchers with computational demands appropriate for national centers need to make smooth transitions back and forth between local and national computing resources or even to use these resources in combination. Many of the problems here are technical—software compatibility and optimization, sharing of files—but to the extent that they rely on identity management and access management, they also have policy components. Though the EDUCAUSE Campus Cyberinfrastructure (CCI) Working Group Task Force (http://www.educause.edu/Campus Cyberinfrastructure%28CCI%29Working Group/10288) is a good start, we currently do not have sufficiently robust organizational structures either for supporting individual campus researchers as they move back and forth across the interface today or for allowing campus IT teams and national supercomputer centers to collaboratively design tomorrow's architectures, tools, and services to facilitate transparent and seamless local-national computational transitions.

Basic data-storage services (simply dealing with storing and fetching bits, as opposed to planning for long-term content-oriented curation and preservation, which will be discussed later) form an interesting complement to computational resources. Historically, there has never been much development of very-large-scale network-based national data-storage services, except when these resources are tightly coupled to the development of similar computational resources. At least some supercomputing applications needed complementary high-capacity, high-performance data storage, but very-high-capacity data storage in the absence of supercomputing wasn't part of the national cyberinfrastructure. Projects

that needed such storage (but not a lot of supercomputing) solved their problems locally, usually with various kinds of portable media. Part of the issue was technical: performance gaps between local storage attached via high-performance channels or busses and Internet-accessible storage servers were large and intractable. Part was a lack of demand: until fairly recently, few thought of cyberinfrastructure as incorporating support for massive long-term data archiving and preservation.

The thinking about storage has changed rapidly. Today, scholars want to be able to store bits on network-accessible servers and to know that these bits are automatically and frequently backed up to one or more geographically disparate sites; they want confidence that their research will survive a hurricane, earthquake, or other disaster. They want storage as a highly reliable service, with the service operator checking the backups, arranging crash- and disaster-recovery mechanisms, and handling the periodic necessary migrations from soon-to-be obsolete older hardware to newer technology. Campus IT organizations are in an excellent position to provide such services, which benefit from economies of scale, and in particular to arrange the validation, auditing, and geographic-dispersion aspects of a modern storage service. As with computational resources, there are real problems getting campus storage services to integrate smoothly with computations hosted at the national centers, and it is still quite common to find data being awkwardly copied back and forth between storage facilities attached to these national resources and local campus facilities, as the locus of use of the data shifts.

Very interesting campus policy issues surround the deployment of reliable storage services, beyond the obvious questions of how best to fund the service and how large to set the base universal service entitlement for the campus community. Should there be a reliable and a (presumably cheaper) unreliable storage service? Should there be multiple levels of reliability offered? The archetypal nightmare is

the researcher, short on funds, who argues that he'll just buy a terabyte hard disk for $400 rather than pay $500/year for reliable storage services—and who then later loses his research data to a disk crash and blames the IT organization for incompetence. How close can economies of scale move the costs of reliable storage service to the costs of individual purchases of (unreliable) hardware? Should researchers be forced to use reliable storage services and, if so, when and how? (One can imagine, for example, audits connected to data-management plans in federal grants as a means of identifying investigators who are putting data inappropriately at risk.) With data-storage services, campus cyberinfrastructure design and deployment begins to interconnect with fundamental campus policies and culture about the stewardship responsibilities of scholars, about contracts and grants compliance issues, and about risk management.

Data Management and Curation
The demands for data management and curation to facilitate data sharing and reuse form one of the fundamentally new aspects of e-research.[8] As both researchers and funding agencies recognize that their data is often going to be of lasting value (perhaps in a wide range of different research contexts), and as they also recognize that much of the outcome of a specific research program may well be documented in datasets and databases (plus, perhaps, accompanying software) rather than in traditional journal articles that simply make reference to the underlying data,[9] there is a growing demand for services ensuring that the data is properly documented, that it is correctly placed into a well-known and well-defined format (using the standards of appropriate scholarly communities when available and applicable), and that it is preserved over suitable periods of time by the use of redundant managed storage and, when necessary, format migrations and, most important, by some organization taking responsibility for the data—technically, legally, and financially—and doing what's

necessary to look after it. Researchers often need help at the beginning of a research project in order to ensure that the data coming out of the project is manageable, rather than simply facing an (at best) costly and time-consuming and (at worst) intractable mess at the end of the project when the data produced by the research needs to be handed off to the preservation

