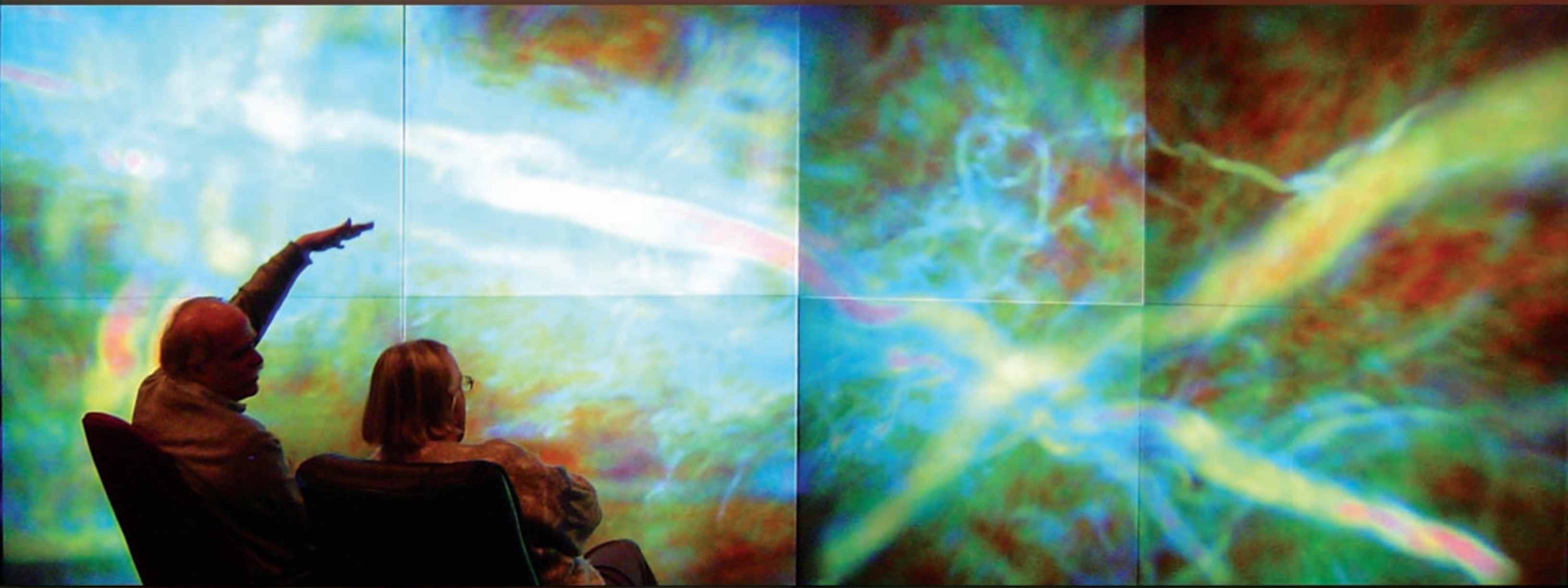
Beyond Being There:



A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations

Final Report from Workshops on Building Effective Virtual Organizations

May 2008

This work was supported by the National Science Foundation under Award Nos. 0751539 and 0816932. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

> “If as it is said to be not unlikely in the near future, the principle of sight is applied to the telephone as well as that of sound, earth will be in truth a paradise, and distance will lose its enchantment by being abolished altogether.”

— ARTHUR MEE, *THE STRAND MAGAZINE*, 1898

The title of this report and the above quotation are borrowed from the seminal paper “Beyond being there.” (James Hollan and Scott Stornetta, in *Proceedings of the SIGCHI conference on human factors in computing systems* (ACM Press, Monterey, CA, 1992), 119–125).

Cover Caption: At the Laboratory for Computational Science and Engineering at the University of Minnesota’s Digital Technology Center, scientists view a movie of a simulation conducted at the Pittsburgh Supercomputing Center. High-bandwidth connections between supercomputers and visualization facilities help researchers explore computationally complex phenomena, such as the fluid dynamics pictured here.

Beyond Being There:

A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organizations

ACKNOWLEDGMENTS

Prepared by

Jonathon Cummings
Duke University

Thomas Finholt
University of Michigan

Ian Foster
Argonne National Laboratory and University of Chicago

Carl Kesselman
University of Southern California/ISI

Katherine A. Lawrence
Scholar at Large

Rapporteur

Diana Rhoten
National Science Foundation

The workshop participants would like to acknowledge the tremendous help and assistance provided by Dr. Diana Rhoten, NSF Program Director. She served as an invaluable resource to the workshop, and we would not have been able to accomplish our task without her unwavering support.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

A virtual organization (VO) is a group of individuals whose members and resources may be dispersed geographically and institutionally, yet who function as a coherent unit through the use of cyberinfrastructure (CI). A VO is typically enabled by, and provides shared and often real-time access to, centralized or distributed resources, such as community-specific tools, applications, data, and sensors, and experimental operations. A VO may be known as or composed of systems known as collaboratories, e-Science or e-Research, distributed workgroups or virtual teams, virtual environments, and online communities. VOs enable system-level science, facilitate access to resources, enhance problem-solving processes, and are a key to national economic and scientific competitiveness. The time is right for taking a more crosscutting, multidisciplinary approach to understanding the basic organizational abstractions, communication models, trust mechanisms, and technology infrastructure required to form and operate effective VOs across a broad range of target domains.

This report is based primarily on a workshop involving 42 people from academia and industry. The goal of the workshop was to share systematic knowledge about the components, characteristics, practices, and transformative impact of effective VOs; identify topics for future research that will inform the ongoing design, development, and analysis of VOs for science and engineering research and education; and create a new cross-disciplinary VO research community to conduct research across a range of important topics. A subsequent workshop brought together more than 200 practitioners and VO researchers to discuss how to build effective virtual organizations, and some of the material from that workshop is represented here.

Current knowledge and practice of VOs is substantial but leaves many avenues for further research. In particular, scientific research communities offer an excellent sandbox in which to study the issues associated with VOs, and they already have provided useful examples. Some of the knowledge we are extracting from such organizations is focused on the social aspects of collaboration, but unfortunately prior knowledge of collaboration in non-VO settings does not always translate to these new venues. Some topics that offer a fruitful starting point include collaboration, emergent organizations, coordination, organizational trust, shared mental models, and knowledge sharing. Further study of technological issues is critical and should draw on technologies used in a variety of virtual settings.

Workshop participants identified a number of research challenges going forward: definitions of VOs, frameworks for comparison, lifecycles, diversity, impacts of research on implementation, technology for knowledge and data sharing, collaboration within and across disciplines, human interaction, scaling, motivation and rewards, governance, and metrics and assessment. Certain development challenges also exist, including the tension between customization and shared infrastructure as well as the deployment, maintenance, and support of infrastructure.

The report concludes with a set of recommendations for how to move forward:

- Encourage cross-disciplinary studies involving both technologists and social scientists working with domain-centered VOs.
- Combine knowledge from multiple studies to present a framework that can inform further VO research and practice.
- Develop a checklist of necessary VOs features—technological, social, organizational, and so on—to ensure that new VOs start off on the right track.
- Design instrumentation, metrics, and evaluation as part of a VO from the beginning rather than adding measurements systems postmortem.
- Support human capital development around VOs.
- Investigate whether technological and organizational factors that support effective virtualization can be standardized or provided as commoditized infrastructure.
- Offer awards for supporting community services at all levels, including the development of new scientific applications, operation of technology infrastructures, and ongoing maintenance of these services.
- Identify incentives and offer rewards for "metacontributors" to VOs—the people who build or reorganize features to make it easier for others.
- Support the development of hardened common tools and protocols for sharing knowledge and data.
- Create proposal funding models that support the use and reuse of VO infrastructures.
- Encourage universities to support VOs with substantial, complementary investments.
- Establish cross-directorate funding opportunities that could more appropriately evaluate and support projects uniting social scientists, computer scientists, and domain scientists.

2. INTRODUCTION: WHY VIRTUAL ORGANIZATIONS MATTER

Virtual organizations (VOs) are a fast-growing phenomenon in all work settings. In many different contexts, people are finding that their goals can be met only by collaborating or coordinating with others located far away. This phenomenon is particularly true for researchers in science and engineering, as the knowledge and capabilities necessary to take scientific understanding to a new level are inevitably dispersed. By organizing their resources with the help of coordinated computer, information, and communication technologies and human infrastructure, scientists and engineers can work together in environments that allow scientific integration, greater access, efficient problem solving, and competitive advantage. This report explores what we know about this relatively new organizational form and identifies future directions for research and development that will enhance our capacity to use VOs to their fullest advantage.

2.1. *WHAT IS A VIRTUAL ORGANIZATION?*

A VO is a group of individuals whose members and resources may be dispersed geographically and institutionally, yet who function as a coherent unit through the use of cyberinfrastructure (CI). A VO is typically enabled by, and provides shared and often real-time access to, centralized or distributed resources, such as community-specific tools, applications, data, and sensors, and experimental operations. Quite often, these resources use high-performance computing (HPC) as a core capability. The term VO can encompass, at least in part, systems known by other names such as collaboratories [105], e-Science or e-Research [47], distributed workgroups or virtual teams [76], virtual environments, and online communities [89].

VOs include a broad range of operational modalities [79]: they can be formal or informal, planned or unplanned, transient or long lived. They may involve, for example, informal exchanges, international scientific collaborations, rapid business innovation processes, or disaster response teams. Most VOs, however, share several common traits. They are—

- *Distributed across space*, with participants spanning locales and institutions;
- *Distributed across time*, with asynchronous as well as synchronous interactions;
- *Dynamic structures and processes* at every stage of their lifecycle, from initiation to termination;
- *Computationally enabled*, via collaboration support systems including e-mail, teleconferencing, telepresence, awareness, social computing, and group information management tools; and,
- *Computationally enhanced* with simulations, databases, and analytic services that interact with human participants and are integral to the operation of the organization.

The recent blossoming of CI and of Internet technologies more generally has put VOs within the reach of most people, enabling both the support of existing communities through technology and the emergence of brand new communities.

VOs enable and are enabled by technologically mediated collaboration, and the relationship between VOs and technology should be discussed from two perspectives: (1) how information technologies are incorporated into, and potentially shape, VO processes and procedures, and (2) how VO characteristics place demands on information technology, and ultimately, how they may shape the evolution of that technology.

In the context of the national scientific research agenda, the first perspective is driven by scientific collaborations organized as VOs. Scientific collaborations are often distinguished from other types of VOs by a focus on shared computational infrastructure, data, and software and simulation as fundamental aspects of the organizational structure and operation. Examples include collaborations formed to create and operate high-energy physics experiments, to analyze data from National Aeronautics and Space Administration (NASA) space missions, or to share data from earthquake engineering experiments [55].

To be sure, VOs in other sectors may make intensive use of computational tools. In the case of scientific collaborations, however, it is frequently the simulation, computational resource, or data set that drives the formation of the collaboration. A consequence is that information technology plays an especially critical role in the formation and operation of VOs for science and engineering, as opposed to VOs established to link communities for other—often nonscientific—purposes.

In these settings, computer science researchers, domain scientists, and engineers have focused on building new technologies, such as shared virtual spaces or new data storage systems to meet the organizational requirements of their distributed collaborations. DeSanctis and Monge [26] observe that technology, organizational structure, and

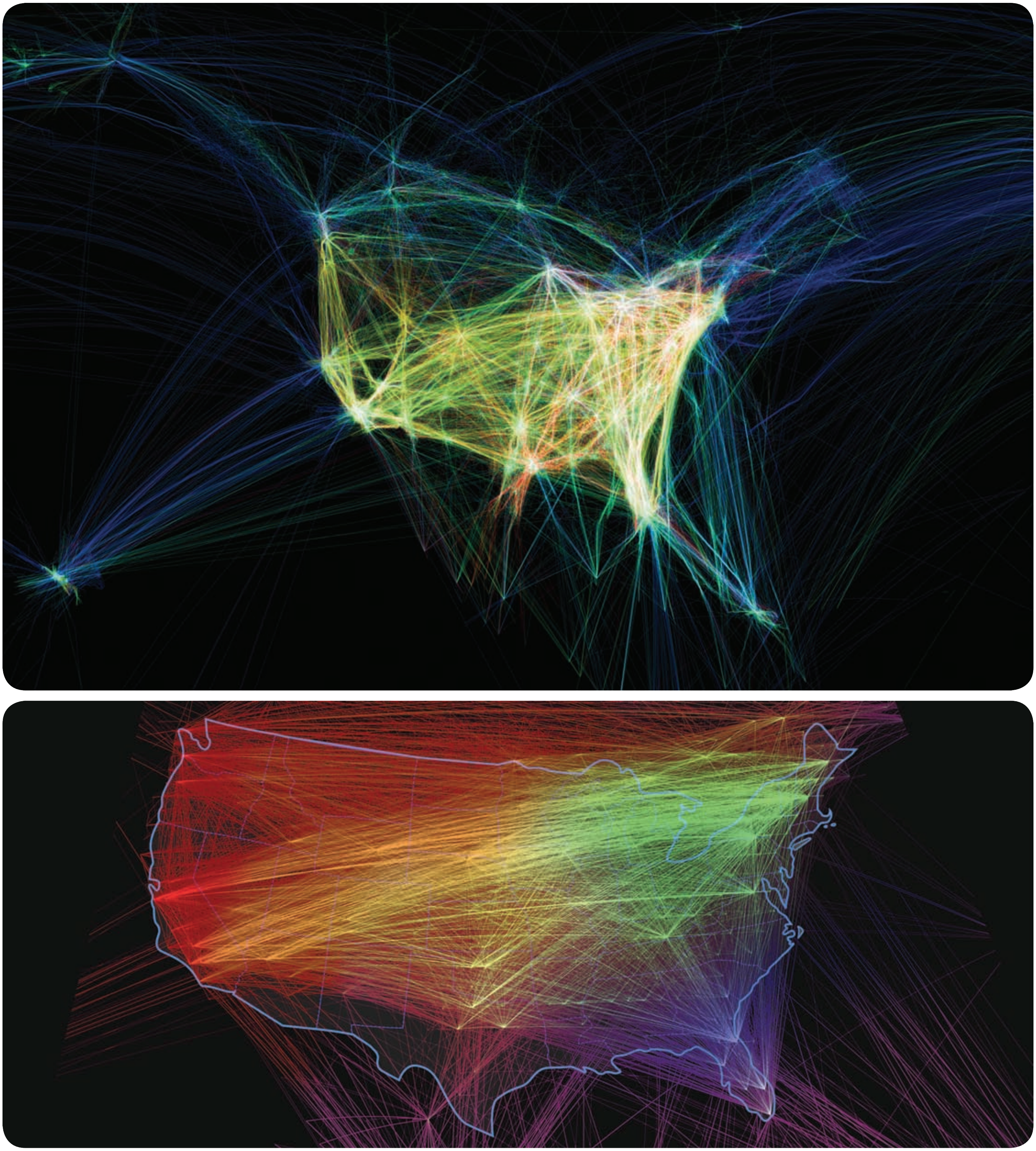


Exhibit 1. Like air travel, the Internet connects people separated by geographic distance. The top image shows 24 hours of air traffic over North America; the bottom image shows a map of the Internet router connections across North America in 2007.

communication patterns are all tightly coupled. However, typical analyses of the operations and impacts of VOs have limited the technological considerations to *electronic communication among people*. Recognizing the contributions that other technologies make to scientific collaborations, researchers studying CI have made this connection between technology, structure, and communication more explicit, investigating new types of information technology infrastructure and social interaction.

At one end of the new technology continuum is the grid, which was created to enable coordinated resource sharing and coordinated problem solving in dynamic, multi-institutional VOs [36]. On the other end, the formation of VOs may be considerably less formal and consequently more responsive to, and influenced by, emerging Web-based technologies. MySpace®, Facebook®, Flickr™, YouTube™, and Second Life® have changed how people congregate, collaborate, and communicate. At this end of the spectrum, VOs may be more like "containers" rather than "vehicles" of collaboration in that they are not necessarily driven by common goals or comparable inputs. Nevertheless, VOs of this type may accumulate the results of many seemingly uncoordinated individual actions, creating a whole that becomes an integrated collection. Examples of such extremely distributed collaborations include NASA Clickworkers and Carnegie Mellon's ESP Game [102], in which distributed collaborators contribute effort but have no intellectual impact on the project. The vast space between this extreme and the tightly integrated, purposeful VO needs to be explored and defined to understand VOs in a coherent and meaningful way.

The second perspective begins with the nature of the collaborators to subsequently identify appropriate technological needs. The precursors of VOs as collaborative structures have been studied within computer science, information science, organizational theory, psychology, sociology, and business and management. For example, work on scientific collaboratories [75] has investigated how distributed scientific communities can more effectively share research results; dynamic business process modeling has produced techniques by which rapidly changing processes can be managed across organizations [26]; work on the grid has produced infrastructure for computationally empowered and dynamic VOs [36]; research in disaster response has investigated mechanisms by which ad hoc organizations form and operate in the absence of any underlying organizational control [67]; studies of social networks have provided insights into how communities scale [74]; and research on distributed collaboration has identified important coordination activities for dealing with distance [23]. Today, research on VOs is growing, spawning conferences on "virtuality" and virtualization, working groups, and the electronic *Journal of Organizational Virtualness*.

