Introduction

"Do more with less."

That is both the challenge and opportunity facing scientists in an era of shrinking budgets, rapidly advancing technologies, escalating environmental concerns—and very high bandwidth. As investigators and educators alike look for new strategies to leverage available resources and bridge miles and time zones, it has also become a guiding principle behind the Chemistry Research and Instrumentation and Facilities programs within the National Science Foundation's Division of Chemistry.

Increasingly, these conditions are leading to cyber-enabled technology: allowing users remote access to instrumentation and the ability to collaborate and manage data from afar, effectively expanding opportunities to pursue state-of-the-art research using more expensive and more sophisticated instruments than many faculty, students, or researchers could hope to acquire in their home institutions. Whether in the classroom or the laboratory, for teaching or research, the promise of cyberinfrastructure lies in its potential to link users with the instruments, data, and collaborators they need, circumventing distance, time, and to some extent, conventional funding constraints. To the extent that maintaining America's national competitiveness depends on connecting professional-level students in science and engineering with the most sophisticated research tools possible in the largest, most supportive environments possible, cyber-enabled instrumentation offers an unparalleled means of providing the access they need and new opportunities for collaboration.

Building on a series of prior panels, reports, and workshops that explored the potential of cyber-enabled instrumentation, some 37 chemists, IT professionals, and developers gathered in Arlington, VA, on July 16–18, 2008, for the NSF-sponsored "Cyber-Enabled Instrumentation Strategic Planning Workshop." Their charge was to assess the current state of cyber-enabled instrumentation in science (with an emphasis on chemistry), including challenges and issues in remote instrumentation; identify best practices and future directions for instrument cyber-enabling, including how to support and prioritize investments in cyber-enabled instruments that best suit the broader scientific community; and generate a series of recommendations, next steps, and future actions. An added benefit was the opportunity to create a network of providers and users of cyber-enabled instruments to propagate innovations and sustain further development.

The desired outcomes included:

- A common understanding of the current state of cyber-enabled instrumentation programs
- Operational definitions for specific terms and phrases related to cyber-enabled instrumentation
- Guiding principles for cyber-enabling instrumentation best practices
- Familiarity with how selected cyber-enabled instruments are used in research and instructional settings/environments
- A resource document for cyber-enabled instrumentation in chemistry
- A cyber-enabling instrumentation network

- Priorities for future instrument cyber-enabling that best suit the broader chemistry community
- An action plan.

Background

Cyberinfrastructure has been defined by the National Science Foundation Office of Cyberinfrastructure as "The coordinated aggregate of software, hardware, and other technologies, as well as the human expertise required to support current and future discoveries." Because cyber-enabling technology has the potential to overcome budget limitations by providing researchers, faculty, and students from multiple sites with remote access to a full range of instruments, it offers a number of advantages over conventional physical access to laboratory instrumentation, including:

- Facilitating collaboration, consultation, and distributed expertise. An institution may not have the resources to hire someone with expertise in a particular area, for example, but using remotely enabled instrumentation and technologies, its faculty can collaborate with experts at other institutions and draw on their knowledge.
- Providing new opportunities for education and training.
- Pedagogical benefits that include the ability to illustrate and validate important concepts, whether in research or while training students. Remote access can introduce students to professional practice in the use of an instrument as well as to collaborative skills.

To open the discussion, Katherine Kantardjieff, Ph.D., professor of chemistry and biochemistry at California State University, Fullerton, presented an overview of cyber-enabled instrumentation—its evolution, its current state, and where it appears to be heading.

Precursors/prior reports

Prior NSF workshops have produced key documents that collectively trace the evolution of cyber-enabled instrumentation. The Atkins report, published in 2003, identified complex opportunities for creating new research environments based on cyber-instrumentation. It also described pitfalls associated with this new concept, including underfunding, both in terms of dollar amounts and duration of awards; lack of understanding of some of the technological features; lack of interoperability between disciplines; and lack of appropriate organizational structures.

Cyber Chemistry Workshop 2004 identified research and education frontiers that would be brought within reach by investments in cyber-enabled instrumentation. The workshop and its report focused more on computing infrastructure for modeling and simulation in chemistry but made statements about remote instruments, noting that 1) enhanced access to remote instruments would benefit the chemistry community and 2) remote access to high-end, state-of-the-art instruments would maximize their scientific impact, serving broader audiences and allowing more widespread use of current-generation technologies.

The Cyberinfrastructure Vision for 21st Century Discovery, published in 2007, recommended the use of cyberinfrastructure-mitigated tools for collaboration in diverse, commonplace environments, including research and the learning environment. It also noted the growing importance of virtual organizations. Among the cyber tools and services it highlighted were laboratory automation, including sample and data management, and remote collaboration tools, such as Web conferencing tools, remote sensors, and remote access/control of instruments. It recommended the development and support of new kinds of learning and research cultures across disciplines, as well as distributed knowledge communities, which transcend campus, departmental, and geopolitical boundaries.

Drivers for cyber-enabling instruments

Remotely accessible instrumentation saves money. One of the primary drivers behind the move toward cyber-instrumentation is financial. Increasingly, tight budgets are bumping up against the high cost of the basic tools necessary to carry out cutting-edge research. Furthermore, not all institutions can afford to hire faculty with expertise in specific areas. As a consequence, sharing facilities and instrumentation through cyber-enabling has the potential to expand the high-end resources available to researchers regardless of institution and also to facilitate collaborative groups whose members are able to share their expertise. CERN's Large Hadron (particle) Collider, for example, a shared facility with unmatched capabilities, is the focal point for a huge collaborative group worldwide. Similarly, two of the few ultra-high-voltage electron microscopes in Osaka, Japan, and Seoul, Korea, rely heavily on cyberinfrastructure for their interconnectivity and collaboration among all the scientists who use them and share their expertise.

There is an increased need for computational capacity. Another driver is the fact that a growing number of science and engineering fields, such as structural genomics and earth science, are becoming very information- and computation-intensive. Scientists increasingly are using computing clusters and grids for modeling, simulation and computation, and, particularly relevant to chemistry, shared and distributed systems of advanced instrumentation.

Technology gets cheaper. A third driver is the progress of information technology generally and the fact that it is becoming less expensive. That cost reduction, in turn, makes it easier and more affordable to share research data, tools, and computing power. At the same time, this affects requirements for data storage, networking, and the capacity and capability of computational hardware.

The evolution of cyber-enabled instrumentation

Thanks to these drivers, the notion of cyber-enabling instrumentation has steadily been gaining momentum, becoming an important option for research and instruction. One of the first demonstrations of remotely enabled instruments took place at the National Center for Microscope Imaging Research in 1992, when researchers demonstrated the first system to

control an electron microscope over the Internet and allowed users in Chicago to remotely manipulate the instrument and refine their data. In 1993, the Computer Science and Technology Board of the National Research Council noted in its report "National Collaboratories: Applying Information Technology for Scientific Research" that computing and communications were becoming essential tools of science, making possible new kinds and degrees of collaboration. The CSTB addressed technical, scientific, and social aspects of fostering scientific collaboration using information technology, exploring issues in molecular biology, oceanography, and space physics, and recommending a partnership between scientists and technologists to develop better collaboration technology in support of science. In 1994, the Mercury Project was published, which demonstrated the feasibility of Internet robots. Crystallography was one of the first scientific areas to exploit remote access to instrumentation and to benefit from the collaborations it made possible; most Department of Energy labs, for example, went online in the late 1990s.

In an educational context, engineering has been a pioneer in the pedagogy of cyber-enabled instrumentation. In 1996, distance-learning activities were incorporated with remote control of instrumentation in undergraduate engineering labs at Oregon State University. Studies looked to see how and what students learned, depending on whether they were proximal to an instrument or remotely connected to it. Remote access modes have been shown to enhance students' ability to identify irregularities in experimental results, and students were more likely to demonstrate an understanding of their consequences. Students engaged in simulation modes, however, displayed a shallow understanding of the real context. Purely remote implementation appeared to emphasize hardware objectives in students' minds, while simulations emphasized theoretical objectives.

Although President Bill Clinton cited the Internet as a valuable teaching and learning resource in late 1990s, the use of remote scientific instrumentation projects for K-12 education, while growing, is not yet widespread. Two of the reasons are most likely cost and the need to conform to state science standards. While precollege teachers may be interested in utilizing cyber-enabled instruments in their classrooms, without materials that help them teach to state standards, which dominate curriculum, instructors probably won't take the time to adopt anything that is offered.

Examples of existing cyber-enabled consortia

With the progression from command line–based remote access to browser-based GUIs and commercial off-the-shelf applications, consortia have formed around various kinds of cyber-enabled scientific instruments. Many of these have emerged within the medical community and have made advances such as remote surgeries and telemedicine possible, at the same time contributing to the ability to deliver remote instrumentation in other scientific settings. Examples from academia and government agencies include:

- **The W. M. Keck Foundation Center for Molecular Structure (CMolS**), at California State University, Fullerton. Conceived in 1992 and established in 1994, CMolS is a core facility serving the 23 campuses of the California State University (CSU), the largest public university system in the United States, as part of the California State University Program for Education and Research in Biotechnology (CSUPERB). CMolS, the first

facility of its kind at a predominantly undergraduate institution (PUI), is a comprehensive X-ray diffraction facility and computational laboratory that is dedicated to molecular structure determination and analysis using single-crystal x-ray diffraction methods and computational modeling. Before full implementation of U.S. Department of Energy collaboratory projects, CMolS pioneered the implementation of remote instrumentation access at PUIs, putting instruments online to the CSU in 1997 with a commercial off-the-shelf product called pcAnywhere, operating under DOS. CMolS has since taught joint crystallography courses with several PUIs, most recently in 2007, when students at the other campuses learned to remotely collect x-ray diffraction data and determine structures on CMolS instruments. CMolS now also facilitates PUI faculty and student remote access to beamlines at major synchrotrons, including the Stanford Synchrotron Radiation Lightsource.

- **Science Teaching and Research Brings Undergraduate Research Strengths through Technology Cyberinstrumentation Consortium (STaRBURSTT-CIC)** grew organically out of preexisting regional instrumentation consortia and was formally set up in 2005. With five core nodes (including the labs at California State University, Fullerton, and at Youngstown State University) and about 150 participating campuses, it seeks to transform the research and educational cultures at predominantly undergraduate institutions. Participants include PUIs, historically black colleges, Hispanic-serving institutions and tribal colleges, and the consortium also collaborates with affiliate members such as Ph.D.-granting institutions, government labs, nonprofit organizations, and industry. Its goals are directed toward systematically and significantly changing the research and educational cultures at PUIs. While initially focused mainly on crystallographic and diffraction equipment, the consortium now seeks to extend the pool of scientific instrumentation it makes available to its members into other scientific fields such as nuclear magnetic resonance, electron microscopy, etc. One such new facility now available to STaRBURSTT members is Youngstown State University's new Analytical Materials Instrumentation Facility. At its opening in 2008, NSF's director, Arden Bement, noted, "With partnerships seeded by federal funding, a region can quickly build competitive research capacity that in turn sparks new companies, new jobs, and a more robust economy."

  During their work since 2005, STaRBURSTT members realized that the same challenges that were limiting the broader utilization of cybertechnology for scientific equipment in an academic setting were also relevant in other environments, such as manufacturing; maintenance, repair, and overhaul; and security and defense. This, in turn, led to a broader understanding of the commercial applications of their work and hence to the establishment of the **CyberLabNet project**, an endeavor funded, among others, by the Department of Defense and commercial partners such as Bruker Instruments. The project focuses on developing and optimizing cybertechnology tools that control scientific equipment in order to pilot the same or closely related tools for other applications. The CyberLabNet software system is being developed using the proprietary technology of cumulus::Archive, engineered by the collaborating software company Zethus Software. Their cloud computing platform will allow for a loose connection of distributed devices

that are self-organized into a whole that can deliver unified services, which is the underlying basis for CyberLabNet.

- **Partnership for Remote Instruments to Study the Structure of Matter (PRISSM),** a collaborative of several predominantly undergraduate institutions, community colleges, and high schools in California that includes California State University campuses, Harvey Mudd College, two community colleges, and 13 high schools. Remotely enabled instruments include an atomic force microscope, scanning electron microscope, confocal microscope, two 400 MHz nuclear magnetic resonance spectrometers, an X-band electron paramagnetic resonance spectrometer, several x-ray diffractometers, ICP-MS, and MALDI-TOF MS. The partnership provides the high schools with computers and projectors to access the instruments remotely in real time and interact with scientists in synchronous mode. Members of PRISSM also design, deploy, and assess learning units and simulations that address California State University Programmatic Student Learning Outcomes and the California Science Content Standards with regard to various aspects of molecular structure. PRISSM has been established and supported by California State University, Fullerton; the California State University Program for Education and Research in Biotechnology (CSUPERB); iLinc Communications; NSF; the W. M. Keck Foundation; and The Boeing Company.