The system of data curation in various disciplines is still developing. In a few areas, disciplinary-based national or international data repositories are funded directly by research funding organizations or other mechanisms. This is the case in crystallography, for example. In other areas, funders manage the repositories directly (e.g., the National Center for Biotechnology

research can generate very long-lived and substantial financial responsibilities for the institution. In cases of grant-funded research, these responsibilities may in fact be legal obligations that are spelled out in the terms and conditions associated with the grant. Even lacking such legal requirements, colleges and universities have a clear policy and ethical obligation to participate in the stewardship of faculty research. However, the extent of this obligation is controversial, especially among institutional administrators who are fearful of being assigned large, new, unfunded stewardship mandates by research funders.[10]

> One of the early visions of the opportunities that advanced networking could offer for transforming scholarly work was the idea of a "co-laboratory."

service. Unfortunately, many researchers still don't realize that beginning curation planning early in the project may save a great deal of expense and pain later. And they may be unfamiliar with how to do such curation planning and may need specialists' assistance across the entire spectrum—from basic data management to more disciplinary-oriented curation dealing with the semantics of the data.

Information at the National Library of Medicine). For many other disciplines today, there is no one to take responsibility except the researcher's campus.

One of the central policy problems with data curation and preservation is that the costs of these activities persist long after the project has ended and the research grant has been spent (if there was a research grant in the first place). Faculty

A second key policy issue is that the most effective curation of many kinds of data requires substantial disciplinary expertise. This may be feasible if responsibility for the data stewardship is successfully aggregated at a national or international level, where there is enough scale to afford the expertise and to build the specialized systems for managing dif-

ferent kinds of disciplinary-specific data. And even in this case, some disciplines may just be inadequately funded. But at the campus level, no institution will have the resources to provide specialized support and expertise for every discipline that is represented on campus, that is producing data to be curated, and that isn't being taken care of by an external disciplinary program. Thus, in the absence of centralized repositories established by funding agencies or disciplinary communities, campuses will need to work together to develop arrangements for pooling and sharing disciplinary expertise.

One of the essential campus challenges here is to create a support organization that can reach out to all scholars on campus early in the data lifecycle with assistance in planning for data management and curation/preservation strategies; this service will need to involve information technologists, librarians and archivists, and disciplinary experts, as well as maintain close relationships with the Office of Contracts and Grants and the Chief Research Officer, among others. In addition, the campus must be prepared to take on institutional responsibility for long-term curation of data at the appropriate point in the lifecycle and must develop organizational capabilities to do this (most likely led by the campus library). The full range of resource-allocation questions immediately appears: How are priorities assigned, and where can funding be found to help support the program? How does the institution decide what should be kept and for how long? What contractual obligations to funding agencies must the campus honor? Are there other ethical and policy obligations that need to be considered as constraints?

Again here we find rich connections to other institutional policy issues. For example, while norms about data sharing are likely to be driven by disciplinary norms and by sharing requirements imposed by funders, the college or university may want to establish policies in this area and will certainly become involved in ensuring policy compliance by faculty. Particularly (and traditionally) in medical and social science disciplines but increasingly in new areas like history, issues concerning human subjects and privacy create major barriers to data collection, retention, and reuse. Most of these constraints originate from local institutional review boards (IRBs), and the situation varies wildly from one campus to the next. Although details are beyond the scope of this article, this is certainly an area that would benefit from both local and national policy review.

Collaboration Environments
and Virtual Organizations
One of the early visions of the opportunities that advanced networking could offer for transforming scholarly work was the idea of a "co-laboratory"—a virtual space where scholars could come together to control experiments, share access to observational instruments, analyze data, and write.[11] More recently, the National Science Foundation, in particular, has been emphasizing the related but broader—and, I think, much clearer—idea of "virtual organizations"—co-laboratories that can be set up, can persist for as long as they are needed, and

then can be broken down when no longer necessary.[12] They might be quite short-lived, to address a specific collaborative activity, or very long-lived, perhaps for the lifetime of a piece of experimental apparatus. They cross organizational and national boundaries and potentially different sectors (industry, academia, government). Virtual organizations can control various kinds of assets (e.g., observational apparatus) and can also produce new assets such as datasets or publications. They may also need specialized support, such as dedicated network capacity between specific participants.