Despite this large body of work, distributed collaboration remains unfamiliar to many, and also too difficult in many regards. Consequently and unfortunately, many VOs fail to leverage existing knowledge about technological systems and social dynamics as they form. For example, scientific VOs are often constructed in a one-off manner—organized around the requirements of specific user communities and the specialized technology to enable their operation. As a result, these organizations do not have the ability to exploit the organizational abstractions and understanding of social processes that have been developed by other research communities, in particular from other types of VOs outside the realm of science. Thus, we find that the construction and operation of VOs is limited by either the lack of technology or the lack of organizational understanding that could be gained from more systematic codifying and sharing of information. Further pursuit of more comprehensive knowledge of VOs could reduce redundancies and the costs of establishing and operating VOs.

In addition, few studies integrate the distinct social, organizational, and infrastructure dimensions of dynamic distributed collaborations. In the remainder of this section, we offer examples of why VOs matter. We assert that VOs enable system-level science, facilitate access to resources, enhance problem-solving processes, and are a key to national economic and scientific competitiveness.

2.2. *AN ENABLER OF SYSTEM-LEVEL SCIENCE*

System-level science is concerned with understanding complex, multidisciplinary, multiphenomena behaviors of large physical, biological, or social systems [35]. For example, global warming is too large an endeavor to coordinate as a single project and consequently requires the network-oriented structure that a VO offers. By its very nature, system-level science relies on extensive CI to support the integration of data and computational functionalities from many sources and to coordinate and consolidate the work of many people. In other words, it requires and is enabled by VOs.

VOs, however, introduce new logistical issues that must be addressed when solving big, thorny problems. In all system-level science pursuits, the participants must reorient their thinking toward working together as a community, which is at a larger scale than that of traditional research projects. When VOs unite people from multiple disciplines, they must find ways to translate vocabularies or standardize terminology to make data and other systems interoperable—even when the disciplines are quite similar. Such VOs may have to cooperate and interact with other, independent VOs, which introduces additional systems into the mix. The infrastructure they build needs to be flexible enough to incorporate new tools and ways of thinking as the questions evolve and to engage multiple stakeholders.

Some examples of VOs that have been created to pursue big questions include the Southern California Earthquake Center (SCEC), headquartered in Los Angeles; the cancer Biomedical Informatics Grid (caBIG™) collaboration operated by the U.S. National Cancer Institute; the Earth System Grid (ESG) collaboration funded by the U.S. Department of Energy; and the international Large Hadron Collider (LHC) collaboration headquartered at CERN in Switzerland (see sidebars 1 through 4 for details). In the case of the LHC collaboration, the project has had to create a VO to build and support the international grid computing infrastructure necessary for their work. This U.S.-based infrastructure, today called the Open Science Grid (OSG), links approximately 80 resources at 50 sites through common software. OSG now supports more scientific fields than just the high-energy physics community and includes some 30,000 processors.

Following these pioneers in earthquake engineering, cancer research, climate research, high-energy physics, and computer science, other communities are now forming VOs to study system-level science. These VOs and others are addressing problems that are too large and complex for any individual or institution to tackle alone. It simply is not possible to assemble at a single location all of the expertise required to design a modern accelerator, understand cancer, or predict the likelihood of future earthquakes. VOs allow humanity to tackle previously intractable problems.

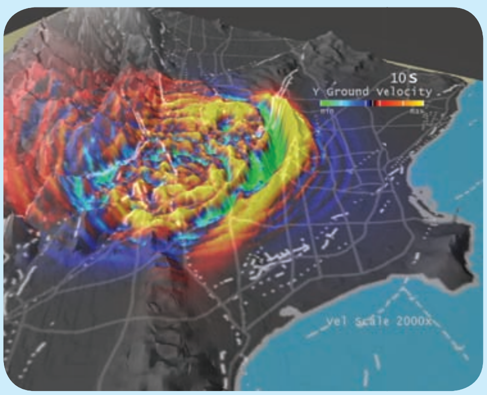


Exhibit 2. Visualizations help earthquake scientists estimate how high-magnitude seismic waves would damage a densely populated region. This simulation of the Puente Hills Fault below Los Angeles, California, calculated velocity components of a magnitude 7.2 earthquake and then displayed the results in a set of animated visualizations (in this image, the Y Ground Velocity).

1. The Southern California Earthquake Center

The Southern California Earthquake Center (SCEC; http://www.scec.org/) was founded in 1991 to better forecast and analyze the consequences of earthquakes, particularly in Southern California. The collaboration involves more than 600 scientists from 16 core institutions and 46 participating institutions. Over the years, they have moved toward doing more of their analytical work through simulations. This work—assessing whether buildings will survive earthquakes—requires the integration of multiple disciplines and the creation of a community modeling environment [53]. Because building failure is catastrophic, they need to trust the data they use in their simulations as well as engender trust in the professional engineers who rely on their analyses. Gathering the data that they use presents challenges for recording, archiving, and attaching metadata carefully and thoroughly. This process is even further complicated by the fact that the occasions to gather data are exactly the moments when their infrastructure is most likely to be compromised and when media and emergency response outlets are most likely to need their input. This presents unique challenges for balancing research desires and disaster responses.

2. The Cancer Biomedical Informatics Grid

The cancer Biomedical Informatics Grid (caBIG™; http://cabig.nci.nih.gov) is sponsored by the U.S. National Cancer Institute (NCI) to support collaborative cancer research by linking researchers, clinicians, and patients. Before the launch of caBIG™ in 2003, cancer researchers worked independently, gathering data that could not be shared across research groups. To address this problem, caBIG™ provides open-source software tools for data collection, management, and analysis, allowing clinicians to gather, share, and analyze data more effectively and efficiently [103]. With these tools, scientists can search diverse resources for specific data sources, process large amounts of heterogeneous data, and coordinate their efforts across institutions [85, 86]. For the pilot phase from 2004 to 2007, the caBIG™ community included more than 50 cancer centers and NCI-supported research projects as well as an assortment of 30 Federal, academic, nonprofit, and industry organizations. The hope is that with improved interoperability and affordable tools, the community can meet the NCI's vision of faster and more effective treatments for cancer in the years to come.

Exhibit 3. The National Cancer Institute offers technology resources, libraries, news, events, and workspaces to members of the cancer Biomedical Informatics Grid through its sponsorship of the caBIG Community Web site.

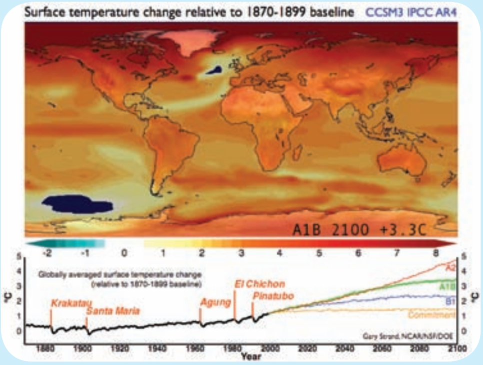


Exhibit 4. This image was produced by combining data from multiple experiments conducted by distributed working groups and based on the Community Climate System Model (CCSM). The CCSM combines a large, sophisticated set of mathematical formulas to accommodate numerous environmental variables. In this case, the model illustrates the global effect of five major volcanic eruptions on surface temperatures.

3. The Earth System Grid

The Earth System Grid (ESG; http://www.earthsystemgrid.org) [12] was established to enable community access to, and analysis of, the large data sets produced by climate simulation models. ESG serves as a gateway to more than 100 terabytes of climate model data and supports more than 6,000 registered users. The project team behind this effort is composed of members from the computer and computational science, climate, data management and analysis, and high-end computing operations communities. The U.S. Department of Energy funded this collaboration to overcome the hurdles associated with making environmental simulation output available to researchers. Previously, accessing and analyzing the vast quantities of data produced by the simulations was cumbersome. To that end, the ESG team has built a system of rotating storage, deep storage archives, middleware, databases, and desktop client applications that alleviate many of the computational difficulties associated with climate analysis.

4. The Large Hadron Collider

The collaborative community that was formed to create the Large Hadron Collider (LHC) and its detectors was born out of necessity. The machines required to study particle physics are larger and more costly than any single nation can support. Thus, the LHC project at CERN, near Geneva, Switzerland, involves thousands of participants from many countries, all of whom will depend on the results of a machine that will have taken nearly 15 years to build and will continue to run for about 20 years. Once it is operational in 2008, scientists will need to access the petabytes of data from anywhere in the world. The project presents logistical and governance issues of how people will share computational resources as well as data, but one of the greatest and most time-consuming challenges is working with new disciplines that have different values and professional languages. Layered on top of this is the necessity of creating unique software that can operate systems that previously did not exist on such a large scale, thereby presenting both social and technical challenges. The process requires learning and adaptation as the project moves forward.

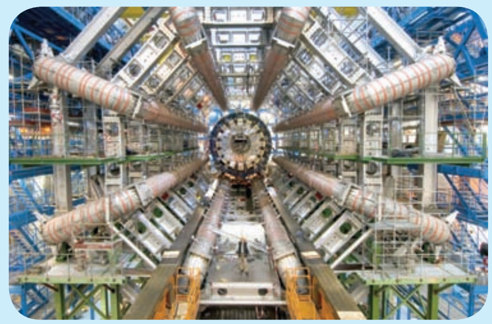


Exhibit 5. ATLAS is a general-purpose particle detector built as part of the Large Hadron Collider (LHC) project. Participants in the ATLAS project include more than 164 universities and laboratories in 35 countries. This detector is 148 feet long, 82 feet wide, and 82 feet high, and weighs about 7,700 tons. For scale, a person is standing near the bottom of the image.

2.3. *A FACILITATOR OF ACCESS*

A second potential benefit of many VOs is the democratization of science. For many years, only researchers at top-tier education institutions or well-funded research centers had access to expensive instruments, high-performance computing, and diverse experts from many domains. Now that the Internet connects institutions, national funding sources aspire to expand access to new communities of researchers and students, crossing existing barriers of institutional size or wealth [5]. In fact, some National Science Foundation (NSF) funding mandates that resources be shared beyond the bounds of the institutions that house them.

VOs open opportunities not only for researchers at smaller colleges and universities but also for students at all levels of the education system. In this vein, new VOs have formed to share data, expertise, and instruments, as well as to build new collaborations for research and education. For example, Linked Environments for Atmospheric Discovery (LEAD; http://portal.leadproject.org) brings together meteorological data, forecast models, and analysis and visualization tools so that anyone can explore weather phenomena, regardless of experience level or understanding of high-performance computing systems. Another project that has enabled access in ways that let community members determine the tools and collaborations that best meet their needs is nanoHUB (http://www.nanohub.org, see sidebar 5) [58], a Web-based resource for research, education, and collaboration in nanotechnology. Likewise, the Biomedical Informatics Research Network (BIRN; http://www.nbirn.net/, see sidebar 6) [29] VO enables emergent, bottom-up collaborations among medical researchers. These projects each serve as a role model of the accessibility forecasted when Internet access began to grow more than a decade ago.

Today, we are told that the modern world is becoming increasingly flat. In practice, significant differences remain in access to expertise and resources, in science as in other fields of human endeavor. The Internet and the distributed social structures that VOs represent are playing a significant role in breaking down these barriers.

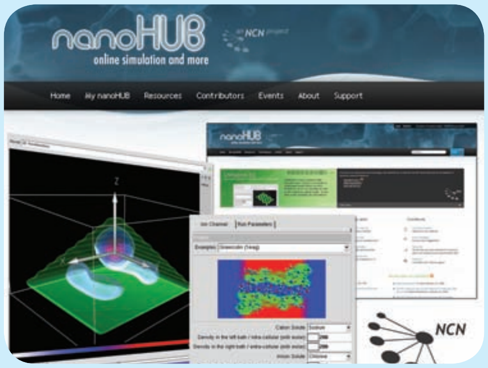


Exhibit 6. The nanoHUB is a rich, Web-based resource for research, education, and collaboration in nanotechnology. The nanoHUB hosts more than 790 resources, including online presentations, learning modules, podcasts, and simulation. The nanoHUB also provides opportunities for distributed collaboration via workspaces, online meetings, and user groups. Resources are used by thousands of users from more than 180 countries around the world.

5. nanoHUB

The nanoHUB project (http://www.nanohub.org) [58] offers Web-based resources to make nanotechnology more accessible to researchers, educators, and students. Although funded by the NSF in the United States, it is used by thousands of people from more than 180 countries around the world. These participants contribute resources such as simulation tools, teaching materials and modules, presentations and podcasts, and animations. When users access these resources, they gain free and easy access to high-performance grids such as TeraGrid or OSG, without downloading or installing any additional software. Even though the site offers an overwhelming variety of options, these resources are ranked according to user reviews and usage statistics—much as Google™ or SlashDot would do—allowing the best materials to rise to the top, and users can customize their portal environment to have ready access to the tools they use most. Thus, in lieu of a hierarchical structure for the materials it offers, the site relies on the community to decide what is best and to tailor the interface to meet individual needs.

6. Biomedical Informatics Research Network

The Biomedical Informatics Research Network (BIRN; http://www.nbirn.net/) [29] is funded by the U.S. National Institutes of Health's National Center for Research Resources to connect laboratories studying a variety of diseases in mice, humans, and non human primates. Because much of the data shared through the network are medical, the data collection and raw data in the VO are strictly governed by the IRB (Institutional Review Board) and HIPAA (Health Insurance Portability and Accountability Act) Federal guidelines, ensuring the privacy of health records. Data must be stripped of all identifying information. All the participants allowed to access the data are members of labs that have been approved according to the Federal rules. Within this externally structured organization, smaller, ad hoc collaborations form among members to conduct data analysis. To support this work, BIRN has created tools for collaboration and data integration layered on top of its extensive CI. Consequently, much of the work structure emerges out of relationships made possible by the VO.

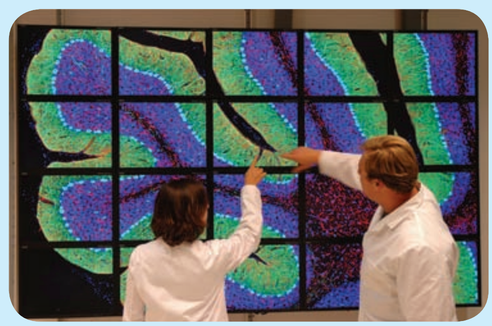


Exhibit 7. The Biomedical Informatics Research Network (BIRN) is fostering large-scale collaborations in biomedical science by utilizing emerging cyberinfrastructure. An essential feature of BIRN is its distributed architecture of shared resources that allows researchers of different disciplines and locales to collaborate on the diagnosis and treatment of disease. Above, brain researchers explore an expansive image of the cerebellum using a computing cluster-driven "biowall."

7. VOs Supporting the Humanities and Social Sciences

The humanities and social sciences have been less evident in these early growth years of VOs, perhaps because the applications are not as obvious or perhaps because computational science is less familiar to these researchers. Nevertheless, some early visionaries have seen the potential of VOs to support research in areas beyond the physical and natural sciences. For example the Humanities, Arts, Science, and Technology Advanced Collaboratory (HASTAC; http://www.hastac.org/) network includes more than 80 institutions, such as universities, supercomputing centers, grid infrastructure groups, institutes, museums, and libraries. This community of researchers, humanists, artists, scientists, and engineers is developing innovative tools that enable education, archiving, and interaction. These tools are made available to those who are considering the implications of history, social issues, and humanistic concerns for the ongoing growth and application of digital technologies.

Exhibit 8. The Humanities, Arts, Science, and Technology Advanced Collaboratory (HASTAC; http://www.hastac.org) is a virtual consortium of humanists, artists, scientists, and engineers from leading researchers and nonprofit research institutions. HASTAC is committed to new forms of distributed collaboration across communities and disciplines fostered by creative uses of technology.

In a similar vein, MATRIX, The Center for Humane, Arts, Letters, and Social Sciences Online (http://www.matrix.msu.edu), housed at Michigan State University, has been leading major initiatives to support the integration of information technologies and the humanities since the mid-1990s. The MATRIX office serves as a hub connecting multiple institutions engaged in interdisciplinary projects, such as building digital collections and developing open-source tools and services for institutions in developing countries. Areas of application have been as diverse as music, speech and audiology, history, education, international studies, and museum studies.

The promise of new technologies is making these initiatives easier to implement, and humanities and social science VOs are becoming more numerous. For example, the Transliteracies Project (http://transliteracies.english.ucsb.edu/category/research-project) connects scholars in humanities, social sciences, and engineering across the University of California system to conduct "Research in the Technological, Social, and Cultural Practices of Online Reading." Likewise, Stanford University's Humanities Research Network (http://www.humanitiesnetwork.org/) provides workspaces for its faculty to conduct collaborative research with colleagues at other institutions. The network offers calendars, wikis, journals, and communication technologies.

Other humanities and social science initiatives have grown internationally. The Alliance of Digital Humanities Organizations (ADHO; http://digitalhumanities.org/) was originally formed to better coordinate two long-standing associations (the Association for Computers in the Humanities and the Association for Literary and Linguistic Computing); today, it connects many more research institutions in Europe and North America. Ultimately, VOs in the humanities and social sciences not only build bridges between disparate disciplines but also expand awareness of the potential of digital media to support research in new fields [1].

2.4. *AN ENHANCER OF PROBLEM-SOLVING PROCESSES*

VOs ideally help people find solutions in a more efficient or cost-effective manner. VOs can make connections and reveal patterns among knowledge and processes that might otherwise remain invisible because technologies capture interactions and events that are more ephemeral in the physical world. When these patterns are better understood, they can be reapplied to new problems.