- The **Delaware Oceanographic and Environmental Research Remote Instrument (DOERRI)**, an autonomous underwater vehicle being used for estuary research and coastal observatory development. One of the research groups using it is headed by Robert Ballard of the University of Rhode Island's Institute for Exploration and Institute for Archaeological Oceanography. Ballard is using it for geological and archaeological research in the Aegean and Black seas, analyzing old shipping routes looking for shipwrecks and other artifacts. The instrument includes remote sonars to map the sea floor, and its sensor systems have analytical chemistry instruments that measure ocean salinity, temperature, and oxygen levels, all controlled by multiple computer arrays that send it out to do its analysis and bring it safely back to the ship every day. An earlier iteration of the instrument was used to locate the *Titanic*.

- The **Biomedical Informatics Research Network (BIRN)**, a National Institutes of Health initiative that supports collaborations in the biomedical sciences using innovations in information technology. The collaboration includes 23 universities and 31 research groups, all focusing on infrastructure development, with the BIRN coordinating center located at the University of California, San Diego. The network has three test bed projects, but the common theme is that all relate to structural or functional brain imaging and the interrelationships of brain imaging of human neurological disorders such as Alzheimer's disease, depression, schizophrenia, multiple sclerosis, attention deficit disorder, brain cancer, and Parkinson's disease.

  The idea behind BIRN is to share knowledge, promote reproducibility, and facilitate new collaborative efforts based on the infrastructure that members have designed and the lessons and best practices they have developed. BIRN participants are constructing and promoting daily use of a data-sharing environment, so although the biological data

utilized by the consortium are from very dispersed sites, they are brought together in single unified database.

One of the programs taking advantage of the BIRN computing infrastructure is the UC San Diego–based Telescience Project, which emerged from research conducted at the National Center for Microscopy and Imaging Research. As noted earlier, in 1992, researchers at the center allowed people attending a research conference in Chicago to remotely control their electron microscope and then, also remotely, refine the data they had using the Cray supercomputer. In the mid-1990s, the center extended access to users throughout the United States. Researchers at the center then realized that if they were going to collect data on such a large scale, they needed to think about computation and storage resources. The Telescience Project is providing a grid-based architecture to combine telemicroscopy with parallel distributed computing, distributed data management, and archiving to support the data collection and analysis and interactive integrated visualization. This approach is providing an end-to-end solution to high-throughput microscopy and is merging technologies for grid computing, remote control, and federated digital libraries with multiscale data, such as cell structure data.

An outgrowth of this project is the Visible Cell, a collaboration between Telescience and BIRN in which scientists are building digital visible cell models using multiscale data sets through interconnected three-dimensional images representing subcellular, cellular, and tissue structures. They are also creating simulations to advance understanding of the impact of structure on function in living organisms and are producing materials useful in K-12 education. Their goal is a better understanding of the relationship between structure and function and applying it to elucidating the neurological relationships.

Additional examples of cyber-enabled instrumentation dedicated to education and training include webSEM, a free-service scanning electron microscope (SEM) connected to the Internet that allows educators from around the world to log in and use the microscope in their classrooms, and MIT iLab, remote online laboratories dedicated to developing and disseminating technology and pedagogy for sustainable and scalable iLabs in monitoring physical infrastructure that can be shared worldwide.

Social considerations

Along with the scientific and technological opportunities that cyber-enabled instrumentation presents, the introduction of remote access to scientific enterprise also has the potential to alter the interpersonal and group interactions that inevitably accompany these new approaches to investigation. In an era when social networking over the Internet is an established phenomenon, cyberinfrastructure inevitably will change the social dimensions of collaboration and learning. Experience is showing that creating a cyber-enabled environment is not simply a matter of directly transplanting activities from a traditional environment. Interactions that take place in cyberspace are inherently different from those taking place in person, whether in the context of a classroom or a laboratory.

VIRTUAL COMMUNITIES

A virtual community is a group that interacts primarily via cyber-mediated tools rather than face to face. Much as people mobilize such Web sites as MySpace and Facebook to connect with one another, scientists already interact in virtual communities for professional purposes. This approach is now being extended to encompass the educational arena with the formation of learning communities where members connect with and learn from others by collaboratively participating in the construction of new knowledge. These virtual communities, regardless of how they come together, will probably drive the synthesis of second-generation Web 2.0 technologies, which aim to enhance the creativity, communication, collaboration, secure information transaction, and functionality of the Web.

Author Howard Rheingold notes in *The Virtual Community*, "People who use computers to communicate form friendships that sometimes form the basis of communities, but you have to be careful to not mistake the tool for the task and think that just writing words on a screen is the same thing as real community." In virtual communities, for instance, there is a notion of hidden reciprocity—that is, when someone sends a message online, he or she expects something back. This expectation, not necessarily articulated, is that something will happen, that the sender will receive something back magically through cyberspace. This is a different dynamic from a conventional social environment, where one can just sit and listen and doesn't necessarily have to give anything back.

Because virtual communities represent both information science and social science, they also reflect a dualism that can have a significant impact on researchers. Those from different disciplines, for example, may define and use the same terms differently and need to find a way to align their usage and understanding. As Mark Ellisman, director of the BIRN Coordinating Center notes: "Although all disciplines can learn from BIRN's experience, the sociological transition from a physical to a virtual research community is a process that each field and discipline needs to go through on its own."

It may be, therefore, that the long-term key to creating and sustaining vibrant virtual research communities is to train a corps of scientists and engineers who have expertise in information technology as well as their primary discipline fields.

LEARNING IN CYBERSPACE

In higher education, students spend more time online than they do with faculty, so a cyber-enabled environment can play a critical role in enriching the learning experience. Just as communities that develop online differ from those taking place face-to-face, however, learning in an online environment is unlike learning in a traditional classroom or lab; it isn't possible simply to post online what takes place in a classroom because working in and acquiring information from cyberspace is an altogether different experience.

One of the factors currently affecting online instruction is that students are "digital natives," while faculty are often "digital immigrants." Having grown up with technology, students know how to communicate online, instant message, and hand off controls to each other; they are unfazed if something breaks down. By contrast, faculty—many of whom came of age before the widespread adoption of computing—may or may not have a comparable level of comfort and expertise with technology.

While students' facility with technology holds great promise, it can be a mixed blessing. At the same time they are accustomed to it, they may be mesmerized and overwhelmed by electronic media. They can conduct searches on their cell phones, for instance, but often cannot filter what the searches turn up. Students—and faculty, for that matter—may fail to realize that to learn, they need to draw on different learning styles—to be "SAVI," where S represents somatic, or learning by doing or by physical activity; A stands for auditory learning—by talking and interaction; V is for visual—learning by watching and listening; and I is for intellectual, or learning by reflecting, thinking, and analyzing.

Everyone has different learning strengths, and most people have a combination of these traits, usually of two. *Connectors*, for example, are AV types; they observe and talk. *Analyzers* are VI types, observing and analyzing. *Appliers* are IS types, learning by analyzing and doing. *Innovators* are SA types, whose strengths are doing and talking.

By their very nature, cyberinfrastructure and multimedia can accommodate all of these types and therefore can be effective at playing to diverse learning modes. But most online instruction and activity today has not yet capitalized on that potential. Instead, these cater primarily to analyzers (VI types). Although most research scientists are VI types, moving forward in developing online educational and training opportunities, it is important to broaden opportunities to reach people who learn by other modes.

Challenges in cyberinfrastructure development

Multiple cyberinfrastructure components must be integrated and coordinated to provide a genuine "at-the-instrument" experience for remote users. These components include:

- Instrument lab resources
- Networking for remote access of instruments and data
- Data storage
- Computation for data analytics (e.g., data visualization) and
- Software applications (e.g., Web portals) that provide user interface to the components.

For successful integration, scientists and technology infrastructure providers must work together to configure network bandwidth, computation, and data storage resources. In addition, scientists and application developers have to work together to design work flows between the instrument labs and remote users. For example, application developers might need to customize Web-conferencing tools for interaction between a remote instructor and multiple remote students accessing a cyber-enabled instrument as part of a class. Or they might need to develop Web

portals that allow searching and sharing of archived data sets as required by the scientists. Hence, developing successful cyberinfrastructure requires a multidisciplinary team comprising a) instrument labs; b) networking, computation, and data storage infrastructure providers; and c) application developers. Cyberinfrastructure development, is a fairly new phenomenon that continues to evolve rapidly with advances in technology, and the learning curve is steep for everyone involved.

The Ohio Supercomputer Center (OSC), a technology infrastructure provider, is working closely with universities in Ohio to enable the sharing of instrument resources with national and international partners. OSC's application development expertise is being leveraged for these efforts, and a "Remote Instrumentation Collaboration Environment" (RICE) software framework has been developed that is being customized for a variety of computer-controlled scientific instruments. Partners include Miami University, Ohio State University, and Ohio University, and the instruments include an NMR spectrometer, unipulsed EPR spectrometer, transmission electron microscope, scanning electron microscope, Raman spectrometer, McGraw-Hill telescope, and nuclear accelerator. RICE can support multiuser desktop sharing and integrates collaboration tools (chat, presence, and control-lock passing) to orchestrate instrument control among multiple remote users. RICE is accessed via a remote instrumentation Web portal developed by OSC, which has Web services to centralize handling of user accounts, user privileges, user authentication, client software distribution, project and experiment management, communications between remote users and instrument technicians, and remote monitoring of experiment progress.

Ashok Krishnamurthy, Ph.D., senior director of research, and Prasad Calyam, Ph.D., senior systems development/engineer of the OSC, presented an overview of their experiences with developing RICE variants for cyber-enabling their partners' instruments and described the policy and technical challenges they have had to overcome. Details of their experiences with their partners can be found in their IEEE e-Science 2008 paper available at http://www.osc.edu/research/networking/projects/rice/riprogram_escience08.pdf.

Among the issues that have emerged at OSC is the fact that cyber-enabling may be written into a proposal, but when the proposal is funded, the principal investigators of instrument labs are not adequately cognizant of the set of steps required to implement it. As a consequence, OSC recommends and has been proactively engaging researchers to consult with the center's staff before submitting cyber-enabled instrumentation proposals that require reliance upon OSC's statewide networking, computation, data storage, and application development expertise resources.

Among the challenges OSC has encountered are:

Technical Challenges

Bandwidth provisioning: Every instrument has unique control software with variations in screen content and control actions. Consequently, the end-to-end bandwidth demands required for accessibility between the instrument site and remote user site are different for each instrument. In particular, transferring optimum-quality screen images in real time from the

instrument in such a way that they are usable for a remote user requires end-to-end bandwidth capacity that needs to be quantified with usability studies.

Collaboration support: For remote observation and operation work flows, in addition to using VNC, the instrument technician and remote users need to be able to communicate effectively with each other. For this, collaboration tools that support peer-to-peer voice/video or voice/video conferences combined with text-chat are useful. In the case of multiuser sessions, collaboration tools need to support a "presence" feature, which indicates who is controlling/viewing the session. In addition, the operator must have the ability to manage control privilege among the remote users; i.e., the operator must be able to grant or revoke control such that only one remote user controls the instrument at any time. Furthermore, Web camera video that is accessible to the remote users via Web browsers can provide a surveillance feature, which allows remote users to view instrument lab personnel or instrument status (e.g., device display panels, sample holder).

Data management: During cyber-enabled instrument sessions where a sample is being analyzed, experiments are run, and data sets (i.e., images or text files) are generated. The analysis may involve visual inspection of the sample, coupled with image captures or experiment configuration script(s) invoked to generate text files. The data archive must handle metadata and provenance of the data sets—e.g., session time stamp, session owner(s), study context, or project name. Such information can be user-specified or parsed from experiment configuration script(s). Given that instrument computers are shared resources with limited data storage capacities, instrument labs need to transfer the data sets of their users to mass data storage systems that are accessible to remote users via Web portals or other file transfer applications.