The software to provide various kinds of collaborative environments is advancing rapidly, with developments coming from areas as diverse as learning management systems (e.g., Sakai), virtual research environments, teleconferencing/telepresence, and disciplinary clearinghouse development tools such as HUBzero (http://www.hubzero.org)—the software that underpins the Purdue University
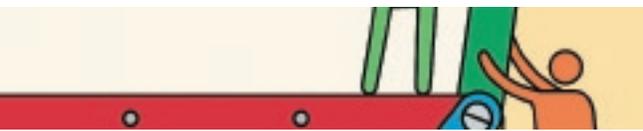
nanoHUB, among other science gateways. These tools are commonly used to collaborate within institutions as well as to support cross-institutional virtual organizations.

The policy challenges that virtual organizations raise for campuses are extensive. Who is responsible for supplying the infrastructure that allows a given virtual organization to operate? Where are the machines that "run" the virtual organization, for example? How are resources allocated to it? What institution should take stewardship responsibility for the outputs of a virtual organization (assuming that no natural disciplinary repository exists)?

Campuses have invested extensively in local infrastructure such as identity management and authentication/

authorization systems. These now often allow the very secure and flexible protection and sharing of local resources within the context of the local campus community. Unfortunately, as the discussion of virtual organizations underscores, the context of the local campus community is no longer even close to sufficient. At least for collaborations among higher education institutions, developments such as Shibboleth can provide the technical basis for talking about interinstitutional resource sharing, but this sharing requires participants in a virtual organization to trust *all* the parent organizations of the participants. Further, for effective collaboration and resource sharing, the parent organizations will need to become comfortable with interinstitutional trust at an organizational level (with everything that this implies in areas such as business processes and regulatory compliance). All of this becomes much more difficult, of course, as collaborations expand—as they often do—to include people from

institutions in other nations, people from industry, and even unaffiliated people (e.g., independent scholars, amateur scientists).

Organizational and Support Implications

Probably the greatest challenge of cyberinfrastructure at the campus level will be the design and staffing of the organizations that will work with the faculty: helping faculty access cyberinfrastructure services locally (and, when necessary, globally); assisting faculty in managing their data—including observational data, the construction of research and reference collections, or data from analysis or simulation—and preparing this data for handoff to the appropriate data repositories and curators at the appropriate time; and aiding faculty in parallelizing computations or organizing data for reuse, mining, and mashups. Staff will be needed to assist in the setup of virtual organizations and also to help with their breakdown. Existing academic computing organizations and libraries will no doubt provide expertise, but this will need to be supplemented with more expertise in disciplinary data, standards, and tools and perhaps also with more capability for consulting on software, data, and information design. Delivering this support to faculty and students across campus and across disciplines—at wildly varying levels of experience, expertise, and sophistication and, most of all, at scale—will be essential. One other organizational choice, which will doubtless be made very differently from campus to campus, will be the degree to which these services are centralized and the degree to which they are modeled on and extended from today's department- and school-level computing support organizations (and even departmental or other specialized libraries). Given these requirements for scale, one final set of questions concerns staff: Where will the necessary staff come from? How will they be trained? What educational qualifications and background will they have, and what academic programs will produce them?