VOs can involve people who otherwise would be difficult to find or unlikely to engage, harnessing their help toward a collective goal. For example, planetary scientists at NASA created a Web site that enlisted nonscientist volunteers to help identify, measure, and estimate the age of impact craters on Mars. The 80,000 people who participated in the first 10 months of the initiative marked nearly 2 million craters at an average level comparable to an expert. Commercial ventures also use VOs to tap into problem-solving mind power. For example, InnoCentive has built a network of 125,000 people worldwide and connects them to companies, academic institutions, and nonprofit organizations in search of innovative solutions. Those with the best solutions receive cash awards for their efforts. Similarly, the Amazon Mechanical Turk acts as a broker of services and payments between organizations needing people to perform "human intelligence tasks" and the individuals available to do those tasks.

In other cases, VOs are simply a better way to leverage limited funding for research. The TeraGrid—an NSF-funded computational infrastructure—integrates an ever-expanding partnership of resource provider sites that offer supercomputers, storage, and scientific analysis tools free of charge to any researcher in the United States. While these individual machines and the network that connects them are expensive, providing such resources to individual scientists would not only be cost-prohibitive but also intellectually restrictive in a host of research areas. The potential value-added of VOs should not be limited to the realm of cost-efficiency, however. VOs may be especially effective at creating and disseminating knowledge as well as enabling discovery and innovation around many otherwise perplexing scientific challenges.

If the United States can build the technologies that enable this integration and also educate people to work well in these environments, then VO competence becomes a big competitive advantage for U.S. science and industry. Both national funding for scientific research enabled by VOs and for scientific study about VOs are vital to accomplish this goal.

Recognition of these possibilities has already placed pressure on the Federal research portfolio

8. The Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS; http://www.sdss.org/) is mapping one-quarter of the entire sky with the ultimate goal of capturing a three-dimensional picture of the universe. This map will identify distances to celestial objects—such as stars, galaxies, and quasars—and their absolute brightness. With this vast body of information, scientists will be able to answer previously unanswerable questions and test theories of how the universe has changed over time. More than two dozen participating research institutions, plus several museums, have joined together to accomplish this ambitious task of capturing and redistributing the data to anyone who is interested.

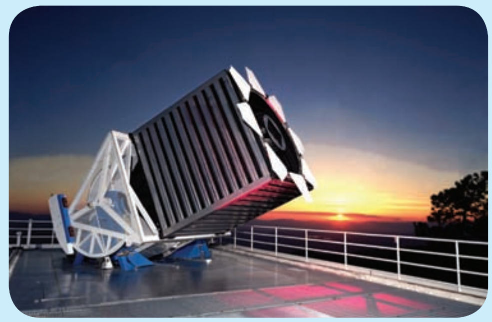


Exhibit 9. The Sloan Digital Sky Survey (SDSS) relies on specially designed hardware and computer software to create a three-dimensional map of the sky. The multifunction telescope, located at Apache Point Observatory in New Mexico, captures information about the appearance, composition, and distance of celestial objects such as nebulae in the Orion constellation.

The SDSS relies heavily on new technologies. The camera uses electronics to capture images, and computers provide substantial information processing capacity to channel data to a wide spectrum of astronomical research questions. Microsoft® has provided hardware, software, and interface designs to make the data and tools available for free to researchers, students, educators, and to the public through the SkyServer (http://cas.sdss.org/dr6/en/). This generosity renders the SDSS an open VO that reaches well beyond the boundaries of the participating institutions and the researchers who are formally engaged on the project. A unique interactive workbench is available through which users can create new value-added data sets and share these with their collaborators. The workbench automatically records the user's actions and can provide a transcript on request.

Over the 6 years the survey data have been available, the Web site has received more than 400 million Web hits. The 1 million distinct Internet protocol (IP) addresses in the logs indicate that the usage goes substantially beyond the professional astronomy community, consisting of about 15,000 astronomers around the world. The Web site contains classroom exercises and teacher guides. The projects encourage students to explore new data, appreciate and understand experimental uncertainties, and make their own discoveries.

2.5. *A KEY TO COMPETITIVENESS*

In the knowledge economy, the traditional enterprise is disaggregating [38]. Virtual reaggregation is one way to be effective in this context. A report by the U.S. National Science Board characterizes this issue:

In recent decades, the speed, complexity, and multidisciplinary nature of scientific research, coupled with the increased relevance of science for industrial technology development and the demands of a globally competitive environment, have increased the importance of technology linkages for innovation and long-term competitiveness [15] . . . the current environment has encouraged an innovation system increasingly characterized by networking and feedback among R&D performers, technology users, and their suppliers and across industries and national boundaries [22, 104]. [73, 44–36]

in computing. A greater potential for national competitiveness relies on larger, longer-term, multidisciplinary projects that will transform the use and application of computer systems that support distance collaboration through shared, virtual spaces. Such VOs offer the chance to synergize the small research projects happening across different institutions by leveraging multiple small venues into larger units of research.

Beyond computing-focused initiatives, other scientific research projects that operate as VOs have the potential to advance a national research agenda that, in turn, produces knowledge capable of supporting that national economy and infrastructure more generally. For example, the Geosciences Network (GEON; http://www.geongrid.org/) integrates heterogeneous data and people through a VO infrastructure. GEON

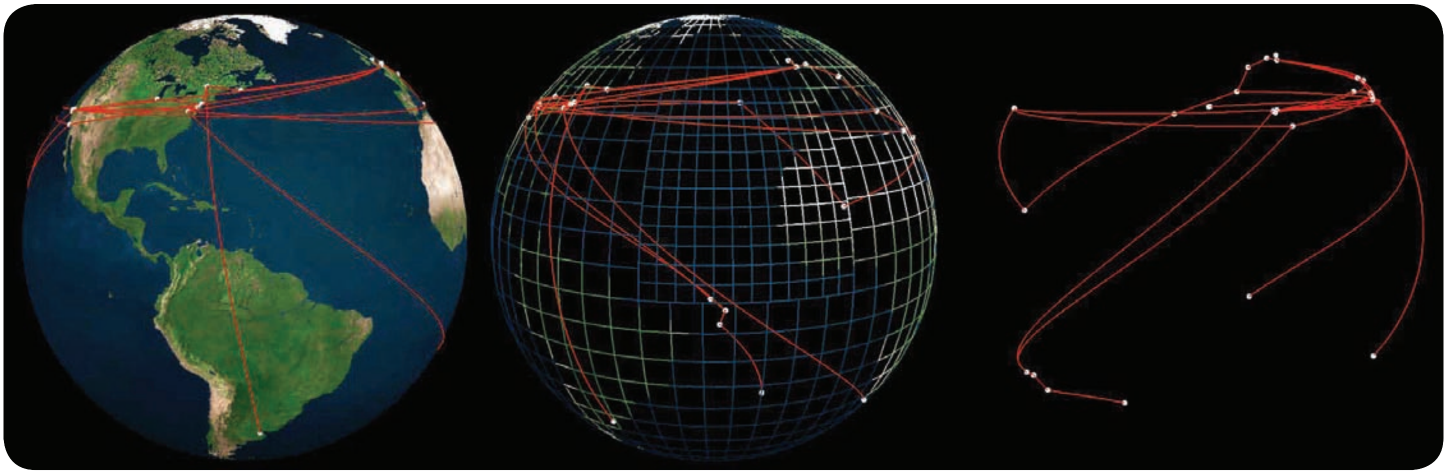


Exhibit 10. This series of images, from the museum exhibit "Ride the Byte," was designed by ART+COM to convey the pathways followed by data packages (bytes) across the Internet. The Internet is a global infrastructure of networked computers. In a global knowledge economy, national competitiveness depends on the speed of data transmission allowed by this infrastructure.

has involved social scientists who observed its development and sought to codify the lessons learned. These lessons about using and sharing data, as well as how to work across multiple, disparate sites, should transfer to both research and industry communities that seek to use information more efficiently and effectively.

9. The Cyber-Enabled Discovery and Innovation (CDI) Program

The Cyber-Enabled Discovery and Innovation (CDI; http://www.nsf.gov/crssprgm/cdi/) program sponsored by the NSF involves the participation of all NSF directorates and programmatic offices. This bold, five-year program intends to award at least $26 million for the purpose of creating paradigm-shifting, multidisciplinary research outcomes through associated advances in computational science. The research enabled by these computational tools, concepts, methods, models, and algorithms should create new wealth and enhance the national quality of life.

One of CDI's three thematic areas is "Building Virtual Organizations," which will bring people and resources together across institutional, geographic, and cultural boundaries to optimize the processes and products of collaboration. This theme underscores the expectations of the NSF that the grantees create partnerships that may include academic, industry, and international organizations. The other two themes are "From Data to Knowledge," which draws on heterogeneous digital data and data mining, data federation, knowledge extraction and knowledge representation, and visualization to promote discovery, and "Understanding Complexity in Natural, Built, and Social Systems," which proposes virtual experiments or computational simulations to analyze systems that are otherwise difficult or impossible to study in the real world. These three themes are closely interrelated, and projects crossing the themes are anticipated to magnify opportunities for truly transformative results.

2.6. *MOTIVATION FOR THIS REPORT*

With these examples, we suggest that our ability to support dynamic, distributed, and technology-enabled collaborations has become critical to national competitiveness in science, engineering, and economic development. Given the tremendous potential impact of VOs—to paraphrase Dan Atkins, director of the Office of Cyberinfrastructure—with so many technological and scientific opportunities flowing together, now is the "time to exploit the advantages." Only by taking a more crosscutting, multidisciplinary approach can we hope to understand the basic organizational abstractions, communication models, trust mechanisms, and technology infrastructure required to form and operate effective VOs across a broad range of target domains.

To prepare this report, in September 2007, we gathered a workshop involving 42 people from academia and industry (see appendix B for a list of participants). We brought together these thought leaders to share systematic knowledge about the components, characteristics, practices, and transformative impact of effective VOs; identify topics for future research that will inform the ongoing design, development, and analysis of VOs for science and engineering research and education; and create a new cross-disciplinary VO research community prepared to conduct research across a variety of important topics. Our hope is that this knowledge will enhance the practices, processes, and outcomes of VOs across the full range of science and engineering domains, identify technologies that need to be on the radar screen, and perhaps create economies of scale that will support numerous projects in the national science portfolio. Subsequently, in January 2008, we brought together a second group of approximately 200 practitioners and VO researchers in a workshop titled "Building Effective Virtual Organizations." The goal of this workshop was to share existing expertise and build a community of practice. Some key insights from this workshop are included in this report.

Exhibit 11. Breakout groups at the September 2007 Office of Cyberinfrastructure Virtual Organizations workshop identified key issues and questions relevant to future cross-disciplinary research on VOs.

Even as we seek to identify what VOs are and how they operate, it is clear that our definitions and understanding need to be flexible. Organizational boundaries are less and less clear, and the concept of "organization" may not be useful for much longer. We often see diverse networks of people with a range of connections, where anything like an organizational boundary is arbitrary. VOs not only work on a person-to-person basis, but also, at times, on an organization-to-organization or process-to-process basis. Some suggest that grids—mappings between physical infrastructures and dynamic organizational structures—could help define the community or network, particularly as they emerge. With such diverse entities, we need to develop new languages and analytic frames to make our discussions productive and relevant. In the next section, we briefly review what we know about the current state of the practice.

3. STATE OF THE PRACTICE

Today's VOs are developing at a particularly exciting point in the history of technology. In the last few years, older technologies have improved in quality and sophistication, ushering in new ways of crossing time and space to collaborate. As people increasingly gravitate to these new tools and technological opportunities, we need to revisit traditional notions of organization and collaboration.

The September workshop began with a consideration of a variation on the CSCW (Computer-Supported Cooperative Work) matrix representing the intersection of same versus different time with same versus different geographic place (see table 1). (For further examples and discussions of this matrix, see Bullen and Bennett [16]; McGrath and Hollingshead [71]) Because some of the latest technologies were absent, the diagram sparked a lively debate.

Table 1: Collaboration tools for synchronous and asynchronous interaction

Geographic Place	Time: Same	Time: Different
Same	*People:* Physical meetings *Information:* Print-on-paper, books, journals *Facilities/instruments:* Hands-on labs, shop, studios	*People:* Shared notebook *Information:* Library reserves *Facilities/instruments:* Time-shared labs, shops, studios
Different	*People:* Audio/Visual Conference *Information:* Web search *Facilities/instruments:* Online, real-time instruments	*People:* E-mail *Information:* Knowbots *Facilities/instruments:* Autonomous instruments, session objects

Source: Daniel Atkins' version of [52].

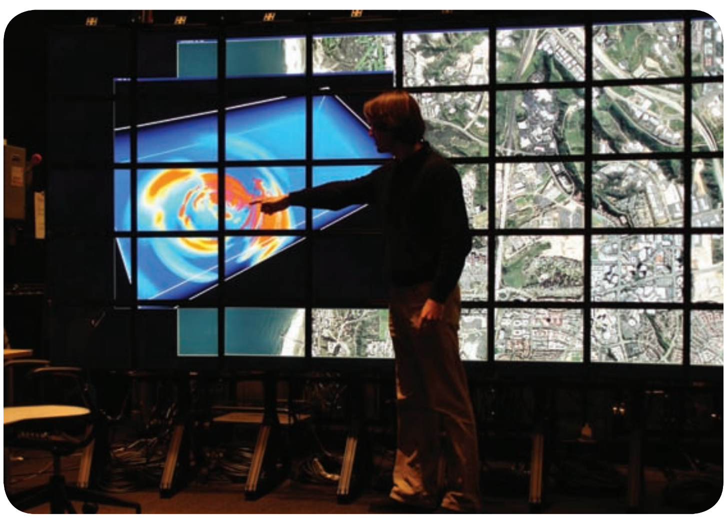


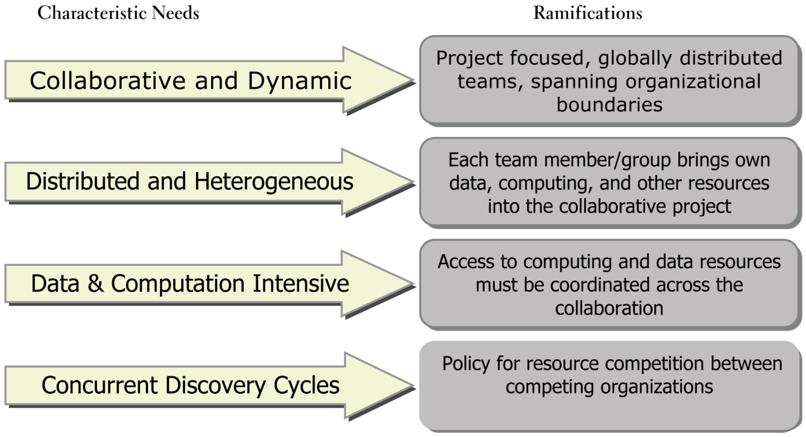
Exhibit 12. Wall-size displays made up of multiple, tiled computer monitors—such as this one at Calit2 (California Institute for Telecommunications and Information Technology) at the University of California, San Diego—allow scientists to interact with data visualizations at a level of detail previously unavailable. The display above shows the TeraShake earthquake simulation, created by integrating data from different sites and disciplines.

September workshop participants noted that the diagram does not include any Web 2.0 technologies nor does it capture the blending that happens as people traverse across and between quadrants. The diagram does not take into account the effectiveness of tools for the different circumstances that VOs might encounter. The vigorous discussion that ensued among participants offered immediate evidence of the need for further research on VOs as well as opportunities to disseminate existing research. Fortunately, despite a scarcity of research on VOs, our general understanding of the demands, social issues, and technology affecting VOs provides a good starting point for exploring their potential. The remainder of this section briefly reviews the state of the practice.

3.1. *VIRTUAL ORGANIZATIONS AND SCIENCE*

Scientific research communities offer an excellent sandbox in which to study the issues associated with VOs. As we described above, science projects today wrestle with answering system-level questions, sharing data remotely, providing access to limited resources, using funds efficiently, and supporting education and outreach agendas. Figure 1 illustrates some specific needs that drive collaboration among distributed researchers in science (and nonscience) domains. The benefit to researchers studying scientific VOs is that the high-end computers and networks used by scientists today offer a preview of the types of information technology that will be widespread in a decade or two. Typically, these scientists are aware of the cutting-edge quality of their large-scale projects and are willing to let other people study them.

Figure 1: Collaboration drivers



It is clear that scientists have the potential to gain much from participating in a VO, but establishing a VO is time-consuming and difficult. It is particularly illuminating to study scientists in this new context, to understand why they chose to form VOs, what motivated the participation, and how they have fared. The Science of Collaboratories (SoC; http://www.scienceofcollaboratories.org/) project was a leader in taking up such questions. The mission of SoC focuses specifically on collaboratories, which they define as "an organizational entity that links a community of individuals working at a distance on common problems or tasks that contains electronic tools that support rich and recurring human interaction and provides common access to resources, including information and instrumentation, needed to engage in the problems or tasks." This definition is close to our previous discussion about what constitutes a VO, although collaboratories tend to involve smaller groups than VOs might. By creating the first typology, the SoC project provided one of the first analytic tools to consider distributed collaboration [81]. The typology reveals seven types: four focus on research, including distributed research centers, shared instrumentation collaboratories, product development collaboratories, and community data systems; and three focus on practice, including virtual communities of practice, virtual learning communities, and expert consultation collaboratories.

As examples of the variety of VOs under way, we describe three new environmental science observatory initiatives—each representing an unusual hybrid of the technology-enabled and the technology-focused collaboration—that are in the planning or early implementation stages: the WATer and Environmental Research Systems Network (WATERS Network; http://www.watersnet.org), the National Ecological Observatory Network (NEON; http://www.neoninc.org/), and the Ocean Observatories Initiative (OOI; http://www.joiscience.org/ocean_observing/initiative). All three of the projects have the aim of establishing networks of sensors to accumulate data that will be available to scientists across the country. Eventually, the VO infrastructures that they build should help scientists work together remotely, analyze data with sophisticated tools, and establish education and outreach connections.