System security: Cybersecurity is a complex subject because as people, instruments, and data become "connected" in the information space, there are ever-increasing threats of cyber attack, loss of or damage to data, and identity theft. It is vital to secure both the network in which the instrument resides and the data sets of the instrument users through judicious application of federated authentication systems and high-quality firewalls The basic TCP ports that need to be selectively opened to computers on the Internet (using, for example, firewall rules) include: a) port 5900 for VNC desktop clients, b) port 80 for VNCs and Web cameras' Web browser clients, c) port 22 for SSH, and d) port 443 for HTTPS. In addition, if voice and video calls need to be facilitated from the instrument lab, TCP port 1720 and a range of UDP ports (vendor-specific) need to be opened. Securing data, which may be distributed across several computer systems, involves authenticating and restricting file system access on the instrument computer with the appropriate permissions.

Policy Challenges

Service-level agreements: As noted earlier, the three primary stakeholders involved in developing and maintaining cyberinfrastructure for cyber-enabled instrumentation are: a) instrument labs, b) technology infrastructure providers, and c) application developers. Given the multidisciplinary issues involved in supporting multiuser cyber-enabled instrumentation, the instrument labs need to establish service-level agreements (SLAs) with both the technology infrastructure providers and application developers for routine cyber-enabled instrumentation

operations and ongoing maintenance. The SLAs serve as a mechanism to convey expectations and identify groups responsible for development and upkeep of the cyberinfrastructure components.

Use policy: Because the use-time of scientific instruments is valuable and there are security/privacy issues involved in cyber-enabled instrumentation, instrument labs should maintain use policy documents. A use policy document could describe a) considerations for obtaining user accounts for remote observation/operation and/or data management; b) scheduling priorities for local/remote users; c) guidelines to enable/disable remote observation/operation to prevent remote users from violating the privacy of local users working on the instrument; d) privileges for expert/novice users in single-user as well as multiuser cyber-enabled instrumentation sessions; and e) procedures for automatic failover to recover from cases where local/remote users become incapacitated during a cyber-enabled instrumentation session. These policies can then be used by application developers to build the necessary software components to implement the use policies and maintain audit trails of usage and problem scenarios in routine operations.

Usage billing: In addition to the high cost involved in initially acquiring expensive scientific instruments, substantial costs are involved in maintaining the equipment, which can have lifespans ranging up to ten years. These costs include the remuneration of personnel responsible for the routine operations of the instrument as well as the fees for physical facilities, networking, data storage, and computation resources. Hence, appropriate "resource units" need to be defined for use-time of the instruments. The usage billing can then account for the fees (e.g., fee/hr, fee/session) corresponding to the resource units consumed. Given that an instrument session setup or remote observation generally requires additional effort on part of the instrument operator, setup and operation surcharge fees could be included in the usage billing.

Future directions

Kantardjieff raised the following key questions about the future of cyberinfrastructure:

- How can we use cyberinfrastructure advancements and capabilities more productively than we are now?
- How do we create institutional and policy frameworks that can facilitate the use of this technology and support research collaboration through this infrastructure?
    - Where does the scientific community need to be going with regard to instrument cyber-enabling?
    - How should these efforts be prioritized for maximum impact on the broader chemistry community?

She then noted that the answers to these questions will likely require focusing on:

- Initiatives or strategic areas
- Resource allocation (broadly defined)
- Intra-agency priorities
- Mechanisms for knowledge transfer and reproducibility. (How do we share best practices and lessons learned so others don't make the same mistakes?)

- Scalability
- Sustainability, which means different things to different people, some of whom will address the issue of resources, while others see the issue as environmental.

Furthermore, rationales for establishing priorities may lie outside traditional criteria for intellectual merit and broader impact, making it necessary to start thinking creatively.

Environmental considerations

From an environmental standpoint, one of the most promising aspects of cyberinfrastructure is its potential to combat climate change. Those turning to cyber-enabled technologies don't have to travel to conduct research or teach classes, which cuts down on CO_2 and greenhouse gas emissions. While cyber-enabled remote access is not a substitute for travel in all instances—it is hard to initiate relationships with people in the cyber world, for instance, and there is considerable benefit to the intellectual stimulation that scientists derive from face-to-face interactions—it can still play a significant part in reducing the scientific community's carbon footprint.

Bill St. Arnaud, senior director of advanced networks for CANARIE, Canada's advanced Internet development organization, discussed how remote instrumentation can help limit global warming. The CANARIE network serves universities, colleges, schools, government labs, research institutes, hospitals, and other organizations in both the public and private sectors. It facilitates the development and use of its network as well as the advanced products, applications, and services that run on it.

According to St. Arnaud, responses to global warming will require dramatic changes in the way institutions conduct business, including research. The information/communications technology industry alone produces more CO_2 than the aviation industry, and it is doubling every four years—a level of growth that is unsustainable. In areas where electric power is coal-driven, one small server generates as much CO_2 as an SUV. On many campuses, cyberinfrastructure is the second biggest consumer of power after heating and cooling. At Canada's Simon Fraser University, a research-intensive institution in British Columbia, for example, the research building, where most of the cyberinfrastructure is located, is the biggest contributor to greenhouse gases. The facility is powered by hydroelectric power, which means the contributions to greenhouse gases would be even more dramatic if it was powered by coal.

This level of consumption is testing the limits of the electrical grid, and power companies are saying they lack the capacity to expand. As a consequence, many institutions are expected to experience power outages; in the next five years, it is estimated that 90% of all companies will experience some kind of power disruption, and one in four companies will experience a significant business disruption.

Governments too are wrestling with global warming and are exploring a number of options for addressing it, including:

- Carbon taxes, which tend to be politically difficult, and because they encounter stiff resistance from the public, are not likely to go forward globally;
- Cap and trade, which is useful for big power stations and emitters but only addresses the supply side of CO_2;
- Voluntary carbon offsets, a new development that represents a very immature market but one that is developing rapidly because there are many ways of making money from it; and
- Carbon neutrality, imposed by law, which is gaining momentum with a number of governments around the world, including Canada, New Zealand, and Australia. In Canada, British Columbia is mandating that all public sector institutions be carbon-neutral by 2010, and other provinces are exploring the possibility of implementing the same policy. This will have a big impact on universities, which will have to contribute to a carbon offset fund to pay for carbon-offsetting activities such as planting trees.

Most current approaches to reducing the carbon footprint are focused on the increased efficiency of equipment and processes. St. Arnaud advised skepticism about claims that this is a viable way of addressing carbon emissions, however. While it is part of the solution, it can also backfire. In 1973, during the last energy crisis, for example, Congress passed the first efficiency act, directed toward cars, appliances, and homes, as a way to wean the country off foreign oil. Thirty years later, what happened was that increased energy efficiency had reduced the cost to consumers for a product or service, so they bought bigger homes, appliances, and cars and drove farther. As a consequence, energy efficiency has resulted in greater consumption because efficiency has reduced the cost. So whenever vendors are promoting energy efficiency, it is important to ask what the long-term impact is.

Canada's PROMPT initiative—next-generation Internet to reduce global warming—a consortium made up of Bell Canada, Nortel, Ericsson, McGill University, and the University of Toronto, with participation from GENI, CAL IT2, Scripps Institution of Oceanography, and other organizations around the world, is looking at all aspects of participants' telecommunication and instrumentation networks to change the way they conduct research and many other aspects of life. Approaches include remote instrumentation and laboratories and redesigning optical networks, wireless networks, and last-mile networks, with the primary objective of achieving a zero carbon footprint. This strategy is based on the rationale that carbon neutrality, while balancing the equation, is not really solving the problem because the basic carbon is still being emitted. The goal is to locate equipment at sites powered by renewable energy, where no carbon is emitted at all—a far more sustainable approach than even carbon neutrality.

Remote instrumentation is a major part of the initiative. With the cost of green power in cities dramatically higher than conventional power, the advantage of remote instrumentation is that lab equipment does not have to be located next door but can be located where there is access to renewable energy. Google and Microsoft are already doing this; Google, for example, bought an old aluminum smelter plant in upstate Washington for a data center. The facility has its own source of hydroelectric power, and the price of energy will be guaranteed for 20 years. It won't be competing with other industry sectors, won't be dependent on a utility company utility, and won't be vulnerable to brownouts.

Canada is doing the same thing, using the CANARIE optical high-speed network to link up universities with these kinds of facilities. Examples include:

- A partnership between IBM and Rackforce, which is investing $100 million in a center that has its own hydroelectric dam, providing cheap, renewable power and long-term sustainability.

- The National Research Council's SpectroGrid system, which provides simple and secure remote access to NMR instruments located at the National Ultrahigh-Field NMR Facility for Solids.

- A data center in Nova Scotia that is being powered by a windmill company instead of connecting to the grid. By connecting directly to the data center and avoiding the utility company, the business will be much more sustainable. Container boxes with computers at several sites in the province will be powered by windmills, and as the wind ebbs and flows, the network will move jobs from one site to another.

- Green data centers built by Indian nations. Native peoples will be among the first victims of global warming, and they are very keen to participate in this kind of research.

- Beam line on your desktop—Science Studio, which will link beamlines at synchrotron and neutron facilities across the country, all built around Web services, so researchers can access different beamlines and do comparative analyses.

- Eucalyptus: A series of data-intensive tools based on Web services so graduate architecture students can link together and collaborate remotely.

- Project Neptune, which is developing remote undersea instruments, funded in partnership with Scripps Institute of Oceanography and CAL-IT2, with support from IBM. This project will have major implications for industry in such fields as gas pipelines and remote mining, which need these tools as well.

The novel way PROMPT is pursuing its goals is through carbon offset credits. As a technology is developed, either to help a researcher move off campus to a renewable energy site, to reduce a carbon footprint, or to commercialize a technology, rather than using traditional licensing and royalty fees, PROMPT will work with carbon offset brokerage firms to aggregate and sell offset credits, which are being traded on exchanges around the world. Remote instrumentation is eligible for these credits if it reduces travel or if it reduces a carbon footprint. Firms such as Google, Cisco, and IBM will purchase carbon offsets from companies using their technology to reduce CO_2. Typically, credits can be sold for anywhere from $2 to $20 a ton of the CO_2 saved, and PROMPT members hope this will be a real driver that helps institutions move in this direction because selling carbon offset credits has the potential to generate a significant amount of revenue.

Several studies indicate that information and communication technology and cyberinfrastructure can be more important than carbon taxes or cap and trade in reducing CO_2. One study from Japan

asserts that these approaches can lead to achieving 90% of the Kyoto Accord target. Another, the McKinsey report (*Reducing U.S. Greenhouse Gas Emissions: How Much at What Cost?*, U.S. Greenhouse Gas Abatement Mapping Initiative, 2007), showed that remote instrumentation and efficiencies can equal the CO_2 impact of China and the United States combined. Clearly, the promise of these technologies is vast, and their impact will be broad. As they are developed for the research community, they will also open up economic opportunities that will help universities, communities, the economy, and society.

Best Practices, Issues and Challenges

Users and institutions, while united by a desire to pursue programs incorporating cyber-enabled instrumentation, may be separated by policies, platforms, objectives, expectations, disciplines, experience, and/or time zones. Whether investigators are initiating a program, planning the necessary infrastructure, ensuring security, working through data storage and archiving issues, negotiating financial agreements among users, collaborating with colleagues, or assessing the project's success, best practices require an in-depth understanding of considerations that may be alien to a conventional research environment or, if they are encountered in a physical lab, grow markedly more complex in a cyber-enabled setting.

It is worth noting that in addition to inter-institutional collaborations, remote access may also be applied to instruments and resources within the institution. Cyber-enabling can greatly increase student access to instruments, for example, particularly for classes that are instrument-intensive,.

Regardless of the context, certain key issues are likely to resurface throughout the planning and implementation process. These include:

- The importance of establishing clear goals for the project and determining appropriate instrumentation for the program
- The need to engage the administration of participating institutions to ensure their buy-in and cooperation, particularly when the instruments they are being asked to support are at another site
- The critical need for strong IT support
- Determining how to meet the needs of users with different levels of experience and expertise and how to train those users, onsite and remotely
- Establishing appropriate levels of security that protect resources and proprietary information but also allow access without requiring users to navigate around burdensome technical obstacles
- Developing policies about intellectual property
- Establishing agreements for cost sharing to ensure the resources to maintain the instruments and
- Recognizing the value of simplicity at every juncture.

Getting Started

In all likelihood, the first phases of launching a cyber-enabled program will focus on establishing a viable group of users and collaborators and obtaining the necessary financial resources to move forward.

Best Practices

USERS AND COLLABORATORS

Before embarking on a project, it is essential to **define the role of research and teaching in the proposed endeavor and to plan appropriately**.