Beyond the human and organizational challenges is the challenge of defining, implementing, and financing a portfolio of fundamental cyberinfrastructure services, available to the entire campus community, to ensure that all scholars, in all disciplines, can make use of information technology and networked information resources in their research and that the results of their research will be available to other scholars—today, tomorrow, and in the distant future. This is the first need, from which follows the second: facilitating the simple and transparent use of components of the national and international scholarly cyberinfrastructure beyond the campus boundaries by scholars who need to use such resources in conjunction with the campus offerings. *e*

Notes

1. Key U.K. documents can be found at <http://www.rcuk.ac.uk/escience/>.
2. Daniel E. Atkins served as chair of the panel that produced the report *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure,* January 2003, <http://www.nsf.gov/od/oci/reports/atkins.pdf>. There are also a substantial number of workshops and reports exploring e-science and cyberinfrastructure developments and opportunities at the level of specific disciplines; these can be found at <http://www.nsf.gov/od/oci/reports.jsp>. I do not know of any document similar to the Atkins Report covering e-science within the disciplinary portfolio funded by the National Institutes of Health, though NIH and its National Library of Medicine have been pioneers in many areas of e-research and cyberinfrastructure development, and certainly e-science is pervasive in work such as the "NIH Roadmap for Medical Research," <http://nihroadmap.nih.gov>. Similarly, other federal science funding agencies—for example, the Department of Energy—are now making significant cyberinfrastructure investments.
3. Francine Berman and Henry Brady, *Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences,* May 12, 2005, <http://vis.sdsc.edu/sbe/reports/SBE-CISE-FINAL.pdf>.
4. See the ACLS Commission on Cyberinfrastructure for the Humanities and Social Sciences, *Our Cultural Commonwealth* (2006), <http://www.acls.org/cyberinfrastructure/OurCulturalCommonwealth.pdf>, and Berman and Brady, *Final Report.*
5. For more on cyberinfrastructure as it relates today to research and education in science and engineering and to teaching and learning in the liberal arts, as well as on how educators view the importance of cyberinfrastructure technologies to research and to teaching and learning now and in the near future, see the articles in the July/August 2008 issue (vol. 43, no. 4) of *EDUCAUSE Review,* a theme issue on cyberinfrastructure: <http://connect.educause.edu/Library/EDUCAUSE+Review/EDUCAUSEReviewMagazineVol/46965>.
6. See National Science Foundation, Cyberinfrastructure Council, *Cyberinfrastructure Vision for 21st Century Discovery* (Washington, D.C.: National Science Foundation, March 2007), <http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf>.
7. NSF Task Force on Cyberlearning, Christine L. Borgman et al., *Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge,* June 24, 2008, <http://www.nsf.gov/pubs/2008/nsf08204/nsf08204.pdf>. An excerpt from the Executive Summary for the report by the NSF Task Force on Cyberlearning is reprinted in this issue of *EDUCAUSE Review.*
8. On issues surrounding data curation, see National Science Board, *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century,* September 2005, <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>, and *To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering,* a Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships, September 26–27, 2006, <http://www.arl.org/bm~doc/digdatarpt.pdf>. See also Tony Hey and Anne Trefethen, "The Data Deluge: An e-Science Perspective," in Fran Berman, Geoffrey Fox, and Tony Hey, eds., *Grid Computing: Making the Global Infrastructure a Reality* (New York: J. Wiley, 2003), <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/research/esci/datadeluge.pdf>, and Christine L. Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* (Cambridge: MIT Press, 2007).
9. See "The Coming Revolution in Scholarly Communications and Cyberinfrastructure," a special issue of *CTWatch Quarterly,* vol. 3, no. 3 (August 2007), <http://www.ctwatch.org/quarterly/archives/august-2007>, especially the article by Clifford Lynch, "The Shape of the Scientific Article in the Developing Cyberinfrastructure."
10. On college and university responsibilities for stewardship, see Clifford A. Lynch, "A Matter of Mission: Information Technology and the Future of Higher Education," in Richard N. Katz, ed., *The Tower and the Cloud* (Boulder, Colo.: EDUCAUSE, 2008).
11. Committee on a National Collaboratory Establishing the User-Developer Partnership, *National Collaboratories: Applying Information Technology for Scientific Research* (Washington, D.C.: National Academy Press, 1993). See also Richard T. Kouzes, James D. Myers, and William A. Wulf, "Collaboratories: Doing Science on the Internet," *Computer,* vol. 29, no. 8 (August 1996), pp. 40–46.
12. National Science Foundation, *Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations,* Final Report from Workshops on Building Effective Virtual Organizations, May 2008, <http://www.ci.uchicago.edu/events/VirtOrg2008/>.