The ocean research community has a long tradition of ship-based research collaborations, so their new challenge is predominantly technical: establishing permanent sensing capabilities across the ocean rather than relying on expeditions. The OOI has spent more than 10 years planning the sets of global, regional, and coastal observatories that will be connected with CI for which "the goal is to facilitate direct and immediate interaction with the ocean. The CI must address the issues of observatory resource management, mission command and control, data management and distribution and the meaningful collaboration across a wide range of disciplines." The CI architectural design for the OOI is a relatively recent component, dating back to early 2006. For the OOI, the development of CI is an early step toward developing more sophisticated capabilities that eventually will allow scientists from many locations to work on shared problems.

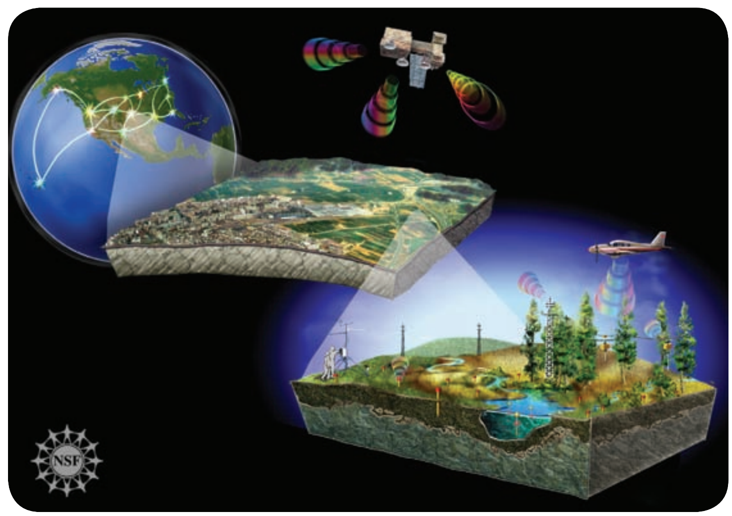


Exhibit 13. This image illustrates the tiered levels of analysis coordinated by the National Ecological Observatory Network (NEON). As a virtual laboratory, NEON will provide nationally networked research, communication, and informatics infrastructure for collaborative, comprehensive, and interdisciplinary measurements and experiments on ecological systems.

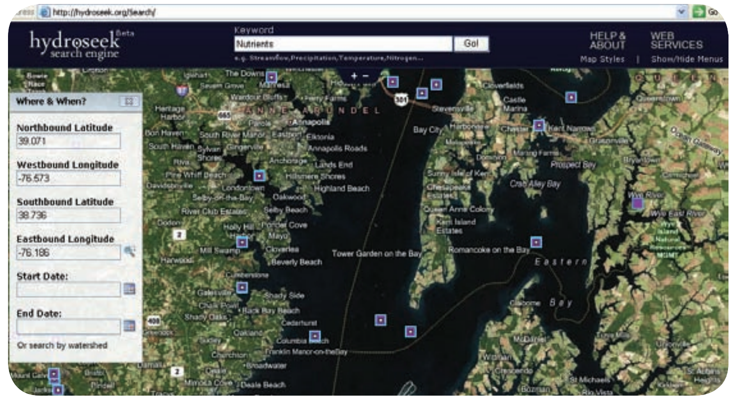


Exhibit 14. The Hydroseek search engine is a Web-based system that allows researchers to search for hydrologic data across multiple data sources and data description systems through one interface. This particular image identifies data stations capturing nutrient data throughout the Chesapeake Bay region in Maryland and Virginia.

In contrast, the ecological and environmental communities have historically worked on single-investigator projects. The participants in the WATERS Network are excited about the potential that sensors will have to generate and integrate data, but because both collaboration and the use of sensors are unfamiliar, they have tremendous difficulty anticipating their future needs and the challenges associated with them. Likewise, the NEON project, which is already planning where it will situate the 20 instrumented core sites across the country, cannot possibly satisfy all the scientists who want a collection of sensors located close to their home base. Simply partitioning the country into ecoclimatic zones presents a substantial task. Making a conversion to system-level science is more than a technical challenge.

While each of these observatory initiatives still have substantial planning and implementation of the VO ahead, early forays by a subset of the WATERS community illustrate how tricky it will be to network its members in a meaningful and effective way. On the technological side, a small team at National Center for Supercomputing Applications (NCSA) has built a prototype gateway to demonstrate the kinds of services that the WATERS CI would offer. They have tried to use this Web venue to keep members informed, but many participants would rather be notified by e-mail than be bothered to learn and monitor a new tool. Unfortunately, as a consequence, these same members complain that they are overwhelmed by too much e-mail. Likewise, the CI project team has experimented with various videoconferencing and collaboration technologies such as the Access Grid, but these technologies are dissatisfying for a variety of reasons, including the difficulty of setting up and learning new technologies and the quality of the transmission.

On the nontechnical side, the WATERS project seeks to integrate members of the environmental engineering community—those who are interested in human-induced impacts on the environment—and the hydrologic science research community—those who are focused on natural processes associated with water resources. While these two communities share some publication venues and professional associations and their members often are educated in the same graduate programs, they pursue fundamentally different types of research questions. How they will bridge these differences remains to be seen.

In these as well as in other examples that go beyond the scope of this report, what VO researchers observe again and again is that these collaborations face a tension between the new modes of interaction that they need to learn versus their existing culture based on face-to-face exchanges. Others note that successful collaborations work, because they have a social relationship to accompany whatever intellectual or instrumental goals they may have [61]. While communities will certainly adopt a new tool that advances their science, a tool may not be enough to motivate a complex, distributed collaboration unless the collaboration is absolutely necessary, as in the case of high-energy physics. In a study of multidisciplinary projects, for example, Cummings and Kiesler [23] found that the more universities that were involved in a project, the less successful the projects were at achieving their self-reported goals such as new ideas, training, and outreach. When considering project coordination, the generation of new tools was the one area in which the participation of multiple institutions was a benefit, suggesting that perhaps distributed collaboration could be a critical element when building new technology. The next section explores in more depth what we know about these behavioral issues that promote or inhibit collaboration.

3.2. *THE SCIENCE OF VIRTUAL ORGANIZATIONS*

Researchers have studied the social aspects of collaboration from a variety of perspectives. But our knowledge of rules of behavior has been developed in settings in which distance collaboration was relatively rare. Those interested in studying human and organizational behavior are now accustomed to hearing arguments about the changing nature of work and collaboration. They recognize the necessity of looking at forms of distributed work and VOs that span geographic and institutional boundaries through the use of information technology. The idea that technology might be able to create a virtual space for interaction fits into a conventional picture of traditional, hierarchical organizations being replaced with dynamic, networked organizational forms.

Recent studies, however, find that these ideas do not automatically translate. Today, researchers face new opportunities for revisiting previously established theories in a new context, and even within a VO context, scholars disagree about what theories hold true. For example, one contingent asserts that physical proximity is important in collaboratories and, correspondingly, that working at a distance can pose challenges with regard to dedication to the project, the building of trust, the allocation of responsibility and leadership, and so forth [56, 78]. A contrasting view is that with better, more transparent tools, more realistic modes of communication, and with the right protocols for selecting collaborators, the seeming need for physical proximity and face-to-face meetings will not be a significant concern [17, 38, 49]. In the future, it is possible that younger people will have been pretrained, in effect, to work at a distance. Such scholarly differences may both be valid, suggesting that perhaps further research needs to examine what features are responsible for these contrasting research findings.

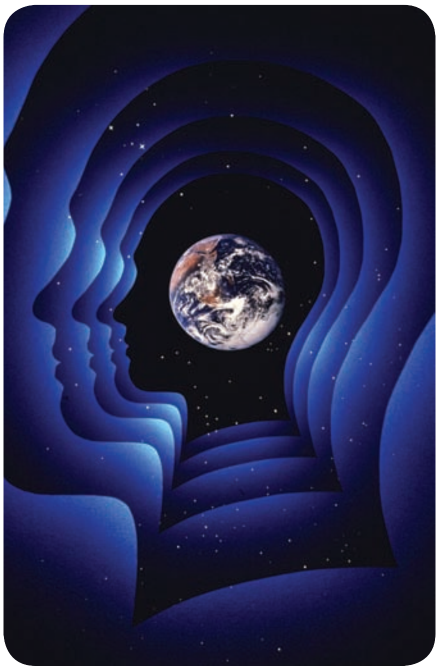


Exhibit 15. A virtual organization is a group of individuals whose members and resources may be dispersed geographically, but who function as a coherent unit through the use of cyberinfrastructure. Focused investments in socio-technical analyses of virtual organizations are necessary to harness their full potential and the promise they offer for discovery and learning.

Most likely, social science researchers will need to revisit topics and concepts that they already have considered. Researchers have studied distributed groups for many years (for a recent review of the research, see Hinds and Kiesler [48]), but today, studies may need to allow greater latitude for multiple definitions of what a VO might be, while making the differences explicit to facilitate meta-analysis. They may need to consider that different levels of analysis are critical to a deeper understanding. New input or contextual variables may need to be considered. For example, whether VO structures are formal or informal and how the boundaries are defined may come into play. The culture and history of a VO may be relevant. Such factors translate into different norms and expectations during the growth and maintenance of a VO, and new process factors might emerge. For example, for many years, the evolution of teams was assumed to follow a path of forming, storming, norming, and performing [99]. Later, new models of punctuated equilibrium suggested that team development is more episodic [3, 41]. Within a VO, the opportunity for asynchronous interaction and changes in composition may surface entirely new processes. Finally, new outcomes may be relevant. Researchers may examine such variables as VO growth, new alliances formed, new forms of knowledge production, and new artifacts alongside such current outcome variables as satisfaction, learning, new capabilities, performance, and productivity.

One clear and current gap in much of the social science research on VOs is that sometimes researchers make the assumption that the technology is more or less the same. Orlikowski and Iacono [84] found that researchers who theorize about information systems do not specify the technology enough and focus on the organizational and human elements too much. Thus, researchers need to unpack the technology more: What are the features of the technology? What is the underlying technological infrastructure? How does it work? The ultimate challenge for researchers is to understand the processes by which people and technology align and coconstruct each other. In this way, we can develop a deep understanding of CI and VOs.

The following sections offer a sampling of additional topics raised by the September workshop participants.

3.2.1. *How Collaboration Happens*

How are VOs created in the first place? This is perhaps the most basic question that we can ask concerning VOs. Initiating and forming a VO should be different from forming a physical organization. The early stages of setting up a VO are also different than maintaining one in terms of the required tools, human capabilities, and skills. Indeed, what compels people to form a VO or join an existing VO does not have a simple answer. Previously, researchers believed that the cultural features of a given discipline—the degree to which members are socialized to work collectively—would explain why individuals would choose to work together

[59], but as the number of fields building VOs grows, disciplinary differences cannot explain all the growth in VOs across varied disciplines. New research on collaboration propensity indicates that collaboration between scientists is more effectively predicted by the nature of the work, such as resource concentration, level of agreement on what constitutes quality research, and the need for and availability of help [13]. If this is true, incentives to form or join a VO may not be easy to manipulate unless the participants have a compelling need for a VO to support their work.

Once formed, VOs are apparently a more complex and heterogeneous form of organization than accounts of traditional organizations, personal networks, or distributed teams provide. (See sidebar 10, The Transition from Informal to Formal VOs.) Scholars have recently made major improvements in our ability to describe virtual teams [42, 76], but the configuration of VOs and virtual teams needs further study. Recent research has found that the "human infrastructure" of the CI that supports many VOs is comprised and supported by a variety of traditional and networked social structures that are constantly shifting and changing [63]. VO boundaries are often unclear, with the result that even one's own membership in VOs is often uncertain and unstable [72]. Likewise, the CI of a VO can include many overlapping networks and can be embedded in others.

Beyond membership and structural issues, VOs are often composed of multiple communities, particularly in early stages of development when both technical and domain expertise are required. New challenges emerge as multidisciplinary teams must find ways to communicate their expertise to each other [62]. This issue is simpler in a collocated situation, where meaning and understanding can be negotiated in face-to-face dialogues. In a distributed environment, VOs rely on mediating artifacts. In the pre-VO era, these mediating artifacts were letters, journal articles, books, and sometimes conferences or classrooms. Nowadays, newer technologies offer richer virtual environments with new artifacts such as video, audio, text chat, and digital documents along with the logs that can trace interactions over time. Even so, research continues to sort out the value and effects of these new technologies (see, for example, Herbsleb, Atkins, Boyer, Handel, and Finholt [46]; Maznevski and Chudoba [70]; Olson, Olson, and Meader [83]).

An additional layer of collaboration is the set of coordination activities used by a VO. Although research has shown that the participation of multiple universities in a research project will negatively affect desired outcomes, certain coordination activities can reduce that impact [24]. For example, transferring knowledge (for example, by cotraining or exchanging graduate students, coauthoring, and giving presentations) and dividing responsibilities significantly mediates the relationship between the number of universities involved and the desired outcomes. Unfortunately, the greater the number of participating institutions, the less likely they will employ such coordination activities.

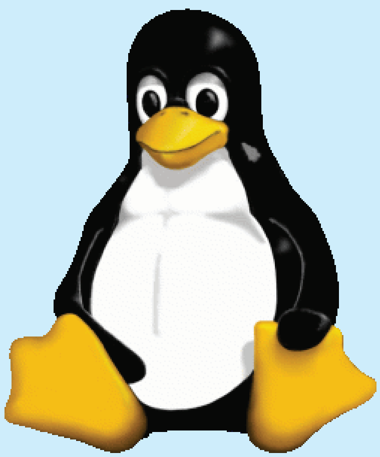


Exhibit 16. The penguin Tux is the unofficial, iconic mascot of the Linux system. Although he was chosen for his fun and friendly appearance, penguin behavior appropriately captures the essence of the massive, informal gathering of like-minded individuals who have together developed Linux. Find Tux among the massive collaboration of penguins.

10. The Transition from Informal to Formal VOs—The Case of Linux and Burning Man

Recent research by Chen and O'Mahony [21] suggests that some VOs—such as the open-source software community that has developed Linux and the artists who started the Burning Man arts festival—thrive when they maintain active debate about and synthesis of competing perspectives concerning ideal ways of organizing. The open-source software movement grew out of a desire to create freely available software code developed by and for the community, and the Linux operating system is one of the largest, most prominent examples of this movement. The Burning Man arts festival began with a handful of artists but grew to an annual, week-long event in the Nevada desert attended by 35,000 people who form a temporary city. (While not a typical VO enabled by technology, the Burning Man community is virtual in that it assembles and disbands in the space of one week each year.) Both these communities transformed from small, informal VOs to large VOs that required substantially more coordination to be successful. Research shows that these communities have come to rely on the flexibility necessary to respect collective and individual interests while establishing formal mechanisms to produce standardization and stability. By allowing the coexistence of these two competing approaches, both communities have successfully balanced the two extremes in their organization. The way these communities and others like them have learned to cope with these opposing tensions could be informative for new and emerging VOs anticipating substantial growth, as would further research to understand how these interests are optimally balanced.

3.2.2. *Emergent Organizations*

The fields of organizational studies and information systems have a healthy literature on the emergent organizations and ad hoc groups that form around temporary or opportunistic circumstances [27, 67, 98]. Some of these groups might be considered to be VOs as well, but most of those that have been studied have been considerably more earthbound. Emergent organizations are often—but not always—self-organizing or grow out of existing social networks. In some cases, such as in disaster response, several established organizations come together and must negotiate their combined organization to link and integrate disparate technologies, procedures, and areas of expertise. For example, disaster response groups, both during and after an event, may include government agencies (e.g., police, fire, the U.S. Geological Survey [USGS]), nonprofits (e.g., Red Cross), national and local commercial businesses (e.g., Wal-Mart®), and community groups (e.g., volunteer groups formed in response to the event). Particular issues in these groups may be extended to VOs, including interoperability, reconciling different goals, shifting compositions, privacy and security, authority, and establishing trust.

September workshop participants raised the issue of legitimate peripheral participation as a topic that applies to both emergent organizations (such as disaster response) and VOs. Legitimate peripheral participation (LPP) [60] denotes how newcomers become members of a community of practice or a collaboration through the process of participating in small, low-risk tasks at the boundaries of the group's task while observing and growing familiar with what more experienced, central members do. These roles can be valuable for the community, but they can be particularly essential for those learning how to be a member of a group. For example, among community listservs, new members are often advised to be "lurkers" to learn the customs and appropriate comments or questions to submit. LPP is relevant to emergent VOs in that, as new members seek to identify their roles within a VO as it forms, they may need to "lurk" at the margins until their appropriate involvement becomes more apparent. To the extent that VOs allow new members to have access without having full privileges (e.g., not having to apply for an allocation to try out supercomputers), LPP could be a fruitful area of research to better understand how members of VOs learn and participate over time, as well as what role LPP might play among members of scientific communities.