- Principal investigators should **identify the potential user base, onsite and remote,** for their project and determine whether that base includes institutions that NSF, NIH, and other agencies are particularly interested in serving, such as predominantly undergraduate institutions, tribal colleges, community colleges, historically black institutions, small branch campuses, or even high schools. At the same time, they should clarify how users might benefit from the endeavor, for if there is no benefit to them, it obviously will not work. If the project includes an outreach component to middle schools and high schools, the planning team should be aware that firewalls at high schools are a tremendous issue and make sure they have a strategy for addressing it.

- Once the user base has been identified, investigators should **develop a specific group of collaborators interested in participating in a proposal to fund the project and, as early as possible, obtain letters of support from both the individual users and the appropriate administrators at the participating institutions**. Establishing formal support early on is essential because it will help ensure that all participating institutions are willing to cooperate to allow their researchers remote access from their campuses campus to the instrument(s).

- **Determining the best people to lead the proposed project is essential.** If the principal investigator lacks the appropriate experience, then he or she should consider handing the lead to someone else if that person is better suited. It is especially critical to **have someone available who has IT experience and has previously implemented remote instrument access**, particularly at smaller schools, such as community colleges. This person may be an IT support staff member, collaborator, or even a colleague who is willing to advise the project team. Without someone in this role, it will be difficult to convince program officers and reviewers that the team has the requisite technical support to ensure the project's success.

INSTITUTIONAL SUPPORT

Institutional support is critical and may pose a significant hurdle, especially for novice applicants or smaller schools that are attempting to cyber-enable an instrument for the first time. It is important that the participating institutions be appropriate for what the collaborators want to do; if they cannot come up with IT support or maintenance or lack the clearances to carry out the proposed program, it will not move forward. Researchers should determine their institutions' policies about collaborations to make sure they will not be penalized for "helping" other colleagues and jeopardize their chances of promotion or tenure. Key authorities at the lead institution should understand the project goals before reading the final proposal.

- **Participating institutions must have the necessary space and facilities to carry out the program,** including IT support, maintenance, climate control, and security and safety

clearances. Investigators should **obtain written commitment of resources up front, from the appropriate administrators,** to ensure clear expectations from all collaborating institutions.

- **Ideally, collaborators and users will come up with some of the program's operating costs,** which means a formal agreement among the participating institutions must be in place as early as possible to ensure sustainability after grant funding ends. Whichever officers have the authority to commit resources at the different institutions must sign off on the agreement. If it is necessary to hire new staff to operate the upgraded/new instrumentation, then the agreement must spell out which institution(s) will provide the resources.

- To complete all the paperwork in time, investigators should talk with the grants and contracts offices of all collaborating institutions early on. The budget process should likewise begin early, to secure institutional support if cost-matching is needed, and clarify allowable costs within the guidelines of an RFP.

FUNDING SOURCES

Finding appropriate funding sources, whether federal, state, and/or private, is essential to bringing a cyber-enabled instrumentation program to fruition. Screening program announcements on a regular basis and looking for RFPs with *cyber* or *remote* in the title or abstract is a good place to start. As noted earlier, however, cyberinfrastructure development is a fairly new phenomenon that continues to evolve rapidly with advances in technology, and the learning curve is steep for everyone involved, including funding agencies. In addition to seeking such sources of funding information as reports and other relevant NSF and NIH documents, potential investigators should conduct Internet searches for similar proposals/grants and talk to program officers at federal agencies. It is an excellent idea to have a one-page executive summary of a proposed cyber-enabled implementation already prepared. Approaching the congressperson representing the institution's district about fund**i**ng opportunities is another strategy that may prove productive; even if the representative cannot provide funds, he or she may be able to direct investigators toward other sources.

- **Because many grants require matching funds, it is important to identify sources of matching funds.** Matching funds may also make it possible to go for a more useful system. If an institution secures a federal grant from NSF or NIH, the state often may come up with matching funds for up to a quarter or third of the proposal budget. Other entities that might provide matching funds include collaborating companies, university endowments, research foundations, etc.

- Seeking assistance from colleagues who have experience writing grants, even if they are not part of the proposed project, can help investigators new to the grant application process zero in on what they need to do. Increasingly, academic institutions also hire professional grant writers to assist with submissions..

IDENTIFYING GOALS

Clarifying program goals and user needs up front is critical to a smoothly running cyber-enabled instrumentation program. The answers to the following questions will guide appropriate planning.

- ✓ Is the aim to make an existing instrument remotely accessible or to obtain an entirely new instrument?
- ✓ Is the project something totally new, an implementation of a known technique in a new field, or a strategy to meet the additional demand for something known?
- ✓ Will the plans be implemented in several steps or in one large leap?

Throughout the planning phase, it is important to make sure participants' goals for the project are largely the same. Establishing this at the outset will help determine how the instrument is shared to ensure sufficient capacity for local and remote users. Similarly, it is important to plan an appropriate scale for the proposed activities to ensure that these are realistic and feasible. Instruments must be suitable for remote control data acquisition. This includes compatibility with current computer hardware and security standards. Provisions must be made for sample handling (shipping, storage, mounting, dismounting).

Once goals and user needs are identified, it is extremely helpful to know what is commercially available already because remote-enabling may be achieved with readily available off-the-shelf products. Contacting vendors about commercially available solutions and assessing their software and hardware for cyber-enabling potential will provide an important knowledge base, as will developing a relationship with their sales reps. The Internet is another excellent source of information about potential solutions.

Issues/Problems

- Inexperienced investigators, who misunderstand funding agency goals or RFPs, who fail to follow RFP guidelines carefully, and who are unfamiliar with FastLane and grants.gov will invariably encounter major obstacles as they try to attract funding for their projects.

- Institutional obstacles may include lack of administration experience, lack of support for grantwriting at small institutions, principal investigators who forget to check with the institution to provide institutional matching funds, lack of release time or other support, and institutional regulations governing remote access or other activity requirements.

- Challenges to identifying appropriate collaborators may include internal users who are not inclined to use remote access and local principal investigators who may not want to share instruments or research activities.

Insights/Observations

- A strong user base is always essential, and, most institutions have a built-in user base that should be cultivated along with an external one. It is important to give the potential user base realistic accessibility and cost projections before they commit to the project.

- It is impossible to please everyone, so investigators should select their collaborators wisely.

- PIs should do their homework and seek advice.

- A sound collaboration will formulate ground rules for instrument use and remote access.

- Knowing and anticipating hidden costs will save time and aggravation down the line.

- Actual problems are often unanticipated, and developing a cyber help desk could provide external institutions with the assistance they need, especially when they are in a different time zone from the institution where an instrument is located.

The Cyber-Enabled Environment

A well-conceived and well-implemented cyber-enabled environment strikes a balance between broad, fair access for users and robust security measures to protect the instrument, data, and intellectual property. It also features a reliable and enjoyable interface so users do not experience significant delays and inefficiency.

Best Practices

SECURITY

- Frequently, outside users are working with proprietary information. To protect intellectual property and to safeguard the system and instruments from hacking and break-ins, an effective cyber-enabled implementation must ensure interoperability and extensibility while maintaining multi-modal security for distributed direct and delegated authentication and authorization. This requires genuine coordination with local IT staff.

- To assure remote users a useful experience with the instrument, its **interface should be both enjoyable and reliable.** If the time penalty associated with using the automated reliable interface is high relative to the proximal console interface, users will prefer to use the instrument console, and the automation efforts will be in vain. (Production instrument software is designed to be used in a closed laboratory environment by professional technicians where physical security is available. Beyond the unsuitability of production interfaces for many educational uses, security issues associated with network accessibility are not generally considered and can become obvious when networked and used in an educational environment.) An unstable user interface is worse than no user interface.

FAIR ACCESS CONTROL

- **A reliable and trustworthy scheduling system must be in place** to guarantee that all users permitted to use an instrument are able to gain access to the instrument when needed. Moreover, the access control system must prevent monopolization of resources by a single user without compromising legitimate scientific needs. As a consequence, any access control system must be based on demonstrated mutual communication and respect among user group members.

PASSIVE DATA BACKUP

- **Data should be backed up to an off-site archive automatically and periodically.** This helps assure compliance with OMB circular A-110, which requires investigators to keep primary data for three years if they receive federal funding for publications. Security, privacy, legal, and regulatory issues regarding the long-term care of archived data need to be resolved, and best practices and associated use cases need to be catalogued. Archiving should be done *redundantly*. Furthermore, investigators must anticipate and plan in advance for handling the data *structures* generated by the instrument, the *amounts* of data to be archived, what will require immediate or frequent *access*, and what will not.

PORTABLE DATA STRUCTURES/ OPEN SOURCE SOFTWARE

- **All data formats in a shared facility must use the most portable and generally readable data format that is possible**. Software platforms for instrument access, data transformation and processing, and data management used in shared instrumentation facilities should be freely available to the scientific community whenever possible. This is in accord with and supports federal agency policies (e.g., NSF, NIH) regarding sharing of research tools and data. Although workshop attendees were not in a position to resolve the tension between vendor-proprietary formats and the needs of the community for more open standards, they did note the need for sustained development of the latter even if used as an archival format to meet agency data sharing requirements.

RELEVANCE TO NEEDS

- **There is a spectrum of need in cyber-enabling instruments related to the value of sharing access.** For example, while there is little need for a cyber-enabled melting point apparatus, there is great value and high return on investment for cyber-enabling synchrotron beamlines. The shared environment should be responsive to user needs, and software should be as user-friendly as possible. .

APPROPRIATE OVERSIGHT

- **The oversight committee of a shared instrumentation facility should accurately and responsibly represent the needs of the user group,** and principal investigator–level users should be active participants in instrumentation oversight.

WEBCAMS

- **Remote users should be able to view experiments via video cameras** whenever possible to enhance the user experience. Telepresence engages remote user and local technicians more deeply in an experiment. End-to-end bandwidth to support an adequate level of telepresence experience needs to be included as a global requirement, as do tools for assessing (and reserving where possible) bandwidth for video and audio interactions. Clearly there should be adequate bandwidth for interpersonal communication as well as instrument access, and data movement and priorities may have to be established to balance these possibly competing uses of bandwidth.

Issues/Problems

BARRIERS TO ENTRY

- **Space, physical infrastructure, facilities, and IT resources must be sufficient to support a shared instrument facility.** Reliable utilities, climate control, and building space are essential. Construction of instrumentation facilities in temporary or non-code-compliant space should not be supported. As noted previously, permanent and skilled IT staff are essential to the success of any cyber-enabled instrumentation facility. In addition to IT technical support staff, this includes instrument technicians to support remote sessions and to provide user training.

- **Policy barriers:** Securing administrative support from remote client users is an ongoing problem for facility administrators,.Administrators who have oversight authority for any instrumentation facility must indicate clear understanding of the facility's work in any letter of support that accompanies a funding application for the facility.

- **Recognition of scholarship and evolution of reward structures in teaching and learning:** The time investment in teaching to support cyber-enabling of instruments—for example, providing remote lectures to classes and users—is not generally recognized as scholarly contribution to the extent that is needed. However, a full discussion of stresses on "traditional" reward structures for faculty involved in highly collaborative teaching and research is beyond the scope of this document to resolve. The workgroup noted that developing teaching and research collaborations based on shared instruments may be an ideal avenue through which to explore new models for recognition of scholarship.

- **Faculty participation and buy in:** Cyber-enabling of a shared user facility must be supported by all of the principal investigator–level users. This must be considered in

constructing the fee structure, if any, for the facility. The "free-rider" problem must be considered and avoided.

SUSTAINABILITY

- **The need to generate revenue for future maintenance costs:** Remote users must shoulder a fair portion of the cost burden of maintaining remote facilities to which they have access. Administrators on both sides of this arrangement must understand and support the mission of the facility. In-house expenditure and recovery models need to be reconsidered to include the cost of labor to support remote access and associated technologies.

- **Carbon zero/energy efficiency:** No strategic plan for high-power-demand instrumentation should be adopted unless it provides for minimizing environmental impact in the long term. A primary positive impact of remote access is to reduce the travel-related carbon footprint of users. Beyond merely recognizing savings at the agency program level, one might consider including carbon-related savings in a cost allocation model for a remote access-enabled facility, a strategy being increasingly adopted by other governments.

PROTOCOL/STANDARDIZATION/INTELLECTUAL PROPERTY

- Intellectual property issues must be addressed in the context of "open science" and extended collaborations implied in research based on shared instrument resources. Agency and institutional regulations as well as legal statutes create a complex web of obligations that must be resolved early and often in extended instrument-based collaborations. As a consequence, the best practices in these areas remain challenging to attain.

BALANCE BETWEEN RESEARCH AND TEACHING

- The balance between research and teaching uses depends on the kind of institution or institutions that will be using the instrument. If the instrument is to be in a predominantly undergraduate institution, where teaching and research training play prominent roles, the instrumentation selected and cyber-enabled, as well as the access schedule and billing, should match that pattern of use. Similarly, if it is to be in a research institution, where it will be used exclusively for research, the instrumentation selected and the access schedule and billing should match that pattern of use. Where the institutional partners are mixed, the participating institutions should identify their objectives, agree upon the balance, and integrate them in the facility planning phase.

Insights/Observations

- Outreach to potential and actual user communities is a difficult task. Engagement with infrequent users is a particular challenge. Pre-planning is required to understand how to minimize user startup costs.

- In keeping with the notion of streamlined user engagement, the facilities and services themselves should be simple to understand and to use.

- Situation-appropriate design, purchase, and use minimize waste and frustration.

- Training and continuing education at the student, faculty, staff, and user levels are essential to maximize broader impacts.

Infrastructure

Simplicity and compatibility are the operating principles when it comes to planning the project infrastructure for cyber-enabled instrumentation. Facilities installing remote access capabilities cannot be burdened with hardware and software that is difficult to install, operate, and maintain. End users will not tolerate software or systems that are complex and difficult to use. The ideal project infrastructure should be platform-independent and accessible by users from multiple platforms. The host infrastructure should be compatible with the guest infrastructure, with guest needs ultimately guiding the host infrastructure configuration.

Infrastructure, including people, hardware, software, networks, and space requirements for remote access, should be included in the overall cost model for the facility and explicitly recognized as ongoing.

Best Practices

INSTITUTIONAL SUPPORT

- **Administration:** The administration should commit to appropriate support *before* a proposal is submitted to NSF or another funding agency.

- **Physical/central plant:** The physical plant should be as "standard" as possible rather than unique or specialized and should be robust to minimize downtime. Physical access and privacy issues engendered by the use of video cameras and audio communication tools need to be included in the physical requirements for the facility.

- **Information technology (IT):** IT networks should be adequate for the requirements of the facility with general rather than specialized cyber-enabled instrumentation modules; they should be able to give real-time feedback on the experiment and results. Although it is not possible to control all aspects of end-to-end connectivity, it is possible to monitor

and diagnose conditions that will affect the quality of the remote access experience. Dedicated IT staff should be available to the scientists.

- **Technical support (**instrument technicians or alternative technical support): It is unreasonable to expect facilities to grow or provide access if IT support is provided only by an individual faculty member. Technical support may include project coordination (shipping samples, scheduling, trouble shooting, training, and help).

- **Cost sharing:** Sharing of instruments and remote access creates new opportunities for cost sharing, for example an instrument facility may collaborate with a computing and storage facility to share costs over the total life cycle of data.

MANAGEMENT PLAN

- **Organization:** Project leadership should engage with local network and computing support operations (IT) to convey requirements and short- and long-term expectations about what it will take for the facility to be successful. The requirements could be organized as "service-level agreements" with corresponding price structures if reimbursement for services is necessary. This information will be critical to the construction of a clear proposal.

- **Experience/expertise:** Faculty and staff at the project site must have demonstrated experience and expertise with the cyber-enabled instrumentation. To effectively and efficiently resolve problems, facility managers should rely on well-known processes such as ticket-and-tracking using standardized tools.

- **Governance:** The principal investigator should develop and maintain a "safe-use" policy document that details standard operating procedures and restrictions for both research and training uses.

- **Leadership and roles**: Project roles should be clearly defined, with some indication of credentials or experience needed. The need for various players in various roles should be well-understood so that the right people are available at the right time.

SUSTAINING THE PROJECT BEYOND INITIAL FUNDING/IMPLEMENTATION

- The infrastructure should be scalable and should be continuously developed. As with maintenance of the underlying instrument, project leaders should secure life cycle funding for the technology remote access capabilities.

MAINTAINING THE PROJECT

- The project should employ scientific and technical support staff to provide experienced scientific user support and 24/7 technical support. Mechanisms for this kind of support will be site specific.

Issues/Problems

INSTITUTIONAL SUPPORT

- **Physical/central plant:** Facility management must collect measurements of system performance (cooling, heating, power) and adapt resource upgrades based on the measurements.

- **Information technology:** Key to a successful cyber-enabled instrument is institutional support for obtaining sufficient bandwidth across and to and from campus. Measurements of system performance must be collected, and resources adapted or upgraded accordingly.

- **Cost sharing:** Guest institutions and users should expect to pay on an ongoing basis so that their usage covers ongoing service contracts for instruments. Use monitoring and chargeback software might be provided, and as needed should account for remote users as well as local ones. Scheduling services need to be accessible to all users, internal and remote.

- **Investigator release time:** When faculty are chiefly responsible for development and management of the cyber-enabling project, frequently release time is promised but not delivered or is not cumulative, or release for many things is counted toward the same block of release.

MANAGEMENT PLAN

- **Training:** Training must be tailored for differing user needs. It should also be coordinated with the recruitment and training of faculty and staff within an institution to help develop an internal user base. Training must also include IT skills requirements for cyber-enabled instrumentation.

- **Governance:** Institutional policies need to be developed to govern how time will be allocated for instrument use between undergraduate students, graduate students conducting research, remote users, the needs for specific times of use by collaborative users, classes, etc. They should include criteria for setting priorities and specify who will establish the policies.

SUSTAINING THE PROJECT BEYOND INITIAL FUNDING/IMPLEMENTATION

- Who will pay for the cost of technology upgrades? Typically, end users are not prepared to bear this cost.

- How will "zero carbon" cyber-enabled instrumentation facilities be developed/funded?

- It is important for hosts to coordinate changes to their infrastructure with their remote users, because altering infrastructure may have a significant impact on the remote users' ability to access the instrument and adversely the project's sustainability.

- Institutions looking to share instrumentation are typically those that are least able to afford it. There is a tendency when preparing proposal budgets to say that there will be no user charges during the term of the grant, because the grant funds will sustain the facility. However, lack of sufficient funding sometimes forces a facility to close after the period of NSF funding ends, thus prompting someone else to start up a new facility. How can a group start with a model that can be continued and sustained?

Insights/Observations

- Larger institutions typically have better infrastructure to maintain an instrument and sustain a facility. Outreach to the community should be a strategic goal in instrument acquisition and operation projects.

- Standardization of IT at all levels, including K-12, is important to allow these institutions to readily connect to cyber-enabled instrumentation hosts. Applications such as chat, VoIP, and Webcams are effective in communicating between hosts and users. As institutions move toward federated authentication schemes, accommodating all levels of remote users should be considered on an as-needed basis.

- Integration of remote access into the higher education and K-12 communities must also go hand in hand with curriculum reform. K-12 teachers are tied to district and state mandated standards. Introducing change into higher education curricula can be difficult due to local policies, faculty time constraints, budgets, and infrastructure. Outreach efforts must include planning for, developing, and implementing new curricula that integrate remote access to fit specific educational goals.

- Lab and service provider staff should attend joint training sessions to educate each other about user and service requirements. After training, remote users should be identified as "Expert" or "Novice" users to determine their user privileges.

- Typically, sustainability plans are not well thought-out in advance. Development of sustainability plans needs to include all participants, from instrument providers to users.

- Software updates must be backward compatible, and it is important to have a process to communicate with users, including who to contact if the upgrade is not backward compatible.

- Many people have an idea of the science they want to pursue, the equipment they want, and the networks that need to be in place, but they must also carefully consider what is required require in terms of the "people" aspects of the project. Users will interact with technical and administrative staff at the instrument facility, and many other support professionals from IT and other areas will be involved in making the facility work. A key factor to the success of a remotely enabled project is also getting the "human" interface right.

- Differences of opinion about cost sharing remain unresolved. Some people say a best practice is no charge to predominantly undergraduate institutions or smaller users. Others say they should be charged. In at least one multiuser facility, when the principal investigator works with a microscopy suite, his or her students get free time on the instrument, while other users are billed during the period of the grant. In a different facility, another approach recoups costs from different sources. If the instrument is used for research, there are chargebacks. If it is used in a training context to teach courses at other campuses, the host facility receives a percentage of the lab fees back to offset the cost of consumables. Much smaller institutions that do not have much research funding expect not to be charged for using the facility. Generating the funds to keep the instrument upgraded typically requires considerable contracting for instrument use by outside entities.

Security

Best Practices

Best practices for security do not necessarily mean "tightest possible security." Too much security in the form of firewalls, log-in restrictions on instrument consoles, encryption, etc. may reduce the usability of the facility and lead to failure of the project. One should aim for the best balance between security and usability. Local area networks and federated authentication schemes can help to provide secure access to computers controlling instruments. Effective policies for data handling, warehousing, and integration across multiple sources should be implemented for managing data security, privacy, and integrity.

If a facility is successful and is going to grow, security will be an integral, appropriate, and well-designed part of it. Security issues and models change, so security components should be replaceable as needed.

SET UP RULES/POLICY FIRST

- It is important to define policy up front for data access and retention, users, and virtual organizations (at both the research group level and the institutional level). For instance, if an instrument is being monitored for quality control, can a user gain access to that data? What are the roles and responsibilities for local and remote users, as well as personnel at

the facility? Establishing clear responsibilities and permissions is particularly important if a group of investigators wants to create a network of similar or complementary cyber-enabled instruments at multiple sites.

- Classification of safety- and non-safety-related issues must be considered early in developing a remote instrument operating plan. Questions such as what happens if network connectivity or power is lost during remote operation must be carefully examined and incorporated into the design.

USER SECURITY

- Where there is a large user base, it is good to develop central resources such as federated authentication systems that identify users, how to segregate and authorize them, and how to track what they did if someone damages the instrument.

- Where connections are peer to peer, there should be a graded system, like a friends list, so access is not so secure that it becomes a barrier. Many people starting out with remote instruments will be working with one or two people, not hundreds, and it is best to keep protocols simple.

- Open standards should be identified and adopted for identity authentication (such as Open ID and identity federations such as InCommon). In general, a Public Key Infrastructure (PKI) encryption and certificates should be used to maximize interoperability.

SYSTEM

- At the system level, it is important to do logging, on the system and from the network, and it should be centralized so the system itself is reporting its state to a third party.

- Instrumental device/system logs should be audited on a regular basis.

- Servers and instrument consoles should be placed in a "secure collaboration zone" outside an organization's normal firewall architecture and having more outside user accessibility by limiting exposure of resources inside the main network.

INSTRUMENT

- In addition to supporting interactions with users during instrument use, a video camera can provide additional security for the instrument and the space around it. Video can also be a useful diagnostic for instrument damage through accident or intentional misuse.

- Contamination issues need to be considered when multiple samples are in use and being exchanged.

- The parameter files used to operate an instrument should be secured against accidental or intentional damage.

DATA

- Passive and unattended data backup is essential.

- Quality assurance of data should be done during data collection.

- Data "owners" need to have control over who can see and use their data sets.

Issues/Problems

IDENTITY MANAGEMENT AND SECURITY

- User identity management is a critical part of security, and yet a tension exists between instruments being a part of a virtual organization vs. being embedded in an organization in a university that has its own user identity approach. How does a facility use security at an institutional level as opposed to what it requires of outside users accessing the equipment, and how is authorization handled if external users cannot participate in the identity management scheme of the instrument's "owning" organization?

- Is federated identity management a possible solution for outside user access? If so, how should identity be integrated/federated into both the institution and the instrument facility?

- Instrument use must be tied back to individual users or groups, potentially through identity management schemes internal and external to the "owning" institution.

INSTRUMENT

- Often instruments have embedded operating systems that are not upgradable. What forces upgrades most rapidly in a cyber-enabled environment are security issues, so there is essential tension between security and the inability to upgrade.

- Forgetting to turn off the remote camera is a security issue because it can violate local user privacy. Some institutions have the camera turned on and off locally.

- Human oversight is invaluable, but how can minimal staff monitor access without it becoming tedious/boring?

- Even more important than the data coming off the instrument itself is information that reveals the characteristics of the sample and that may be considered extremely closely held data—about sample preparation or, for instance, if there are industrial partners in the program. If that information is revealed at the instrument level, it also needs to be protected at the instrument level.

- Physical and network security of instrument consoles and associated servers needs to be accounted for.

- Accessibility of instruments behind firewalls or through VPNs may be limited, and the impact needs to be evaluated and mitigated.