3.2.3. *Organizational Trust*

An article in *Harvard Business Review* more than a decade ago [45] suggested that virtual teams cannot build trust. While more recent research suggests otherwise, building trust within a VO certainly takes a long time and, because it is dynamically reevaluated with each interaction, remains fragile [50]. Trust is built on a foundation of interdependence and interaction that builds a sense of shared identity and familiarity [94, 95]. So, for example, when people see others executing their roles competently, predictably, and reliably, that builds trust. Trust in VOs may be different than trust in physical organizations, however, and therefore presents opportunities for considering how trust can be built other than through familiarity. For example, on eBay®, reputations are built through the captured opinions of others. However, even such systems can be manipulated and thus unreliable. Yet, because trust is dynamic, different mechanisms may be necessary at different stages of a relationship; what initiates the connection may be different than what sustains it. People may trust others in different ways, to different degrees. For example, cognitive trust levels might be different than affective trust. Also, people tend to trust those more like them, meanwhile assuming that those not like them are different. Finholt and Birnholtz [32] have shown that differences in professional cultures increased the chances for misunderstanding and mistrust. Overcoming genuine distrust in virtual teams, however, is a subject that remains to be studied.

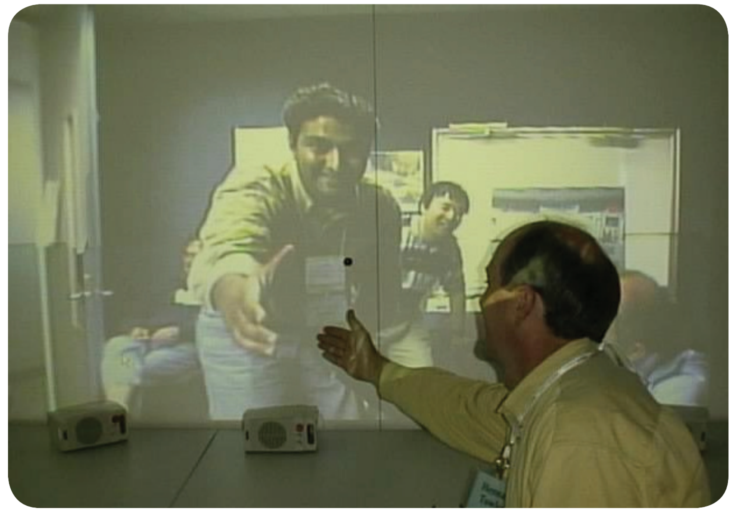


Exhibit 17. This early version of an environment for "virtual handshakes" combined computer-based technologies to help geographically separated collaborators feel as though they were working together in a shared office space.

3.2.4. *Knowledge Sharing*

Research on knowledge sharing investigates how teams share and coordinate their expertise. In particular, transactive memory—the knowledge of who knows what in a group—has been found to improve team performance [6, 65]. Researchers evaluate transactive memory with metrics such as the degree to which people share a conceptualization of who knows what, the accuracy of that impression, and whether the expertise is actually shared [67]. Likewise, shared mental models, which includes commonly held knowledge about tasks, team members, goal, and strategies [18], also improve team coordination [30]. This research has bearing on VOs in that virtual teams may face different challenges as they build transactive memory and shared knowledge among themselves. For example, many things may prompt the withholding of expertise, including selective conservation of attention, low motivation to learn about others, and low interdependence, all of which can be aggravated by the reduced accountability associated with a lack of face-to-face contact [44, 93]. Also, to the extent that some VOs are emergent, less is known about how such ad hoc or emergent groups coalesce to achieve transactive memory and knowledge exchange. VOs present further obstacles in that the typical cues used by collocated groups are unavailable to those in a virtual environment; therefore, media richness may come into play. Appropriate support for learning and knowledge transfer in the context of VOs—for example, identifying whether audio or video support is more critical—would vary according to a person's current level of expertise [43].

3.3. *ENABLING INFRASTRUCTURE AND TECHNOLOGY*

VOs need technology to function and are themselves often concerned with the development of technology. Already, we have no shortage of technologies intended to enable, or available for use in, VOs. Advanced networks between universities and research institutions support and demonstrate state-of-the-art technology using high-definition video conferencing, data sharing, data visualization, and even virtual reality immersion that comes close to "being there." On the lower bandwidth spectrum, tools for course management (Sakai, WebCT™, Blackboard®, Moodle™), multimodal Web conferencing (WebEx™, Microsoft NetMeeting™), and instant messaging or video (Microsoft Instant Messenger™,

11. Second Life—As a Phenomenon and a Subject of Study

Second Life (SL; http://secondlife.com/) is an Internet-based, immersive virtual world in which "residents," represented by self-designed avatars, can interact with other residents and with objects created by themselves and others. These residents can socialize, explore, participate in group activities, and create and sell or trade property and services. SL was launched by Linden Labs in 2003, but growth has skyrocketed since early 2006. SL is also open, at a discount, to academic institutions, some of which host virtual classrooms, library reference desks, education tools, meeting spaces, and museums.

Exhibit 18. Second Life provides an immersive virtual environment for people to collect and collaborate, mediated by their avatars. This presentation on problem-based inquiry is being given to a small but distributed audience as part of a course about the potential of Second Life as a virtual environment for learning.

Some see immersive interfaces such as SL as the culmination of an information technology evolution, whereby technological interfaces have become more oriented toward simplifying and enriching the user experience to better match real-world experience. Typically, first movers are younger, open-minded, or less risk averse, but SL has proved to have great appeal to users of all ages. SL and other immersive environments (some of which are called Massively Multiplayer Online Role-playing Games) have the potential to establish real connections between the virtual and the physical world, bringing in data from sensors, conducting alternative simulations of reality, and allowing users to interact in ways that are possibly more effective than prior collaboration tools. Social scientists see the potential for conducting novel research and virtual laboratory experiments in SL, because the environment allows inexpensive simulation and observation of events that might be impossible in the real world—such as economic or political manipulations or contagious disease transmission—as well as access to larger pools of potential participants [9]. Recently, Bainbridge documented the range of research that uses virtual worlds as a substitute for otherwise difficult real-world environments, but he also points out that virtual worlds are exciting new territories for research in their own right [9].

Exhibit 19. Although the World of Warcraft is structured around fantasy adventures, the massively multiplayer online game offers the tools and experiences of a prototypical virtual organization to its community of more than 10 million subscribers worldwide. Avatars, such as this dwarf hunter and gnome warlock, coordinate activities through communication tools, division of labor, and an economic system that includes production, barter, and auctions.

Recently, Linden Labs made its open-standards client and server software available to other programmers interested in developing add-ons and alternate client interfaces. Their Application Programming Interface (API) allows broader external development, too. These methods of accessing virtual worlds not only open up opportunities for building research and education environments inside SL, but also make it easier for researchers to study the phenomenon of SL—what works, what does not, and why.

AOL Instant Message™, Skype™, Jabber™) offer environments and tools to facilitate synchronous and asynchronous communications. Grid technologies enable the federation and remote use of diverse resources, and grids are in turn supported by "middleware" (see sidebar 15).

As these technologies become more stable and accessible, we see new opportunities to build and use common infrastructure, thus achieving economies of scale and reducing the cost of creating new VOs. There are two basic approaches to achieving this goal. One angle is to plan a predetermined system that thoughtfully integrates resources with top-down notions of how they will be used. The second approach is an emergent model that assembles technology that brings people together and then creates more structure once it is evident how they are optimally using the technology.

An exemplar of the former approach is TeraGrid (http://www.teragrid.org) [20], the NSF-sponsored scientific discovery infrastructure that provides an integrated computational resource through 11 partner sites. TeraGrid connects high-performance computers, data resources, analysis tools, and high-end experimental facilities through high-performance network connections. Although TeraGrid has added new partner sites since its founding in 2001 (and continues to do so), and the resources provided at these sites are heterogeneous, the system is carefully coordinated through the Grid Infrastructure Group, working in partnership with the resource providers. Depending on the nature of the resources, the systems make use of shared middleware while also providing unique resources. In this way, TeraGrid provides consistency that can be exploited for grid computing while also allowing for users with more specialized needs.

Another similar but somewhat less "heavy" infrastructure is the OSG (http://www.opensciencegrid.org) [88], an international consortium of research institutions, funded by the NSF and the Department of Energy (DOE) SciDAC-2 program. (OSG grew out of the pioneering Grid2003 infrastructure [37], created by the NSF Grid Physics Network [7] and International Virtual Data Grid Laboratory [8] projects, and the DOE Particle Physics Data Grid projects.) The distributed resources on the grid are independently owned and managed, and individuals access the resources by joining a VO that is registered with OSG (see sidebar 17). OSG is structured as a community of communities, and its functionality is driven directly by the science stakeholders. Thus, while the OSG resources provide a standard software toolkit, VOs are free to add software to support their own needs.

Exhibit 20. This map shows the eleven TeraGrid resource providers in 2008. The facility continues to grow and enhance its resource capabilities.

Both TeraGrid and OSG support science gateways. Through TeraGrid, science gateways provide community-focused venues for the development of research tools (see sidebar 13). OSG's science gateways—which map onto OSG VOs—similarly provide access to OSG resources. OSG and TeraGrid collaborate on middleware, security, education, and training. They are also coordinating efforts to enable application communities to act across the federated infrastructures. For example, OSG users will be able to access TeraGrid resources through the OSG gateway currently under development between Fermilab and NCSA.

At the opposite end of the spectrum, exemplars of a more lightweight approach than grid-enabled environments are popular, Internet-based social networking systems, one of which is Facebook (http://www.facebook.com). Facebook was launched in early 2004 to connect students within recognized education institutions, and today it also includes

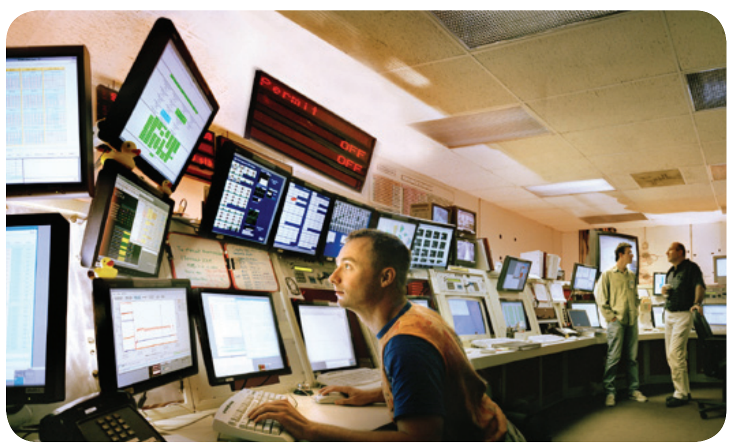


Exhibit 21. One resource provider to the Open Science Grid (OSG) is the FermiGrid, which consolidates the computing resources of the high-energy physics laboratory Fermilab. The efficient and effective use of their resources allows FermiGrid to donate spare central processing unit hours to other VOs that depend on OSG for computing power.

members of recognized companies and nonacademic institutions. Though it began as a hobby project in a dorm room at Harvard, it spread quickly to universities across the world, and claimed more than 59 million users by the end of 2007, while growing at a rate of 250,000 new registrations daily. Within this free, ad-supported system, users create personal profiles, through which they can connect with friends, post photos, write blog entries, form groups, plan events, and play with a variety of free widgets built to work with the site. These widgets are an interesting illustration of the emergent quality of Facebook. The company opened the internal workings of its API to developers who can now develop additional tools that members can add to their profiles.

The products of Web 2.0 spaces such as Facebook are not the result of any real common purpose. In fact, people create collections, directories, and other seemingly integrated sets of resources while working strictly for their own purposes, or for purposes shared by a small group within the whole. The integrated, searchable, browsable collections produced within Web 2.0 environments (particularly sites such as Flickr, YouTube, del.icio.us, and Connotea [66]) are created much more as a result of the nature of the tools provided than of any intent or common purpose among the contributors. In the Web 2.0 world, what sticks is determined by what is successful. Although none of these popular sites offer tools for scientific use, some developers are making initial efforts to think about scientific applications, such as with myExperiment Virtual Research Environment (http://www.myexperiment.org), a new beta site designed to let people share digital items such as workflows with their colleagues.

Exhibit 22. Facebook personal profiles let members share photos and videos, post messages to and from friends, share personal interests and hobbies, and list their friends and social groups. Extra tools add functionality such as games, travelogues, and virtual pets.

Regardless of the approach, building any integrated infrastructure is a difficult and costly endeavor, both technically and socially. Participants still have to invest enormous amounts of time and effort on cobbling together, launching, and sustaining VOs. As they seek to connect people with heterogeneous needs, they face the problem of creating standards and ensuring that they can be adopted across a diverse population. They have to consider many different types of potential needs and interests in any governance arrangements. They must resolve certain issues up front, such as who can join and who owns the intellectual property generated on the site. At the same time, flexibility for the

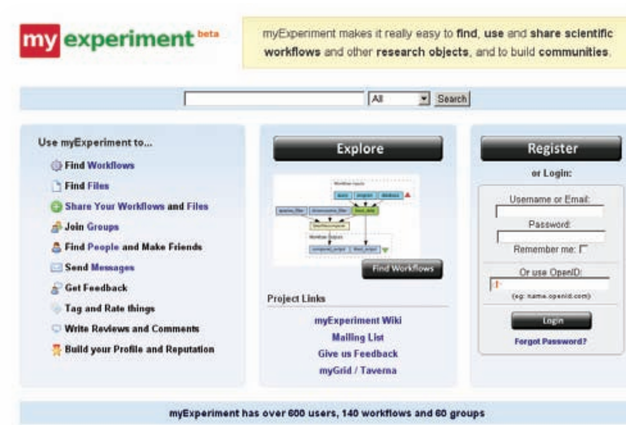


Exhibit 23. The myExperiment Web site, targeted at research scientists, is a Web 2.0 initiative to create the kind of community sharing and connection common among sites such as Facebook and Flickr. It lets members join groups, send messages, and find friends, but it also offers methods for sharing and evaluating research objects, including tagging, rating, reviewing, and commenting.

future and advancements in technology need to be considered. Earlier adopters have to be patient with the instability of the system, and all newcomers face a learning curve that may take time and training.

Some question whether new standardized infrastructures are even necessary. Often, it is not clear whether the system has created a "new" community that could not have accomplished their task in any other way. For example, some collaborations may simply move to a new tool. Has the infrastructure truly transformed collaboration? Others question the value of linking heterogeneous systems into a single global infrastructure. Even though science and engineering researchers find universal access appealing, businesses have yet to decide whether there is a need, and even whether it is desirable, to have a single grid. Some suggest that research and business are better served if multiple, heterogeneous grids connect as they see fit.

In fact, sometimes ironing out the details is simply impractical. Temporary VOs may face timeframes that require immediate action. For example, agency responses to public health crises require a fast and coordinated effort among organizations that may not have worked together previously. The head of the Centers for Disease Control and Prevention (CDC) described the effective and rapid international response to severe acute respiratory syndrome (SARS):

Exhibit 23. During the outbreak of severe acute respiratory syndrome (SARS) in 2003, a worldwide temporary VO emerged to share information and coordinate an appropriate response. The NSF-funded Pacific Rim Applications and Grid Middleware Assembly (PRAGMA) helped Taiwan's National Center for High-performance Computing set up Access Grid sites (see sidebar 14) so that quarantined doctors could consult with specialists at other institutions.

> National and local health agencies across the globe have disseminated up-to-the-minute information tailored for clinicians, public health officials, health care workers, travelers, household contacts, and many other affected parties. . . . Use of the Internet has sped information exchange and helped overcome the problems presented by asynchrony in the activities of investigators in many time zones. Scientists at the international collaborating laboratories are exchanging laboratory results and images on a secure Web site. [40, 2030]

Such capabilities, however, may also be the product of lessons learned. The difficulty of diagnosing the West Nile virus outbreak just four years earlier was hampered by inadequate identification of and coordination with relevant parties [87, 101]. Clearly, experience helps.

We do know that those who choose to build VOs to support a national science agenda should consider several factors. First, they must consider the relationships between technical and organizational features. Neither portion can be left as a black box [106]. One opportunity is to marry technology with organizational knowledge to better understand how science—both technically and socially—is done. Second, developers of heavy-duty systems can learn from the agility of online communities. Beyond the social networking sites, this includes everything from limited enterprise efforts to teacher professional development communities to virtual markets to multiplayer gaming systems. These networks have solved issues associated with managing identities, forming groups with controlled access, monitoring reputation, managing rapid growth, and so on. Third, VO creators need to plan for the long-term sustainability of their project. VOs simply cannot be left to operate without maintenance, but VOs and their funders need to consider where the money will come from when the initial grant runs out [92].

12. eBay—Virtual Economy as VO

The virtual marketplace eBay offers insight into the issues of scale, data management, trust, coordination, and other basic issues of VOs. It is a community in the sense that the 244 million registered users interact, leaving a record of good and bad behavior. It is a monster to manage and support—106 million items are available at any one time, 6 million of which are added and removed daily. This amounts to 34 billion SQL (Structured Query Language) transactions.

Exhibit 25. Electronic commerce, more commonly known as e-commerce or eCommerce, is the buying and selling of products or services over electronic systems such as the Internet. The amount of trade conducted electronically has grown exponentially since the spread of the Internet. Total e-commerce sales for 2007 were estimated at $136.4 billion, an increase of 19 percent (±2.8%) in just one year from 2006.

Just as the transactions, communications, and data are challenging to track, so too is the task of managing a large, distributed, and disaggregated infrastructure of people around the world. Many of the business departments are too much in silos, and bringing them together to achieve greater efficiency is one challenge that eBay faces.

Paul Strong has been modeling workflows and processes and mapping it onto their infrastructure with the goal of determining whether value is added by the variety of tools eBay uses. He finds that he cannot assume that particular systems are optimally managed the same way over time. For example, some tasks that had to be run offline in batches are now better run as online, dynamic transactional systems, boosted by increased computing power.