SYSTEM

- Users must comply with the local security policy for all servers and desktops involved in a remote instrument facility. This policy must be understood, and risks with respect to operation of the facility must be mitigated.

EXPORT CONTROL

- Used sample activation and return of the sample are an issue: How are samples shipped, tracked, etc. to meet legal requirements for shipping (EPA, health)?

- Export and control laws are very particular and can get a university and principal investigator in trouble. Of particular concern is the federal government's "deemed export" rule. If a graduate student is from a country the U.S. has designated as a terrorist country, the situation is particularly uncertain. This is a troublesome and not uncommon situation. Remote access may also involve situations covered by export restrictions, so these cases and how to respond to them should be thought through in advance.

Insights/Observations

- All instruments are remotely operated; the question is whether it is from 30 feet or 300 kilometers away. That difference is a social policy frontier, not a technical frontier.

- Cooperation is best motivated if it is in the program announcement. Trying to achieve it after the fact may or may not ensure buy-in from local organizations.

- The instrument community needs to be able to trust remote users not to be malicious and have in place an education, training, and monitoring capability to support users adequately.

- Frequently discussions start out addressing one instrument provided to many users, but often the instrumentation will build, so it may be that there are two best practices: one for novices, but another for a larger center with many users.

- What are the options when a principal investigator's collaborators are in a country the U.S. government doesn't like, and he or she wants to give them access?

Data Storage and Archiving

As with traditional instrument facilities, when setting up a remote instrument facility and resource, future archival and storage needs at the facility must be considered up front, and they have to be built in, or users will inevitably complain that their data was erased when the next person came in. Many facilities view the instrument as their primary focus, not what happens to the bits of data after they are acquired. This needs to change, especially in the context of traceability of results and reuse of primary data.

Best Practices

Planners should consider initial and future storage and archival requirements of an instrument facility or resource within a framework that includes:

- Design of an overall architecture that meets the facility and user community needs and that includes facility and user storage components and networks that connect them. Design of overall architecture that meets facility and user needs is an essential part of these requests. Sensitive data must be segregated from non-sensitive data, and who has access to these must clearly be established. Data relevant to teaching should be maintained separately from data relevant to research. Data implicated in intellectual property must be carefully protected, and users should be aware of who has access to it and for what purposes.

- A plan for scalability and evolution, including:

    - **Capacity planning:** Instrument plans should include capacity planning for storing, transmitting, and computing with the data that are produced. Systems should accommodate scaling to estimated future requirements.

    - **Technology refresh planning:** Because storage technologies change, continuity of access to data means that planning for technology upgrades should be considered. In general, a life cycle funding approach should be taken with the remote access components of the facility.

- A plan for continuity of operations/data assurance and survivability, if these are important, including:
    - Data backup at redundant sites
    - Failover systems for networking, storage, and computing

- Replica management systems for keeping and finding copies of data sets

- A plan for privacy, security, and legal/regulatory issues, appealing, if necessary, to federal regulations and implementation for solving them, that includes:
  - Layers of access—who can access what and when?
  - Experimental data embargo for publication or limited use and then publicly available
  - Role-based access
  - HIPAA (and possibly FRPA) considerations and possible adoption of ISO 17799/27001 data management standards as needed

- Use of standards for information capture at the source, including:
  - Laboratory information management systems
  - E-notebooks
  - Adoption/use of common file formats across the user community

- Use of provenance standards for annotating raw and derivative data
  - Raw data should be made available as a prerequisite for publication to prevent fraud and enable independent analysis.
  - OPM (Open Provenance Model) could be used for tracking data and its transformations.

- Use of open standards that enable data sharing and metadata capture and searching
  - Adoption/use of common, easily searchable metadata schemas across the user community
  - Improvement of provenance data depth by including workflow step(s) to extract metadata from data files as data are processed
  - Tagging data before archiving to ensure search and retrieval

- Use of open and standard storage technologies and techniques where possible
  - Use of open and/or widely used and robust storage protocols and distributed storage systems such as WebDAV, GPFS, HPSS, GridFTP, CIMA (as a storage/transport protocol), OGSA-DAI

Issues/Problems

- Reliability and robust access
  - Replication—how can copies of files be stored in multiple locations for data assurance?
  - Caching—how can local copies of files be made rapidly for computation without risking damage of primary data?
  - Backup—how can data be secured as backups?
  - Disconnected use
    - How can users on, say, airplanes work with data from instruments?
    - How can data from instruments be accessed without a network?

- Operational costs and issues (hidden costs)
  - Capacity expansion
  - Technology refresh

  - Redundant power
  - Security
  - Bandwidth
  - Personnel costs

- Curating and archiving
  - Media replacement
  - Format conversion
  - Verification of data through media transfers and format conversions
  - Deciding what to store for how long (data lifecycle)
  - Long-term archiving versus rapid access
  - Limited storage at high-throughput facilities
  - Integration and coordination of data across sources
  - Data backup and recovery.

- Privacy, security, legal/regulatory, ownership
  - Who gets to use data and when?
  - Who controls who has access to and received what data?
  - Non-repudiation (proving data is valid, recorded by a given investigator on a given date)
  - Research compliance issues—e.g., human subjects institutional review boards (IRB)
  - Data access policies derived from local institutions, agency policies, regulations, and laws
    - Enforcing data embargo policies
    - Making data public when publication occurs

- Other issues
  - Training requirements and implementation, safety training compliance
  - Incentives for recording metadata
  - Interoperation of data, data formats, and processing software across systems and across time
  - Initial costs vs. long-term benefits

Insights/Observations

- Instruments and other real-time data sources are major drivers of current and future storage requirements. It is estimated that real-time data will outstrip storage capacity by the year 2011.

- Issues are still complicated, and design has to be done within a framework and with *expert* advice. There are two types of experts: those who want to develop a solution and those who want to explain how complex the problem is, how hard it is to solve, and how far it is possible to go in terms of security and data storage, so a facility ends up with more than it needs.

- Even the most successful academic systems have only been implemented at a handful of sites. Robustness and continuity of operation for a software system are related to size of the user base.

- Many solutions have been developed to work in specific labs. Redundant development means that systems are highly variable in terms of robustness, scalability, cost, and the ability to meet federal budgetary requirements.

- A set of standards would allow early adaptors to begin implementing remote access systems for new instruments that can reasonably be expected to meet their needs for the lifetime of the instrument.

- Proposals for remote access and cyber-enabling schemes should have a clear plan for data management, including long-term operational and archival storage requirements and solutions. They also must indicate how users will move data from the instrument (as part of a total curating plan for data produced by the instrument/facility).

- Most common data management violations depend on what is going on. Major violations tend to be improper destruction and anonymization of data (incorrect redaction of patient data in medical and biological studies, or anything having to do with IRB research).

Fiscal and Economic

Cyber-enabling an instrument increases the user base and, potentially, productivity or throughput on the investment. It reduces operating costs and ensures that smaller institutions and industrial users have access. The greater the potential user base, the greater the potential usability of the instrument.

Determining what resources are available to support a cyber-enabled instrument program is essential to its success. By far, the greatest concern related to the economics/fiscal responsibility of sharing an instrument is sustainability; it is, in fact, the most important component of fiscal responsibility beyond purchase of the instrument.

Best Practices

USERS

- If a facility has industrial or outside users, an institution should consider charging them more than it does in-house users to help maintain the instrument and to avoid charging for outreach or purely educational efforts.

- Solid administration commitment (cost sharing, time release, room renovation or space allocation, etc.) should be in place prior to grant submission.

- To maximize instrument use, the principal investigator should be prepared to advertise to possible users and use wise time management and user allocation as needed, running longer experiments run at night, for example, scheduling runs that need staff during the day, or setting aside specific times for graduate student students to use it.

- Many instruments sit unused between 60% and 70% of the time. The first focus should be to increase institutional use of instrumentation using remote technologies. As an administrator, it is easier to justify use when the instrument is used a lot, and it is also easier to support. Only then should efforts be made to go off campus.

- Identifying the user base and creating a representative advisory committee will help ensure the long-term success of the project.

INSTRUMENTS

- The planning team should negotiate the best/longest warranty for large equipment as part of the purchase price, perhaps incurring no charge for the maintenance agreement for the first year and extending the agreement to a second year.

- A facility should locate instruments centrally to maximize investments in cyberinfrastructure technology. This has advantages relating to support/maintenance.

- Providing access at a reduced cost, reflecting reduced staff-user interaction time required for remote users, may be prudent.

- Targeting repetitive tasks for appropriate automation allows for expanded capabilities and more productivity with a single instrument, whereas a more specialized function may not be appropriate for automation.

OUTREACH

- Targeting outreach communities up front should be designed into the project and cyberinfrastructure time allocated for proposed outreach activities.

- Budget and design (i.e., bandwidth, browser, software, etc.) should be planned so that target outreach communities have little or no out-of-pocket costs for participation and little or no up-front time commitment.

- Exploring opportunities to generate IP and patents, create startup companies, or commercialize new products and ideas (licensing, SBIR, etc.) may advance sustainability by providing income. It may also aid in negotiations with vendors by improving the value of the vendor's product or by determining which vendors will be willing to work with the facility as it develops software or other enhancements for the products.

Issues/Problems

- If an institution is going to get an instrument, there needs to be a plan for institutional support to maintain it. There is no use getting an instrument in a lab or as a university facility and not having potential users know about it.

- To keep the instrument running, facilities should carefully consider scaled user fees (low or no costs for smaller colleges or educational centers) and the staff support needed. If a facility charges for educational use, should there be no charge, or should the charge be coupled to undergraduate fees at the participating institutions?

- K-12 and smaller colleges have particularly limited resources.

- When users are accessing instruments located at other institutions, administrators do not understand the endeavor and why they should support it when they cannot even see it. It can be helpful to sign memoranda of understanding with other institutions so it is clear what percentage of time they are using the instrument and why they are supposed to support it.

Insights/Observations

- In reaching out to K-12 and smaller colleges, successful cyberinfrastructure will facilitate access and be sensitive to the priorities and limited resources of those two communities.

- If faculty want to make sure all undergraduates have some access and understanding of instrumentation, remote access may be the best way to facilitate that, and it provides a new avenue to teach instrument-intensive courses. It will, however, present more complex challenges concerning policies, security, and pedagogy. Faculty need to rethink how they teach and incorporate instrumentation.

- It is an exceptional instrument that is in constant use. As noted earlier, some instruments sit idle as much as 70% of the time. One way to address this may be to single out the less-used instruments for cyber-enabling to increase their use. If, for example, it is possible to get students using instruments 24 hours a day, it will be a better use of what an institution already has.

- It is helpful to remember that unlike other instrumentation grants, personnel costs are allowable for cyberinfrastructure proposals.

Collaboration

Collaboration is central to a successful interactive facility, and communication, in turn, is key to collaboration at all levels of a multiuser cyber-enabled operation, whether its focus is education

or research. A facility may serve hundreds of users, some of whom show up once a year, while others check in every hour to track experiments. That variance in frequency of use, like the diversity of its user base, can affect how a collaboration proceeds.

Best Practices

- **Effective operator-client communication:** Communication is essential whether a facility is seeding or terminating collaborations. It is important to communicate across levels of expertise, including K-12 and the graduate level, if the collaboration includes those groups.

- **Keep it simple:** When a facility collaborates with users at different levels, it has to decide the best way to collaborate and communicate with them. Regular alerts and notifications via shared calendars, e-mail, and threaded discussion groups can be effective.

- **Accomodation and modification:** Cyber-enabled facilities should be flexible about making changes in the way that users are trained or instructed, expertise is distributed, contributions to and ownership of work are acknowledged, and the goals of the facility are assessed. Accommodations can be made to instructional methods and materials, the cyber-environment, time demands and schedules, and communication systems. Modifications may include changes to interfaces, content, policies, and skill requirements to use the instrumentation remotely.

Issues/Problems

- **Communication:**
    - Selecting collaborators
    - Establishing ground rules
        - Instrument/software requirements
        - Operator/user
        - Ownership of and credit for work
    - Terminating collaboration

- **Operational costs:**
    - Fee structure
    - Consumable supplies
    - Cost sharing
    - Instrument maintenance
    - Shipping issues

- How do you market and then support your capabilities?

- Although the instrument(s) may be cyber-enabled, users will likely need to meet with colleagues, whether it is in person or virtually in a collaborative cyberspace, to train on software or to be able to examine the data they have collected.

- Training on a cyber-enabled instrument can be challenging for some instrumentation if users cannot adequately visualize the instrument. This may require visiting the host facility for in-person sessions. How to develop effective remote training remains an issue, as does covering travel costs that individuals and institutions incur for training purposes.

- Not all samples are the same, and there is an overhead to preparing an instrument to analyze different types versus similar samples. Automated sample changing robotics are needed to effectively handle diverse samples..

- Scheduling issues that need to be worked out include prioritizing local vs. remote users, and determining and using excess capacity—daytime vs. evenings vs. weekends, for example. How can these off-peak times be made available to users?

- Intellectual property issues that will affect collaboration include licensing agreements, intellectual ownership, and confidentiality agreements. If a facility is collecting and sending serious data to a researcher and having to interpret it, for example, should that role be reflected in publications as coauthorship, or should it be acknowledged at the end of a publication?

- Replacement of an old instrument is one justification for starting a cyber-enabled program. But how is replacement defined? It may be more economical to "buy into" the use of a cyber-enabled instrument than to buy one new.

Insights/Observations

- Dissemination of best practices and lessons learned facilitates successful projects.

- Collaborations inevitably involve various levels of expertise.

- The needs or the research community are different from those of the educational community, so training collaboration is inherently different from research collaboration (asking questions during a remote use session versus controlling the instrument).

- Advertising and promoting the cyber-enabled instrument to potential users such as K-12, community colleges, museums, and community connections will lead to new and broader collaborations.

- Cyber-enabled instrumentation presents an opportunity to promote the sciences to younger audiences to help increase the number entering STEM fields.

- The centrality of participation and commitment may extend beyond cyber-interaction to include face-to-face interactions such as meetings.

- People change institutions, and the institutions involved have different capabilities. If a facility can provide a base of support, they can carry out their research anywhere.

- The cyber-experience must be enjoyable, engaging, and efficacious.

- There need to be appropriate rewards (fiscal, professional, personal) for all involved in the project; however, the system itself needs to be changed to reward collaborative research. ***In particular, if junior faculty members spend too much time helping other people through the collaboration, they may fail to get tenure themselves because the system does not recognize the value of collaboration*.** There also need to be rewards that exist in parallel for non-tenure-track participants such as staff scientists, to assure them a career path for their involvement in the project.

- There must be a means for professional development and continued learning.

Assessment

Program evaluation should consider the use of the cyber-enabled instrument(s) by multiple levels of users (K-12, undergraduate, graduate, researchers/faculty) and whether they are being used in educational or research contexts.

Best Practices

GOALS AND DOMAINS/ASPECTS

- **There needs to be a description and consensus around the goals for the project.** The goals, whether research or educational, should be well-defined and clearly articulated with detailed outcomes and deliverables. Sometimes there is a lack of consensus or lack of goals for a project with many stakeholders, players, and institutions; an evaluation needs to take this into account. To the extent that the participants can define the goals well and deliver on them, however, a much better evaluation will result. One of the evaluation points is always how well program outcomes match goals. If there are no clear goals, it will not be possible to address that point.

- **It is important to develop appropriate domains/aspects for the projects and indicators that can readily be used and that are aligned with the project goals.** (Domains are aspects or topics by which the project is assessed. Indicators help determine measurements.) Assessment will take place by examining indicators relative to the domains and making sure the domains and measuring devices match the goals. Sometimes a project will have good goals, but the domains and indicators are not aligned.

PEOPLE

- Good evaluators hard to find, and qualified expert are critical, so **forming a community of resources, through word of mouth or through other colleagues, can be useful** to identify someone who has knowledge and a track record to effectively execute and/or manage an evaluation project.

- **The assessment structure should evolve into communities of practice to pursue longitudinal studies to show growth in an institution or institutional collaboration. A** community of practice can be a group of users, K-12 collaborators, or administrators who support the project's work.

- It is important to track users, including those from underrepresented groups.

DATA COLLECTION

- **Data collection for evaluation and assessment should start early.** Sometimes evaluation takes place as an afterthought or without a lot of concerted effort. If the idea is to change the way people think about cyber-enabled instrumentation, however, in order to show change, there needs to be a baseline.

- **It is useful to consider different forms of assessment and evaluation.** Is a pilot assessment necessary to see if what is being measured will produce the necessary information? A pilot can be as simple as surveys that show how many people are using the instrument(s) or an increase in use or user satisfaction. These can take place during the progress of the grant as well as part of the summary evaluation. (As noted previously, it is important to establish a baseline before making in-progress measurements.) Ongoing and summary evaluations are both useful.

- **Using both qualitative and quantitative methods,** and going back to users for more input in an iterative cycle, will help ensure the accuracy of the observations and the validity of the evaluation and whether the findings can be generalized. It might also help define domains and aspects further.

- **There needs to be user input on usability.** Finding out if a collaboration is actually working can help lead to changes before many users have a bad experience or get turned off or the collaboration turns out not to be working as intended.

- **Assessment can be used for continuous quality improvement and better sustainability.** Dissemination of best practices and lessons learned, even if negative, leads to improved practice and better results and reduces wasted time reinventing the wheel. It is equally important to disseminate excellent results; otherwise efforts are fragmented, and investigators don't learn from one another.

- **Broader impacts should provide access to tools, people, and resources and in a way that can be measured.** For larger groups, surveys can work; for smaller ones, case studies might be appropriate. Agency guidelines regarding broader impacts are often vague, and some may be hard to measure, but it is important to have some metrics.

EVALUATION OF EDUCATION AND TRAINING

- **PIs should check with their institutional review boards, which review all research having to do with human subjects, on the need for review of the project's evaluation plan.** Each institution will have its own policy, but it is better to err on the side of caution because an IRB misstep can shut down a project quickly. Typical classroom activities usually gain rapid approval.

EVALUATION OF FACILITY PERFORMANCE IN RESEARCH USER SERVICES AND SUPPORT

- Data from the institution's administration may be available that will help estimate usage by students.

Issues/Problems

- There is currently very little data on the end-user learning environment.
- Data reliability is impacted by the quality of user training and/or user support.
- Cyber-enabling projects for a given instrument may yield results that are not reproducible at different labs.
- Effective management of a given data set is contingent upon establishment of and adherence to data release policies related to the institutions and projects involved.
- Socialization between instrument support staff, remote access facility developers, and users is required for full collaboration. The sociological transition from physical to virtual research community depends on scientific field and discipline.
- Evaluators are often not from principal investigators' home discipline.
- Thoughtful pre-planning with regard to intended and future uses of data is essential to provide for appropriate data archiving and management of data quality.
- How is it possible to know that students are learning how to use the instrument?
- If a study is using a control group, it will be difficult to select randomized populations in schools.

- It may be necessary to measure user base size, reduction in travel, etc., to know the success of the project.

- User expectations will vary, and "one-size" will likely not fit all.

- Correct staffing levels are critical for good service and good use of financial resources.

- It is important to assess the project's environmental impact.

Insights/Observations

- Collaboration-based projects require complex project management.

- There should be a clear institutional policy for users to acknowledge use of cyber-enabled facilities to the wider community.

- Evaluation can contribute to the development of contingency plans for situations in which outcomes, products, and services are not achieved or delivered.

- It is essential to provide the best possible end-user experience in all aspects, from performance to reliability to ease of use.

- Pre- and post-testing with regard to the content delivered by the instrument data, as well as IT skills required to utilize remote access may be useful for both K-12 and postsecondary contexts.

- Some aspects, such as informal learning, language barriers, learning issues, personal issues, etc., may not necessarily be observable.

- Financial accounting and feedback are essential; it is critical to keep good records.

- A clearinghouse of NSF project reports and publications would be helpful.

- Successful cyberinfrastructure will be taken for granted.

- Because evaluation experts are rare, a clearinghouse for finding qualified evaluators would be valuable.

- NSF's guide to evaluation in education projects may prove helpful for assessment of broader impact: http://www.nsf.gov/pubs/2002/nsf02057/start.htm.

- Funding for evaluation may be available at the institutional and/or governmental level—e.g., the NSF Innovation through Institutional Integration program (cross-directorate).

- It is necessary to evaluate capabilities at the facility and of the user at the site.

Future Directions

The opportunity to conduct scientific research in a more inclusive, cost-effective, and sustainable way will continue to shape future plans for cyber-enabled instrumentation and drive investment in cyberinfrastructure. Varied approaches—such as retrofitting existing instruments, developing modular components that work across platforms, providing funding in phases, creating a clearinghouse for information about relevant resources, integrating cyberinfrastructure projects with digital repository projects, and coordinating projects across disciplines—collectively will help advance the field.

To develop the human expertise and prepare a cyber-engaged workforce that possesses the knowledge and skills needed to design, deploy, adopt, and apply cyber-based systems will require a parallel focus on the human dimension to complement these developments, yielding new information about the best ways to incorporate cyber-enabled instruments into the classroom and laboratory. Methods to assess cyber-enabled projects will promote successful projects and share lessons learned.

Reviewing proposals

In evaluating proposals for cyber-enabled instruments and facilities, reviewers should remember that cyberinfrastructure implementations will vary for different institutions and users, but they must, first and foremost, meet the needs of the targeted users. If an individual instrument is to be cyber-enabled versus cyber-enabling a facility or institution, for example, or if the instrument or facility is to be centrally supported versus principal investigator supported, NSF funding should reflect these differences.

In proposals submitted to the Division of Chemistry, cyberinfrastructure should primarily support chemical research broadly, with other considerations secondary; proposal scientific merit and the ways in which the cyber-enabled components will *facilitate and enhance* this scientific activity should be the number-one criterion for funding.

Furthermore, the division should recognize that cyber-enabled research will not succeed without a commitment to cyber-education. Higher education as a whole does not currently do a great job of training students how to use instruments in a face-to-face environment. In the interest of time, instrument use is often treated as "black box." This deficiency should not be replicated in how they are trained to use instruments remotely. Students need to see instruments early and often in order to become proficient researchers.

Criteria for prioritizing submissions should include:

- Scientific merit
- Development of new capabilities
- The number of research projects impacted and efficiency of instrument use
- How extensively the instrument capabilities can be included in education.
- Portability of access—e.g., can an instrument be accessed anytime from anywhere?
- Sustainability that goes beyond financial and also presents a green IT strategy

- Adequate IT infrastructure to facilitate issues related to inter- and intra-institutional firewalls and connectivity.

Where appropriate and where there is a clear, demonstrated need, cyber-enabled projects should be encouraged. In some contexts, instrument requests with cyber-enabled components may be given higher funding priority. Grant review panels may be provided some guidance about instruments or mixes of instruments that may be effectively cyber-enabled, or for which there is a need for cyber-enabling. Where appropriate, program officers may consider encouraging the formation of instrument-sharing consortia, either before or after proposals are written based on cyber-enabling, to give instrument investments more impact and to stretch available funding. In situations where groups of users propose similar instruments, additional funding may be allocated to bring instrument capabilities to groups that are not selected for funding.

Instrumentation

RETROFITTING/UPGRADING EXISTING INSTRUMENTS

Cyber upgrades of existing instruments should be encouraged when they are realistic and cost-effective, leveraging investments in instrumentation that an institution has already made. Grants to cyber-enable current instrumentation will provide NSF with an efficient, economical model.

MODULAR DEVELOPMENT

Developing modular, reusable hardware, software, and training components that can migrate from environment to environment is a cost-effective and efficient way to speed the implementation of cyber-enabled instrumentation projects. The availability of a complete, scalable package of the elements needed to create a cyber-enabled facility will encourage institutions and collaborations to pursue very small projects. These "canned" products and modules can then be used for additional projects to avoid duplication of work. Ideally, these modules should be developed in such a way that they can be integrated into existing instruments to cyber-enable them. Tools based on Web 2.0 technologies will increasingly be required for federated authentication, with the value-added benefit of facilitating participation of K-12 and other users with limited resources.

Creating Community

CYBER-ENABLED INSTRUMENTATION RESOURCES

NSF should encourage the development of a series of informational resources for institutions and researchers seeking to create or expand a cyber-enabled environment. These might include:

- A cyber-directory, clearinghouse and/or Web site, searchable by geography, discipline, and instrument type and capacity, that includes best practices and technology.

- A *Journal of Cyber-Enabled Instrumentation*
- A consulting database of expert/users/groups available as resources for those embarking on a cyberinfrastructure program for the first time
- A virtual community of users of cyber-enabled instrumentation through the use of Web 2.0, RSS, social networking tools, etc.
- Regular workshops, tutorials, and symposia at ACS, ACA, PittCon, and other professional meetings
- An annual workshop dedicated to cyber-enabled instrumentation projects, technologies and lessons learned.