Another layer of challenge faced by Strong is how to manage a small virtual team in Russia and China. He finds that coping with time zones is much more difficult than the physical distance. Additionally, he must learn cultural preferences. Because creating team spirit is tough and requires constant effort, he recognizes that sometimes the best person to hire for a distributed team is not necessarily the best person technically, but the person who is best at working as part of a VO.

13. TeraGrid Science Gateways

TeraGrid Science Gateways (http://www.teragrid.org/programs/sci_gateways/) expand the ways the grid computing infrastructure is available to scientists, educators, and students. These gateways provide a front end that allows users with varying levels of experience to access features of the TeraGrid (and sometimes other services). These users tap into a community allocation that is managed by the gateway developers, who provide the necessary interface for accessing and supporting TeraGrid services. Gateway providers are typically more familiar with the needs of their user communities, offering Web portals, applications, service brokers, and connections to external grids and data sources that are appropriate for their communities. At present, gateways exist for the fields of astronomy, biology, chemistry, computer science, earth science, engineering, materials science, and physics. In 2007, Gateway members made up the fastest-growing group of TeraGrid users and represented a substantial portion of the entire TeraGrid user community.

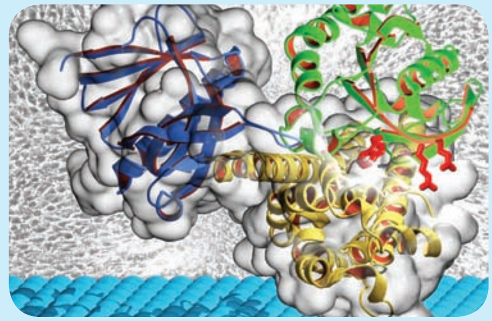


Exhibit 26. Although the majority of TeraGrid users rely on a direct, "command line" interface, Science Gateways represent an area of new and rapid growth. Between 2006 and 2007, Gateway Users increased by 430 percent to just over 500 users. Gateways enable complex analyses and visualizations, such as this protein complex image created with the tools of the TeraGrid Bioportal.

14. Video-Based Collaboration Technologies

As the Internet provides greater data transfer capacity and technologies offer faster data compression, so too have collaboration technologies evolved to support collaboration in more realistic and effective ways. In particular, videoconferencing systems appear to be the most appealing substitute for face-to-face meetings, but their technology requirements, image quality, and cost still present a hurdle for many groups. A few systems in particular have made an impression in research and business communities.

The Access Grid (http://www.accessgrid.org) was an early form of group-to-group videoconferencing that brought together multiple sites in the same virtual venue featuring large-format, multimedia displays. Access Grid connects multiple sites in a virtual meeting room, displaying multiple video images on the screen and allowing access to shared presentations and other files. Access Grid has been particularly popular among universities and research institutions, with thousands of systems deployed worldwide.

Exhibit 27. Weather data and visualizations surround researchers using the Social Computing Room at the Renaissance Computing Institute (RENCI) Engagement Center at the University of North Carolina, Chapel Hill. The room uses 12 projectors to create a 360-degree display, allowing researchers to share data and visualizations in novel ways, experience virtual worlds, and interact with colleagues over long distances and gaming scenarios. The 360-degree display can be operated as an extension of a desktop machine and applications.

Other, similar systems also support the academic research community. For example, the OptIPuter or *Opt*ical networking *I*nternet *P*rotocol comp*uter* project (http://www.optiputer.net) supports data-intensive scientific research and collaboration by allowing researchers to simultaneously view large-scale data sets in real time over dedicated optical networks. These scientists meet at OptIPortals—large tiled displays that enable the collaborators to view and analyze high-resolution visualizations of data from multiple storage sites alongside documents and slide presentations. Likewise, the EVO (Enabling Virtual Organizations) system (http://evo.vrvs.org) provides the high-energy physics community with a collaboration system that automatically adapts to the participants' network configuration to provide reliability and quality regardless of budget. The system provides many features including instant messaging, video or teleconferenced meetings, file sharing, and whiteboard functions.

Unfortunately, many videoconferencing solutions are limited by poor-quality images or awkward room arrangements that reduce the impression of face-to-face communication. Now, new commercial vendors are creating even more sophisticated virtual interaction capabilities. For example, Cisco's TelePresence and HP's Halo provide an in-person meeting experience in rooms that transmit life-size, high-definition video images as well as spatially discrete audio feeds between multiple sites. Because the rooms are set up identically to create the illusion of a seamless meeting table, the experience is realistic—even allowing eye contact.

As the personal and environmental costs of travel for face-to-face meetings become more of a concern, the financial expense of installing videoconferencing systems seems more reasonable. Both HP and Cisco emphasize on their Web sites the time and productivity lost to long-distance travel. More recently, scientists are acknowledging the large carbon emissions associated with extensive airline travel and are searching for alternate solutions [64]. These issues have the potential to spur further development of affordable, high-quality video collaboration systems.

15. Middleware and the Globus Toolkit

Many VOs, such as BIRN, TeraGrid, and SCEC, are distinguished in that their operations require access to underlying computational capabilities. Indeed, in many instances, the VO forms around the need to share computational or data resources. For example, while SCEC is a VO with many concurrent activities, projects such as CyberShake [25] depend on the ability of SCEC to share program code, supercomputer access, and data access across the organization.

Exhibit 28. The Globus Toolkit is one example of middleware services that bridge between computational resources and user-centered applications.

In the past, such distributed resource sharing was done using ad hoc mechanisms that were established on a project-by-project, resource-by-resource basis. As VOs become more pervasive, however, the on-demand and dynamic nature of VOs makes such one-off methods unsuitable as a standard of practice. Rather, a VO is best served by a widely deployed, interoperable infrastructure that enables autonomously operated computational, storage, networking, and other IT resources to be shared across the participants of a VO.

This reasoning led to the creation of grid infrastructure whose stated purpose was to support “coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations” [36]. Grid infrastructure, or middleware as it is sometimes called, has three characteristics that are essential to support the resource sharing requirements of VOs [34]. It (1) coordinates resources that are not subject to centralized control; (2) uses standard, open, general-purpose protocols and interfaces; and (3) delivers nontrivial qualities of service.

Today, grid infrastructure has been applied to many different use cases. For example, Internet-connected virtual supercomputers “scavenge” cycles from loosely coupled, distributed communities to support analytical projects such as SETI@home (http://setiathome.berkeley.edu/) [4]. However, while use cases may vary significantly, research over the past 10 years has shown that it is possible to define a common set of interoperable abstractions that enable VO members to share data across a diverse set of resource types and operational policies.

The Globus Toolkit (http://www.globus.org/) [33] is an example of how a single set of middleware services can be used to support a variety of use cases and VO types. Indeed, the Globus Toolkit is used by BIRN, SCEC, and many other VOs mentioned in this report to provide shared access to IT services across the participants of the VO, and by infrastructures such as OSG and TeraGrid. Globus is not a VO application, but rather it is a service-oriented infrastructure that provides core mechanisms for authentication, policy enforcement, computational resource management, data management, and monitoring. The availability of these mechanisms facilitates a layered architectural approach, enabling tool and service designers to create alternative sets of system or application software that can address the specific requirements of different classes of VOs. Thus, systems such as Kepler [2] and Taverna [77] can be developed to enable the creation of scientific workflows, and systems such as Swift [107] and MPICH-G2 [54] can be developed for parallel programming, over distributed resources.

Middleware comes in other varieties as well. The Condor (http://www.cs.wisc.edu/condor/) [97] system addresses challenges associated with federating and managing computers, while Shibboleth (http://shibboleth.internet2.edu/about.html) interfaces with campus directory services for authorization. Apache Tomcat is an example of middleware that is oriented toward the delivery of Java-based Web server applications, offering security and application delivery. Portals, such as Sakai (http://sakaiproject.org/) and GridSphere (http://www.gridsphere.org), offer a Web-friendly face to a variety of applications, including other levels of middleware. The Storage Resource Broker (SRB) [11] enables access to, and management of, large quantities of data. All of these systems can be, and often are, used in conjunction with Globus to create complete VO solutions. These and many other middleware initiatives are open source, or at least are free.

16. User-Oriented Software

The complexity of the software interface is the big challenge of using much of the sophisticated technology made available in VOs. While middleware bridges some facets of content and computation functionality, most middleware, with the exception of portals, still requires computer-savvy users.

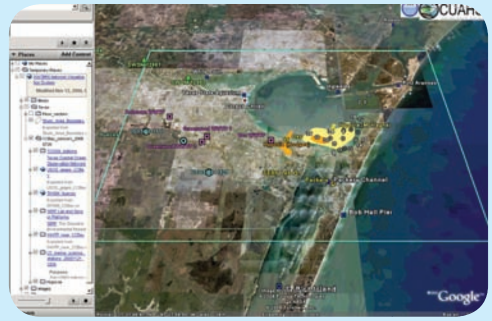


Exhibit 29. This mashup combines Google Earth with information about environmental sensor networks maintained by several university-based, Federal, and local government institutions. Water or environmental researchers can use this application to readily identify sensors in a focal region (here, Corpus Christi Bay, Texas).

Portals, in fact, are one of the older technology solutions for making VOs user friendly. Portal technology is a content aggregator; the content of the portal is displayed in sections called "portlets," which can be customized by the individual user. MyYahoo and iGoogle are commercial examples of customizable portlets, and GridSphere is a similar, open-source portlet environment used in research settings. A newer, Web 2.0 method of user-friendly content aggregation is the "mashup." Mashups aggregate data elements from multiple sources and display an integrated result through a simple graphical user interface. Users may not even realize that their content is coming from multiple sources. For example, researchers associated with the WATERS Network project (described earlier in this report) have adapted Google Earth to display hydrological sensor information in their exact geographic locations. Whereas portals present each source as a discrete block of information, a mashup appears unified. The difference is much like a television dinner with compartments for each part of the meal versus a casserole in which all the ingredients are mixed together.

Another solution to the complexity of VO technology is to use off-the-shelf products. In some cases, middleware can deliver content to be used with downloadable Java client applications or other commonly used products. Some VOs rely on commercial Web-based collaboration tools or suites of document-editing programs, while others use open-source versions of these tools to avoid licensing fees, which can be cost-prohibitive and difficult to estimate with a blurry membership population.

17. The Real Os in VOs

The challenge of being virtual is understanding what roles nonvirtual organizations play in their formation and upkeep. At the January workshop, Ruth Pordes, executive director of the OSG, described how the members of her VO have come to understand these relationships. She explained that VOs (the virtual organizations) depend on Os (the nonvirtual organizations). Specifically, VOs delegate a physical identity to an O, which verifies the identity of the people participating in the VO and which is also responsible for developing the skills and careers of these participants. VOs delegate to Os the purchase and maintenance of physical hardware, and in turn, VOs own the rights to use or access these physical resources. Surrounding these transactions, the VO administers the boundaries, including any agreements established to operate the VO.

Exhibit 30. This image depicts the virtual network of international climate change organizations on the Web. This was created by conducting a colink analysis of significant climate change URLs, using the IssueCrawler, Java crawler, and scaleable vector graphics (SVG) visualization software by the Govcom.org Foundation, Amsterdam. Visualization provided by Reseaulu by Aguidel.com, Paris.

4. RESEARCH CHALLENGES

We have discussed not only successes but also challenges. In brief, we know that VOs are still far too hard to form and operate. We are painfully unaware of which factors make them work or not work. We see many opportunities for understanding VOs from the perspective of several disciplines. Social scientists and technologists have a great deal to learn from each other, and it is important that social scientists and technologists work together as partners to develop new knowledge about how VOs are created and sustained. End users—bringing diverse perspectives from science, engineering, humanities, and business—can also contribute valuable insights. To develop a rich literature about and for VOs, each of these groups should serve as equals—not in diminished roles as "business analyst for hire," technology "plumbers," or passive recipients of technology. Likewise, VO research must recognize and make use of the diversity of theories, methods, and areas of expertise represented in all these areas of study.

One current tension is the sense of urgency that practitioners and funders feel as they anticipate building or sustaining VOs. Unfortunately, it will take significant time to do thorough VO research using traditional social science methods. Perhaps a practical solution is to approach VOs both bottom up and top down.

From the bottom up, technologists are asking the question, "How can we leverage the Internet, Grids, Web 2.0, Virtual Worlds, sensors, and so on to enable collaboration and innovation at a massive scale?" They want to know now what infrastructure they should be building, including tools, platforms, and standards, and they tend to experiment with many possibilities and see what happens and what works. The social sciences and related communities take a top-down perspective. They ask, "How do people communicate and work together? What gets in their way? What do they need to do differently?" The answers to these questions will change as new technologies unfold.

A key question is whether VO research can meet in the middle, allowing for both the planning and the tinkering that are critical to successful technologies [100]. The technologists will keep building tools to enable collaboration, while the social scientists will continue to identify how people actually use these new technologies. The social and the technical perspectives must combine to make the experience a positive one for the user communities that are relying on these systems to make their own research and discovery happen.

The following main topic areas cover issues and prominent questions that September workshop participants believe that VO researchers should consider as they move forward in a partnership. These include issues of definition, comparative frameworks, lifecycles, diversity, codifying research knowledge, technology for knowledge and data sharing, collaboration within and across disciplines, technology-mediated interaction, scaling, motivation, governance, metrics and assessment, and logistics. Cross-topic themes include stability versus emergence, formal versus informal systems, trust, organizational design, and improved participation.

18. VO Research in Europe

European research institutions and their members are acutely aware of the centrality of VOs in the near future. They recognize that projects that are global in scale and last multiple decades require investment and sustenance beyond what individual countries can support. The activity associated with these grand projects, coupled with growing demand more broadly for computational infrastructure, suggest that investment in VO capabilities may not be keeping pace with actual or anticipated needs.

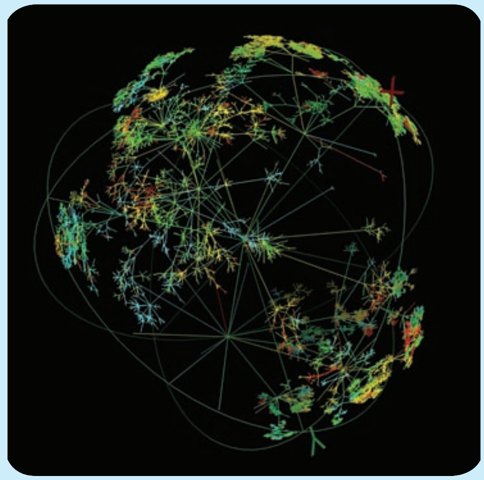


Exhibit 31. As the Internet has grown and its connectivity has collapsed distance between countries and continents, an interesting challenge is to measure and model the new virtual topology. This visualization depicts a single cycle of delay measurements made by a CAIDA Internet performance monitor in Herndon, Virginia, on February 2, 2002. The graph was created using the Walrus graph visualization tool, designed for interactively visualizing large directed graphs in three-dimensional space.

The European Communities have already commissioned several reports to analyze opportunities and propose plans for implementing what they call "e-Infrastructures" [10, 28] to support "e-Science"—what is more commonly termed in the United States as CI or high-performance computing, sometimes called grid computing. These reports are targeted at policymakers, funders, industry, and the broader community of stakeholders in European society. Authors of the reports have drawn on interviews with and surveys of many experts in different fields of research, including social sciences and the humanities, environmental sciences, energy, biomedical and life sciences, materials sciences, astronomy, astrophysics, nuclear and particle physics, and computation and data treatment. While physical and natural sciences are the most obvious beneficiaries of VOs and the underlying infrastructure, focused effort has also been applied to understanding VO relevance to the humanities and social sciences [10] and to scientific fields that are not as deeply steeped in scientific computing as others. Some projects on VOs in different settings—both scientific and other disciplines—have been funded accordingly.

Many themes are pervasive across these reports and have much in common with the issues and topics raised in the United States about the future of CI. For example, Europeans are considering the interaction of many technological components: networking infrastructure, middleware, supercomputers, data, collaboration tools, sensors, and large instruments termed "research infrastructure" [31], which is similar to the NSF's MREFC. The Europeans see research infrastructure as a way of solving key issues such as global warming, energy supplies, clean water, terrorism, demographic changes, and social issues. Such challenges will require interdisciplinary expertise and new ways of analyzing data.

Looking ahead, these reports emphasize concern about pipeline issues, not only to develop young people to become interested in scientific computing, but also to attract the interest of established researchers, particularly in fields that previously have not used e-Infrastructure. One way of attracting new young people to pursue careers in research may be to more effectively connect research organizations, industry, and education institutions. Researchers also see that it is critical to enlist the comprehensive participation of designers, developers, and end users in developing systems, and for broader adoption, technology centers must offer sufficient, ongoing organizational support for users. Other big hurdles to adoption concern data, particularly issues of data confidentiality, sharing, and curation, as well as the interoperability of different data systems.

Indeed, European thought on this topic is similar to the discussions in the United States; however, unlike us, Europeans already recognize how crucial it is to involve the participation and partnership of many stakeholders, organizations, and sources of funding. Connecting with European initiatives through conferences and partnerships should advance research needs on both sides of the Atlantic.

4.1. *DEFINING VIRTUAL ORGANIZATIONS*

To develop a more coherent body of literature on VOs, researchers need to identify and define what they mean by VOs when they study them. In particular, they need to clarify their unit of analysis—both social and technical. Social units of analysis may be individuals, teams, scientific disciplines, individual VOs, or even ecologies of VOs. Technical units of analysis may include specific tools or objects, virtual or immersive environments or "worlds," specialized niches, or collections of such virtual environments. These social and technical units may be combined in multiple ways. VOs tend to change their composition and connections over time, so representing this evolution may be required. Thus, a key question is, "What are the ways to identify the boundaries of VOs and are there differences in the units of analysis?" This type of question opens opportunities for multilevel analysis, looking at effects produced at the disciplinary level, institutional level, team level, and individual level [57].

4.2. *A FRAMEWORK FOR COMPARISON OF VIRTUAL ORGANIZATIONS*

Once researchers have defined their terminology, they need to develop a framework for the comparative analysis or evaluation (on a social, technical, or practical level) of VOs, both within and across them. Such a framework would ideally specify the data that one would want to have about all VOs studied under NSF auspices (without limiting the investigations of individual cases). The framework should account for diversity in terms of size, "success," purpose, interdependence, membership composition, disciplinary heterogeneity, and so on. Other ways of characterizing VOs may include how they are structured in terms of the types of participation they involve, where such inputs come from, and what structures are used to accommodate this participation or input.

To develop this approach, many options are available. Because VOs are held together with technical systems, these systems offer the potential for self-documentation, automatically capturing detailed data about work and interactions for use in research (with appropriate privacy protections). Methods like social network analysis would be excellent for looking at social dynamics within VOs, especially large ones. Intensive qualitative investigations can reveal unanticipated details of structure and culture. Social network analysis can support qualitative research by initially identifying fruitful areas for targeting the more cost- and time-intensive qualitative data collection efforts. One challenge of comparison is that specific kinds of VOs (in terms of combinations of traits) are likely to be few in number. Fortunately, recent developments in comparative methods (e.g., fuzzy set qualitative case analysis) can handle investigations in which only qualitative data and limited diversity of cases are often available; these same methods can also handle quantitative data and many cases [39, 90, 91]. This is a good opportunity for developing more sophisticated computational methods.

4.3. *LIFECYCLES OF VIRTUAL ORGANIZATIONS*

To support VOs effectively, researchers need to better understand their lifecycle—how they begin and evolve, dissolve and reconfigure—at different time scales and over an extended period of time. While research often focuses on the features of successful VOs, failures are important to study as well. Lifecycle research applies to VO development at all stages and with all degrees of success and failure.

Multidisciplinary research is particularly important for understanding the lifecycle of a VO as it is built. Research shows that information technologies and organizational features interact as they develop [106], thus VOs must be codesigned by end users, technologists, and social scientists. Codesign includes the user-centered design of the interface and underlying infrastructure but could also include designing in mechanisms to collect various kinds of data for coanalysis as the VO develops [108]. In relation to the lifecycle of a VO, technologists need to know what is required to be in place before, during, and after the lifetime of such an organization. Likewise, designers and developers need to consider the *social context* that exists before and during the lifecycle of a VO. Large-scale collaboration projects suggest that the success of the VO depends not only on the CI available before collaboration begins, but also on the social and collaboration imperatives that may (or may not) exist among the potential participants of a VO. Studies are needed to understand which social structures are appropriate to support different types of collaboration and to support work throughout CI lifecycles.

VOs do not always start from scratch. Some VOs coevolve alongside another system that was created previously without a VO in mind. Researchers need to better understand the impetus behind such a development and the path dependencies. For example, when are they prompted by an external force (such as the Federal regulations governing BIRN, described above) and when do they spring up opportunistically? What characteristics distinguish these types of VOs and what do these characteristics imply about their formation? To what extent do coevolving structures need to be interdependent? Are VOs more likely among communities with a history of collaboration or are other factors more important [13]?

Finally, once VOs are established, they are rarely static. Thus, researchers should consider what structures need to be stable to allow emergent organizations to continue to evolve and what structures get in the way. As a VO community grows, how will it scale effectively to accommodate the increased complexity of task coordination, social interaction, and technical capacity?

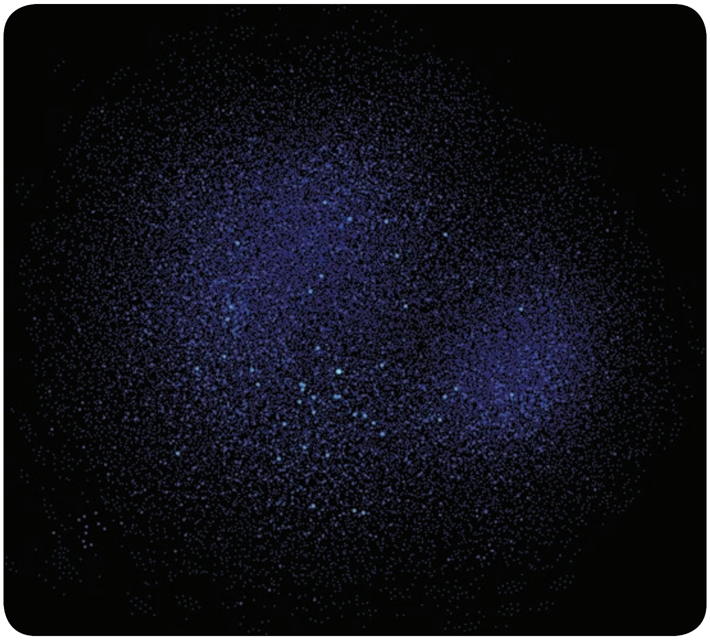
4.4. *LEARNING FROM DIVERSE VIRTUAL ORGANIZATIONS*

VOs are diverse and we risk overgeneralization by treating them as one kind of thing. We need studies to understand the nature and origin of their diversity. This means that researchers—including those funded by NSF—must look beyond just scientific communities to other models and contexts.

For example, when expressing how Web-enabled collaboration has changed the way people work over distance, most examples during the September workshop did not come from science domains, but rather from other domains such as business and enterprise, Web 2.0, open-source software, or other kinds of open development. Scientific communities are relatively risk-averse and cannot be expected to be on the cutting edge because failures would be too expensive. This conflict points to the value of looking at domains other than science, especially at those that exemplify the new paradigms, and comparing those domains with scientific domains and with each other. In some cases, these domains may resemble VOs in only some features; consider, for example, Usenet groups, online communities, Mechanical Turk task groups, and wikis.

Diversity of size is also an issue. Scientific VOs such as those in the fields of high-energy physics, astronomy, biology, and medicine may be many times larger than VOs in the engineering and business or nonprofit domains, but important similarities and differences across domains may give insight into factors leading to success and failure in these contexts.

A third source of diversity is the origin and structuring of VOs. New forms of collaboration provide a simple, flexible tool at the start and allow the collaboration to emerge bottom-up, incorporating grassroots opinion. In other cases, top-down intervention by moderators or conveners is crucial. In many of these cases, the timing of network effects cementing the popularity of a particular community may have more to do with its success than anything inherent in the structure of the VO. In other cases, a preexisting community of practice or social network provides a ready-made audience in ways that a task-oriented VO may not.



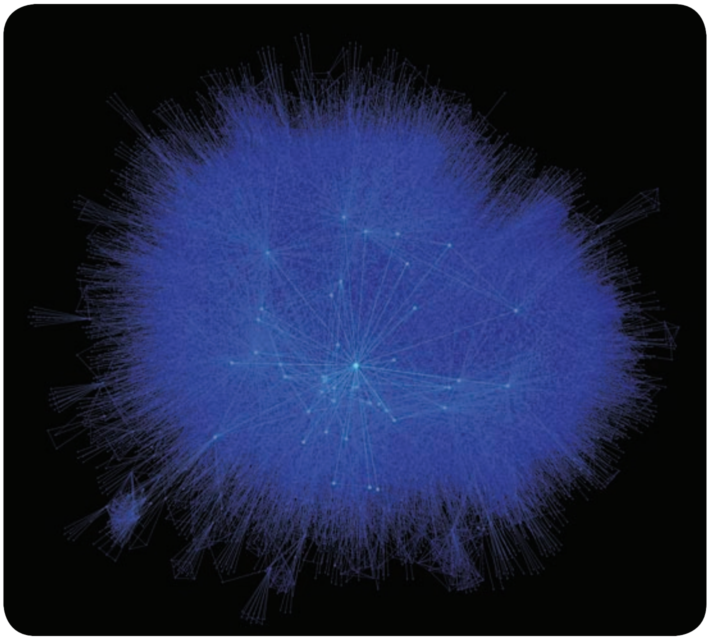


Exhibit 32. Virtual organizations make social networks visible. These two images portray the complex connections between people (dots) and their connections (lines) as captured on the Friendster® social networking site. The first image shows the distribution of individuals who are part of a potential network, whereas the second image reveals the connections between these individuals as a complete network. Individuals and connections closest to the central person are brighter than those who are several steps removed.

Exhibit 33. InnoCentive makes the most of virtual communities and Web-enabled collaboration by helping organizations with challenging problems connect with individuals who can potentially solve those problems. Both parties are kept anonymous until the final solution is selected, which typically yields a reward of $10,000 to $100,000 for the solver. Open innovation networks such as InnoCentive add to the diversity of VOs available for study.

4.5. *IMPACTS OF RESEARCH ON IMPLEMENTATION OF VIRTUAL ORGANIZATIONS*

Any analysis of VOs has implications for the design of future VOs. Already there is a perceived need to codify the explicit and tacit knowledge associated with creating and operating VOs, and the question remains as to how to educate people to develop and participate in VOs. In particular, it would be valuable to offer guidance on these questions to those forming and operating VOs by compiling existing research (as described briefly in this report and other published venues) and conducting additional research (finding answers to other issues raised in this section). These groups have practical decisions to make that have financial, social, and intellectual implications for immediate issues, such as the technology they use or develop and the support for whatever face-to-face meetings are necessary. as well as for larger issues of governance and strategic planning.

Guidance could take many forms. New VOs may need to consider a checklist of crucial issues that, in some cases, might become principles, rules, or requirements. Just as NSF and other sources of funding have fairly strict criteria that must be met before supercomputing resources can be allocated to a project, these criteria may help ensure that the resources will not be wasted and that there is a reasonable probability of real results. Thus, a proposed VO may need to demonstrate that they have the capacity to work effectively together and have thought carefully about how the VO they create will be structured to accomplish their goals. One checklist can be found in the "theory of remote scientific collaboration" (TORSC), which outlines five categories of important prerequisites for successful collaboration: the nature of the work; common ground; collaboration readiness; management, planning, and decision making; and technology readiness [82]. (See sidebar 19 for further details.)

Technical designs need to consider social as well as organizational and structural issues, while allowing for the users to adapt the system to fit changing needs. However, a future research question is how to balance (or make use of) general purpose infrastructure versus field-specific infrastructure. In business settings, some think that the project cycle should be decoupled from the infrastructure building cycle because the evolution of infrastructure operates on a time scale that is not consistent with the shorter time scale of projects.

Some suggest that an eventual goal for researchers in VOs would be to provide toolkits that would help establish and run VOs more rapidly. These toolkits would need to be robust and simple to use, taking into consideration that the people who manage them may not be technology specialists with computer science doctorates or training as system administrators. If such a plan were implemented, the following issues should be addressed:

- What patterns of organizational structure or process need to be considered to provide adequate VO templates for the variety of projects that exist? What traditional components of management need to be considered and replicated in a VO? Where does a VO fit in an existing organization?
- What technology features are necessary to support the "5-minute" VO? In other words, what would be the common set of infrastructure services that would allow collaborators to quickly and easily form a VO? Some of the features that need to be considered through the interface are as follows: settings for administrative roles, resource use, and desired policies; membership access; and configuration of the resources, data, and applications to allow or restrict access as necessary.
- How can tools support the social aspects of a VO? Options might include collaboration visualization tools that show metrics of activity, analyze use and attention, monitor and support social networking, and apply peer pressure. Privacy issues become paramount in this situation.
- How could toolkits be provided? One possibility is to allow third parties to provide them, much as Amazon has commoditized computational capacity and storage through Elastic Compute Cloud (Amazon EC2) and Simple Storage Service (Amazon S3). Advantages of using a third party may include lower cost, greater flexibility, and overcoming trust negotiation.

19. A Potential Checklist for VOs-to-Be

Exhibit 34. An online survey will allow geographically distributed collaborators (and the researchers who study them) to better understand whether they are technologically prepared to work together. The TORSC acronym stands for Theory of Remote Scientific Collaboration.

The Nature of the Work

- Participants can work somewhat **independently** from one another.
- The work is **unambiguous** (although a complete lack of ambiguity is impossible, clear understanding of the task is desirable).

Common Ground

- The task is so simple (or simple enough) that **vocabulary** is not an issue. Alternatively, other systems provide a bridge between vocabularies.
- **Previous collaboration** with these people was successful.
- Participants share a **common vocabulary.**
  - If not, there is a dictionary.
  - If not, there is a culture that actively helps people understand.
- Participants share a **common management** or working style.

Collaboration Readiness

- The culture is naturally **collaborative**.
- Participants have a **motivation** to work together that includes mix of skills required, greater productivity, they like working together, there is something in it for everyone, *not* a mandate from the funding source, the only way to get the money, asymmetries in value, and so on.
- Participants **trust** each other sufficiently to be reliable, produce with high quality, and have their best interests at heart.
- The goals are **aligned** or congruent in each subcommunity.
- Participants have a sense of **group self-efficacy** (able to complete tasks in spite of barriers).

Management, Planning, and Decision Making

- The principals have **time** to do this work.
- The distributed players can communicate with each other in **real time** more than 4 hours a day, or at least have systems to make communication more feasible as needed.
- There is critical **mass** at each location.
- There is a **point person** at each location.
- A **management plan** is in place.
- The **project manager** is respected and has real management experience.
- A **communication plan** is in place.
- The plan has room for **reflection** and redirection.
- No **legal** issues remain (e.g., IP).
- No **financial** issues remain (e.g., money is distributed to fit the work, not politics).
- A **knowledge management system** is in place.
- Decision making is free of **favoritism.**
- Decisions are based on **fair and open criteria.**
- Everyone has an opportunity to **influence** or challenge decisions.

Technology Readiness

- Collaboration technologies provide the right functionality and are **easy to use.**
- If technologies need to be built, user-centered practices are in place.
- Participants are comfortable with the collaboration technologies.
- Technologies give **benefit** to the participants.
- Technologies are **reliable.**
- Agreement exists among participants as to what **platform** to use.
- **Networking** supports the work that needs to be done.
- **Technical support** resides at each location.
- An overall technical **coordinator** is in place.
- Special issues—
  - If data sharing is one of the goals, de facto standards are in place and shared by all participants, and a plan for archiving is in place.
  - If instrument sharing is part of the collaboration, a plan to certify remote users is in place.

Source: Adapted from Olson [80]. For further information, see Olson, Hofer, Bos, Zimmerman, Olson, Cooney, and Faniel [82]. Used with permission.

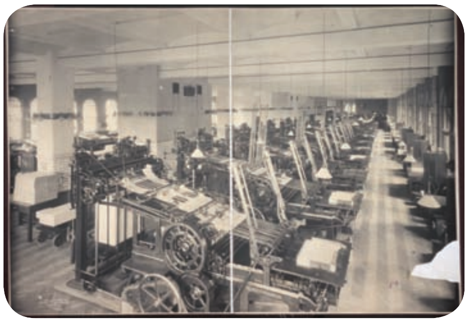
4.6. *TECHNOLOGY FOR KNOWLEDGE AND DATA SHARING IN VIRTUAL ORGANIZATIONS*

Additional research is necessary to better understand how VO technology can support knowledge and data sharing. These issues fall into two rough categories: knowledge access and data management.

The first category concerns how people identify where to find knowledge and learn from others. Knowledge sources often include other people but also may include networks [14], teams, documents, technologies, or software. The question is how people might use technology to identify these knowledge sources, specifically how they could visualize and navigate this information. In the September workshop, there was general agreement that instrumentation of VO infrastructure was critical to social science research on collaborative activities. Suggestions on how to perform this instrumentation were offered from several angles.

One suggestion came from software engineering, which has long used tools that create an extremely rich history of all work that is done, all interactions among people, and how the technical work is related to the objectives of the collaboration. We can draw on this engineering experience to design tools that create rich data repositories and to create analysis techniques that produce interpretable results.

Another, somewhat novel, source of ideas about instrumentation was social network analysis. Analyzing individuals, the artifacts they touch, the connections among people and artifacts, and so on is at the heart of collaborative work, and a rich set of tools is available that, to date, has not been used extensively. Finally, there were suggestions that "process" data captured for social science purposes may also be useful for providing novel collaborative functionality (see, for example, Cataldo, Wagstrom, Herbsleb, and Carley [19]).



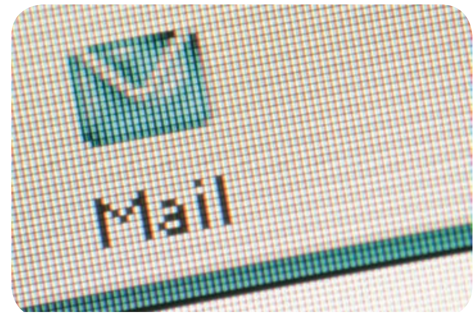


Exhibit 35. Scientists have always used technology to support knowledge and data sharing at a distance. As these technologies have evolved over time, so too must the research about the sociotechnical dimensions of distributed collaboration.

The second category is how to handle data. Many factors influence the degree to which a VO can offer an integrated function for accessing and sharing data. We need to know more about what conditions need to be set before, during, and after data are used. Active research considers standards and their impact on data types, metadata, and quality as well as the degree to which architectures must be centralized or not. Standards will likely vary across disciplines depending on existing customs and outside regulation (e.g., IRB). In fact, CI and advances in high-performance computing that are incorporated into VOs suggest that our notion of data may need to be expanded to consider the models, simulations, and computations that use data to be data themselves. How to engender trust in data sharing is an open question, as is how data can be both stable and agile at the same time. Although these data issues are not unique to VOs, they do present some new challenges in the data's role as a mode of collaboration and communication among members of the VO.