INTEGRATION WITH DIGITAL REPOSITORY PROJECTS

Exploring opportunities to establish stronger cyber-links with digital repositories such as ASDL, NSDL, and MERLOT will expand the usefulness of cyber-enabled instrument facilities and the research that they spawn by disseminating best practices, both in implementation and pedagogy, as well as lessons learned.

COORDINATED PROJECTS ACROSS COMMUNITIES AND DISCIPLINES

Approaching cyber-enabled projects from a community standpoint, whether that community is strictly scientific or comprises average citizens targeted by outreach efforts, may be an effective way to leverage resources. This approach necessitates looking at projects that are more science-based than instrument-based—focusing less on an individual technique than on a whole field of science. One project, for instance, catering to scientists, might create a community research database. Another might center around the theme of atoms, incorporating simulations, measurements, and experiments, while introducing elements of chemistry, physics, and materials science. The public might participate in various ways in sample collection, which students could then analyze—activities that would simultaneously engage average citizens and interest students in science.

Assuring Viability

PHASED FUNDING

Grant funding for cyber-enabled implementation projects may be highly effective if done in phases. A project might receive funding for a pilot or test bed in early stages, for example, and once it has shown some success in the initial phase of development, it could then compete for funding for a full cyber-enabled implementation. Once the cyber-enabled instrumentation is demonstrated to operate smoothly, it could compete for further funding to introduce outreach components in K-12.

The advantage of a phased approach is that it will likely increase the level of success; by going through test phases, a project will have the opportunity to reach the point where it is successful in later stages because it has already had successes in earlier phases.

ENSURING SUSTAINABILITY

The unique requirements of the cyber-enabled environment demand a strong, ongoing focus on sustainability, broadly defined. Because of the complexity inherent in creating cyberinfrastructure, however, sustainability must be considered differently for cyber-enabled instrumentation than it is for conventional instrumentation.

A sound sustainability plan will take into account issues such as the cost savings incurred by having more efficient instrumentation or a lower cost per sample. Equally important, sustainability must be viewed in terms of sustaining the instrumentation itself, and support for the instrument should be built into a grant. A successfully funded project also will consider the fact that turnover of computational hardware, due to technological developments or age, is more rapid than the instruments themselves, and will build in appropriate funding to address this disparity. Furthermore, a sound sustainability plan will address environmental consequences, such energy consumption and the value of reducing the carbon footprint. Issues related to inter-institutional sharing of resources will also be addressed—for example, development of agreements, recognition of teaching remotely, etc.

Education and Training

LEARNING IN A CYBER-ENABLED ENVIRONMENT

Because we need to understand better how people engage and how they learn in a cyber-enabled environment, there is considerable opportunity to explore the pedagogy of this field to refine training on the instruments, adjust instructional techniques, and develop appropriate metrics and assessments to accommodate learning via remote access. NSF should consider funding studies that address such questions as:

- What constitutes a good environment at the user level to support cyber-enabled instrumentation and encourage use of the instruments? How does this contribute to student learning?
- How does learning occur in a cyber-enabled environment, and how do different learning styles come into play? How is this different from learning in person?
- How can instructors change their pedagogical strategies to teach people in a completely cyber-enabled environment to use instrumentation in way that is safe for the instrument and time-efficient for the bandwidth that is available for collaborating institutions?

- How might having reusable cybertools and components at all levels (instrument, user, institution, and center) contribute to learning?

Resources to support the pursuit of these issues might include appropriate and adequate technical staff support; a cyber learning commons; instructional designers; formative and summative assessment, as well as formal and informal assessment; and aligning training with future workforce needs in research, industry, and the broader community..

In summary, given current opportunities and constraints, the prospects for making instrument resources and the data produced by them easily and widely available through cyber-enabling will 1) promote better and new kinds of scientific research collaborations; 2) maximize the utility of instruments; and 3) provide novel, high-impact education and training opportunities. Reusable, widely available methods for remote access to instruments will also permit the broader use of unique analytical resources. What is more, by facilitating consultation and distributed expertise, cyber-enabling will encourage more rapid evolution and improvement of novel analytical techniques.

The workshop participants strongly support a program to develop reusable approaches to cyber-enabling instruments, and to promote its widest possible application. The full extent of the program's impact will be clear in the myriad uses to which a generation of widely available instruments will be put.

Glossary of Terms Relevant to Cyber-Enabled Instrumentation

Archive—The structured storage of data in format for future retrieval.

Assessment—Collecting, distilling, and analyzing the performance of people, equipment, facilities, remote systems, or designated tasks through quantitative and qualitative methods; also includes gauging changes in attitudes, behavior, effect, etc., in particular contexts.

Bandwidth—The maximum amount of data that can be transferred over a communications channel in a unit of time. Usable bandwidth is the capacity to carry data and information at certain rates for an intended purpose or use.

Bottleneck—A hindrance to productivity due to limitations in human expertise and system capacity.

Cloud computing—Virtual computing on demand using non-geographically anchored computing resources. Often associated with commercial services.

Cloud storage—Data storage in a noncentralized grid at multiple physical sites that should include multiple redundancy.

Collection—A set of data that have been categorized for future search and retrieval (see *taxonomy*).

Cyber-based—A task primarily carried out on a network, sometimes remotely.

Cyber-enabled—A task that could or already does exist without a computer network, but that has been adapted and/or enhanced through the use of a network.

Cyber environment—A cyber-based ecosystem for storing, managing, and analyzing data from remote instruments or for facilitating cyber-mediated human interaction.

Cyberinfrastructure—The coordinated aggregate of software, hardware, and other technologies, as well as human expertise, required to support current and future discoveries (NSF).

Cyber security—Protecting and/or monitoring of computer infrastructure in terms of intrusion prevention, malware detection and removal, etc., as well as guaranteeing the integrity and veracity of collected data through the use of authentication methodologies such as digital signatures. This can also include aspects of controlled access to data and resources for users by various means.

Cyber terrorism—Deliberate attempts by malicious entities to disrupt or incapacitate network/computational infrastructure as a force for political change and/or financial gain; an extreme form of cybercrime.

Cyber tool—A stand-alone or downloadable computer application that relies in large part on a network to facilitate a task.

Data acquisition—Starting with a prepared sample, the physical measurement taken, recorded, and stored in raw format for later processing.

Data acquisition framework – An extensive library and device I/O approach designed to support remote data acquisition, control, and presentation of device data. Elements of this framework include the instrument (driver), parsing data strings from instruments into individual data elements (i.e., individual readings), calibration, event detection (optional), automated first-level data quality control, and archive.

Data visualization—Display of raw or processed data in a form that facilitates interpretation.

Digital library—A structured collection of data stored in a digital format.

Distributed resources—Resources located at multiple sites.

Evaluation—Using assessments to judge how well program or project outcomes, outputs, or deliverables match program or project goals or objectives.

Format—The way in which data are digitized or expressed.

Global—That which is inclusive and not exclusive of political, social, economic, and cultural groups and regions.

Grid—The physical and virtual network of all resources available, including instrumentation, data storage, processing, and data transfer that allows for cyber-enabled activities; all aspects of cyberspace with the exception of humans.

Grid management—Operations required to maintain and develop the grid.

Interface—A specification of the interaction between disparate entities. *Instrument interface* is a GUI (graphical user interface) that provides user control of sensors and actuators for physical measurements. *Module interface* provides encapsulated and reusable functionality with defined inputs and outputs. *Human instrument interface* addresses psycho-audio-visual issues to efficiently facilitate user and instrument tasks. *Web services* specify programmatic interfaces to instruments and facility services/resources that support workflows that are secure and empower multiple user communities.

Interoperability—The ability to integrate a single system with multiple platforms.

Metadata—Data having multiple validated searchable tags that can be utilized for future data retrieval.

Mining—Searching a data repository or archive to obtain information or knowledge.

Modeling—Describing mathematically a physical, social, or biological system so that it can be mathematically or algorithmically formulated.

Parallel processing—Executing a computational problem by using multiple computers to solve portions of the problem. Typically programs have to be written (or rewritten) specifically for parallel processing.

Program—An organized list of instructions that, when executed, causes the computer to behave in a predetermined manner.

Programmability—The capability within hardware and software to change; to accept a new set of instructions that alter its behavior. Programmability generally refers to program logic (business rules), but it also refers to designing the user interface, which includes the choices of menus, buttons and dialogs.

Remote access—An umbrella term that encompasses remote observation, operation and status monitoring of computer-controlled equipment. *Remote observation* is the term for visualization of instrumentation and data. *Remote control* is the ability to operate instrumentation from a remote location. *Remote monitoring* is the term for the ability to monitor experimental status.

Remote control protocols – A set of rules or procedures for transmitting data between electronic devices, such as computers and instruments.

Repository—The physical location of stored information.

Resource scheduling—A combination of reserving human and device calendars.

Resource sharing—Resources shared by multiple groups.

Scalability—The ability to expand or contract without a fundamental change in the infrastructure.

Scheduling—The process of deciding how to commit resources between a variety of tasks, such as the priority with which resources are allocated, or the order in which I/O requests are submitted to a device.

Sequential process—A program or algorithm that expresses actions occurring one after another. The algorithm may be inherently sequential (non-parallelizable) or capable of execution on a parallel computer.

Serial process—A process that occurs sequentially using the results of a set of operations, in order, each which must be known before the next action is taken.

Service-oriented architecture (SOA)—Creation of software as a set of components that interact through standard protocols. SOA means services can move, be distributed, be replaced by similar

services, and/or be written in different languages. SOA is current best practice for flexible systems and for supporting software reuse.

Sharing—Making resources and data available to a community.

Simulation—The imitation of some real thing, state of affairs, or process. It can be used in many contexts, such as simulations of technology for performance optimization, safety engineering, testing, training, and education.

Social computing—An intersection of computer science and social science. Social computing often focuses on the social impacts of new computing modes such as the creation of social networks and virtual communities.

Standard data formats and I/O protocols—A proprietary file format and the rules that specify how data contained therein are read or written from/to the client device.

Storage—The process of physically locating data.

Sustainable CI—Sustainable cyberinfrastructure, operationally and/or environmentally. *Operational sustainability* is defined as adequate funding for instruments, maintenance, access, and consumables or otherwise meeting users' needs through transparent processes. *Environmental sustainability* is defined as achieving carbon neutrality and approaching a zero carbon footprint.

Tagging—1. Community control of annotation of content such as associating keywords or names with content (images, protocols, courses, instruments, and other forms of data) that don't show up in data itself. 2. Setting privileges for specific user access to data and resources. 3. Embedding metadata—for example, ADA access (disabled users). 4. (*Specific to information science*) Adding identifying data (metadata) to other data (usually nontextural) so that it can be retrieved/consolidated for analysis. 5. (*General*) attaching informational objects to physical or information items in order to facilitate identification, access, and retrieval. Examples include chemical taggants, pet chipping, and blog tagging.

Taxonomy—The method used to categorize data for future search and retrieval.

Telepresence—A set of technologies that allows a person to participate in a process at a location other than his or her physical one. Observing or operating an instrument over a network. The use of networking or audio/video conferencing applications by a remote participant.

Throughput—A measure of service quality for user productivity, network health, and computation time.

Use—Refers to applicability of remote access for research and training for various communities (K-12 through graduate level, industry).

User —A person or group who uses an instrument or facility. A client, a consumer of data or customer of service.

Utilization—The ratio of service usages and availability.

Validation—The multistep process to ensure the integrity of data.

Virtualization—An emulation of an entity by a different entity. Examples include virtual environments (such as SGI caves), virtual instruments, or virtual computing (such as Xen of VMware). Virtualization can also include virtual operating systems, virtual computing, virtual networks, and virtual environments.

Virtual community—A group of people who primarily interact via cyber-mediated tools rather than face to face for social, professional, educational, or other purposes**.**

Virtual experiment—An experiment such as a computer simulation that simulates most aspects of the physical environment.

Virtual instrument—Simulating a physical instrument through computer code and data acquisition. Although *virtual instrument* is sometimes used to refer to a remotely operated instrument, the latter is correctly referred to as a "cyber-enabled instrument."

Virtual laboratory—A laboratory environment (typically on a computer platform) that allows control of laboratory elements from a different physical location

Virtual organization—A group of people and/or facilities connected via a network that interact around shared tasks, often with real-time connections between collaborators and with cyber-mediated tools. A group of individuals whose members and resources may be dispersed geographically, but who function as a coherent unit through the use of cyberinfrastructure (NSF). A virtual community with an organizational structure.