4.7. *SUPPORTING COLLABORATION WITHIN AND ACROSS DISCIPLINES THROUGH VIRTUAL ORGANIZATIONS*

In the near future, VOs might involve at least three disciplines: technology specialists, social scientists, and the end-user community. To even begin, potential VO participants within these three general categories may struggle to find collaborators in an unfamiliar field whose work is relevant. Once they find each other, they face the challenge of working with and reconciling their differences.

Some of the differences include vocabulary, culture, reward systems, data sharing, orientations toward the use of technology, and experience using relevant technologies. Likewise, specific domain knowledge requires translation for those outside the field. Yet these differences and gaps in understanding may not even be evident at first. The question remains as to how disparate groups come to common

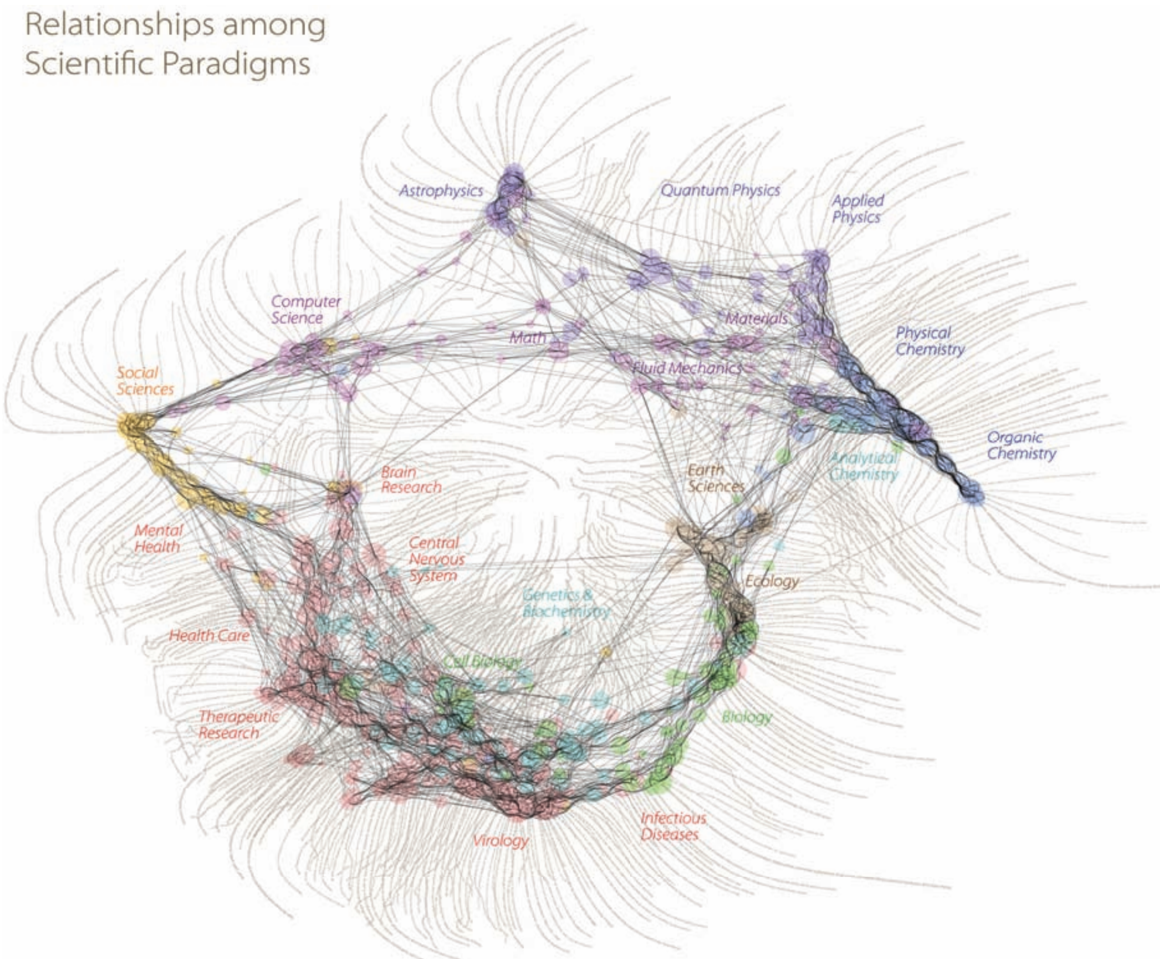


Exhibit 36. The Topic Map was constructed by sorting roughly 800,000 scientific papers (shown as white dots) into 776 different scientific paradigms (red circular nodes) based on how often the papers were cited together by authors of other papers. Links (curved lines) were made between the paradigms that shared common members, and similar paradigms are nearer one another. Labels list common words unique to each paradigm.

ground for understanding and accessing this tacit knowledge. A community needs to be so dedicated that they will see the need for collaboration and put up with these hurdles.

Layered on top of a VO's internal cross-disciplinary challenges is the relationship of the VO to the larger scientific community. A community's own legitimacy and its relationships with established institutions may affect its function. For example, some disciplines are loosely organized and may not have effective social connections for getting a VO off the ground. The ecology of prior VOs may influence and constrain the composition and effectiveness of a new VO. In fact, the development of new "watering holes" may not only change the patterns of established groups, but also encourage the formation of new groups.

Those who work with multiple disciplines to help create VOs to suit their unique needs may find challenges for translating what works in one discipline or domain to what would work in another. For example, physicists expect to write code and prefer an information technology infrastructure that makes this easy (e.g., command line interfaces). Scientists in other disciplines, on the other hand, may expect point-and-click functionality. Building a common infrastructure that makes both comfortable would be difficult.

Yet another layer is that the VO may be transforming the nature of existing communities—particularly how science is done—but we do not yet know how a VO with transformative goals might need to be designed differently from a VO aimed at steady-state science. New communities may be formed as they share facilities at a distance or share data across disciplines.

4.8. *INTERACTION BETWEEN PEOPLE VIA VIRTUAL ORGANIZATIONS*

VOs continue to struggle with the problem of making technology-mediated collaboration comparable to face-to-face interaction. The lack of physical presence raises a host of issues for the day-to-day collaboration that takes place, and conversely, few studies identify how virtual might be superior to face-to-face interaction (for an exception, see Hollan and Stornetta [49]). Three prime areas for studying interaction include coordination, situational awareness, and cohesion.

Without collocation, coordination is more complex. People may have different cognitive and task dependencies in a virtual environment, and technology may help us instrument interactions to better understand how coordination happens. However, only a few studies [51, 63, 79] have investigated the particular coordination and social practices within CI.

Distributed groups have much more difficulty with situational awareness, but we need to know more about supporting it. For example, research has to be done to better understand how we can make attention-getting efforts (and understanding what others are focused on doing) more natural in distributed groups. Also, we need to examine how people shift attention between people and projects when they inhabit lots of them.

Also, VOs may lack the cohesion associated with a shared institution or location. People may feel less connected with and less invested in a VO without the direct social connections. We do not yet know how media richness might improve these connections. If members do not have formal authority and may be "equals" across sites, leadership must also be different.

Each of these issues has implications for whether VOs should be striving to operate like face-to-face organizations do or whether they should be trying to create a system that is altogether different. Although the answer to this question depends on what we learn by studying coordination, situational awareness, and cohesion more fully, VOs provide a venue for experimenting with new formats that do not have analogues in the face-to-face world.

4.9. *SCALING VIRTUAL ORGANIZATIONS*

The question of scale in VOs is crucial. In smaller VOs, personal traits and contacts seem to weigh heavily and expertise may be relatively broad for each individual. In larger VOs, expertise in one slot or another may need to be more specialized and personal traits may become less significant along with organizational protocols and explicit role assignments. So-called massive VOs may function completely differently than small ones. Thus, we need to study the unique determinants and dynamics of massive VOs.

The expectations for massive VOs may also shift the technology that is used. For example, physicists are accustomed to collaborating on huge projects, required by the size of the equipment needed for their observations. Other fields face this trend, although the scale is still much smaller. These differences in scale require different technologies, especially when we see a phase shift (e.g., I can know each individual or I cannot). Also, the capacity necessary to support huge "clickwork-style" projects would be different from one that assumes more sporadic access or data management.

Exhibit 37. Just as the scaling from small to "massive" VOs brings changes in function and structure, scaling of computational power requires dramatic shifts in its use. The processor in a standard notebook computer might run 500 million floating-point operations per second (FLOPS), whereas the IBM supercomputer known as Blue Ice (left) can compute 12 trillion FLOPS. To optimize these capabilities, researchers must write new computer codes.

4.10. *MOTIVATION AND REWARDS FOR PARTICIPATING IN VIRTUAL ORGANIZATIONS*

Given that collaboration is not mandatory in many sciences, what motivates people to participate in VOs? We need to better understand how to design sociotechnical systems that leverage or enhance existing motivations, given the nature of the science.

VOs may need different technologies at different stages of their lifecycles because there are different social needs, incentives, and motivations over the course of a project.

In some cases, external incentives and rewards will be the draw, whereas for others, the costs of or the barriers to participation are stronger determinants. We might investigate the effects of social factors, task interdependence, community heterogeneity and homogeneity, shared identity, and perceived boundaries between organizations. Also, many people simply do not understand what a VO (particularly grid computing) offers, so in some cases, better education and information about what VOs can offer is needed. Ultimately, some people may need to be co-opted while others may respond better to long-term courting.

Individual characteristics may also play a role. The type of people who have already self-selected into a VO would be predisposed to make it work. These predilections may be discrete or a combination of attributes, such as being an early (versus late) adopter of advanced technologies, having a propensity for cutting-edge research, or displaying risk-taking behaviors. As we enfranchise groups that have traditionally been on the periphery of research, we must understand and adapt to cultural preferences that will improve participation.

Finally, the disciplines initiating a VO need to consider how other disciplines would like to be involved and treat those needs respectfully. For example, computer scientists do not want to be treated like car mechanics, nor do domain scientists want to struggle with a beta version of software that will be abandoned once the proof-of-concept is done.

4.11. *GOVERNANCE OF VIRTUAL ORGANIZATIONS*

A key issue for VOs is how to form governance agreements. Not much research has considered what they should cover (see, for example, Markus [68]; Markus, Steinfield, Wigand, and Minton [69]), but possibilities include membership, leadership, IP ownership, security and privacy, and the relationship to physical technical architectures. When some VOs begin, formal governance agreements may not be necessary, but over time, more management and thus more structure may be needed. Other VOs begin with the knowledge that they will be large and complex and will require a more top-down approach to governance. However, some issues may be difficult to foresee and must be addressed through bottom-up policies. To identify some of the options and approaches that have been used effectively, further research may consider comparing and contrasting existing governance agreements for VOs such as the OSG, TeraGrid, NEES (Network for Earthquake Engineering Simulation), and NEON.

4.12. *METRICS AND ASSESSMENT OF VIRTUAL ORGANIZATIONS*

How do we define and evaluate success or failure in VOs? The appropriate processes and metrics for evaluation may vary as much as the VOs themselves and should be considered as a stream of research. Meanings and metrics might focus on community-specific objectives, education and outreach, networking outcomes, and social or technical factors. Metrics may need to shift over time. For example, before a VO starts, it may be important to evaluate whether it is collaboration-ready. In the middle of the development of a VO, formative evaluation might focus on whether a group is going in the right direction even if they are not yet there. Other evaluative areas might include the following:

- Impacts on competitiveness (do VOs provide effective mechanisms for constructing flexible arrangements that lead to more efficient transaction and effective innovation?)
- Democratization of technology (do VOs lower traditional barriers to underrepresented groups?)
- Innovation in research (do VOs enhance access to diverse skills and expertise to advance transformative discovery?)
- Value-added for participants (do VOs improve the research and learning opportunities for scientists and students?)

Studying failures is as important as studying success. Failures may help identify the barriers associated with different disciplines (such as data confidentiality or incentives). Failure may also be multilayered. A VO may exist to achieve the goals of a project and the VO may be effective even though the project fails for reasons beyond the VO. Alternatively, a VO may appear to fail on one facet but may be a tremendous success in an unexpected facet. With the sheer diversity of VO purposes and forms, one would have to expect that some will produce good outcomes while others will not fare so well. Workshop participants discussed the value of "crash teams"—much like those that study plane crashes at the National Transportation Safety Board—that would gather data from failed (or nearly failed) VOs, before the associated people and data disappear, to better learn from these failures.

Some issues faced by VOs are not part of the research agenda. These issues are practical and, although they may elicit the interest of those interested in policy or the history of science and technology, they need to be addressed as soon as possible to avoid elevated development costs. The two primary issues are the viability of shared infrastructure and the deployment, maintenance, and support of VO infrastructure.

5.1. *THE TENSION BETWEEN CUSTOMIZATION AND SHARED INFRASTRUCTURE*

One concern shared by NSF and the people who build VOs is the tremendous cost and time that goes into developing custom infrastructures for each new VO. Many people would appreciate the opportunity to borrow or repurpose code that has been developed for purposes similar to their own rather than "reinventing the wheel." Meanwhile, once VOs are built, their often-unique systems have to be maintained, often indefinitely. Thus, many people encourage the development of shared infrastructure that enables VOs to draw on standard components. This would not only allow VOs to be built more quickly—the 5-minute VO—but would allow maintenance to become more centralized and flexible. The flip side to this argument is that VOs are too complicated, dynamic, and emergent to be pinned to predetermined systems. The main areas of this debate focus on standards, commodity infrastructures, control issues, and policies and contracts.

5.1.1. *Standards*

Global connectivity and data exchange require the existence of clear common standards to reach necessary levels of interoperability. However, companies—and sometimes research centers—are reluctant to move to open standards because their survival depends on proprietary solutions.

We need to raise these groups' confidence in the benefits of using standards versus proprietary solutions, and this will be possible only if we show and demonstrate new business models that achieve better incomes through the use of standards. Many of the resistant enterprises are those that are key players in standardization organizations (e.g., Open Grid Forum [OGF], European Telecommunications Standards Institute [ETSI], the World Wide Web Consortium [W3C]). They have the power to speed up or slow down the establishment and adoption of worldwide standards. Standardization initiated by major players in the information technology industry and recognized alliances would speed up the process of reducing doubts and would foster adoption. Some September workshop participants also believe that European Technology Platforms (ETPs) will have an important role in reaching common agreements, supporting technology developments, and fostering standards adoption. Another approach to dealing with the long timeframes necessary for adoption is to look at industry success stories, which sometimes are based on previous research projects, to provide de facto standards that can be converted to official standards.

5.1.2. *Commodity Infrastructure*

September workshop participants see potential for commoditizing VO infrastructure services. These services could be "rented" on demand just as any other resource might. Thus, these resources would be available when needed and cost only as much as they are used. Such systems have proven to be viable, as seen in Amazon.com's recent commoditization of computing power and storage through Amazon's EC2 and S3. The contracts associated with a "free market" would likely be simpler and more standardized. Already, Yahoo! Inc. is providing academic researchers with free access to its open-source program Hadoop and a supercomputing cluster to support the study of Internet-scale systems software. This initiative offers a commodity-like infrastructure independent of the NSF- or university-funded supercomputing centers that typically provide resources to researchers.

Exhibit 38. NSF's partnership with Google Inc. and IBM via The Cluster Exploratory (CluE) program will enable the academic research community to conduct experiments and test new theories and ideas using a large-scale, massively distributed computing cluster. CluE will accelerate research on Internet-scale computing and expand access to research infrastructure for academic institutions across the nation.

The question is, "What would be necessary to build such a capacity?" If NSF were to provide this infrastructure, we would need a "meta-VO-infrastructure"—factories to create those services. More sophisticated virtual machines will make this a viable alternative for many VOs, which likely would be able to share resources across VOs more smoothly. Because they are not tied to the real hardware, this would offer economies of scale and a higher return on investment.

5.1.3. *Control Issues*

Organizations consider themselves much more vulnerable if they expose or outsource internal resources of the company, and they are skeptical about the virtualization or automation of services. They prefer to maintain in-house control of their systems, or they work with other organizations that are familiar or have a good reputation. In many companies or institutions, the perception of the application being directly linked to the "box" (something tangible) needs to be removed.

5.1.4. *Policies and Contracts*

Those building VOs typically do not have experience with governance, policy administration, and contracts. Participants suggested that new VOs be provided with documents that suggest, for example, "Here is what your governance document should cover and here are some examples." The agreements should spell out the purpose of the collaboration, the investment of the individual partners, those responsible and accountable within the collaborating organizations, and a high-level adherence policy. Likewise, security policies for users of the VO will need to be considered in the higher-level documents.

Even if this advice is not yet codified, recommendations can be bootstrapped from typical business contracts. Thus, participants in a VO are tied to these policies to gain access to the resources. The policies will need to be general or flexible enough to accommodate the inevitable development of new tools. These documents may need to be adapted to suit the unique regulations of the participating organizations.

There are also legal issues to be considered as VOs become more prevalent. A significant one is intellectual property. The use of commoditized infrastructure may, to some degree, alleviate at least one area in which intellectual property complicates collaboration agreements. At present, many universities are so keen to encourage technology transfer revenues that they make partnerships too complicated or they disallow university employees from making software free under open-source standards such as the GNU (GNU's Not Unix) general public license.

5.2. *INFRASTRUCTURE DEPLOYMENT, MAINTENANCE, AND SUPPORT*

Two key issues of infrastructure deployment, maintenance, and support are the cooperation of vendors and VO builders and the availability of people to run the systems over time.

The acquisition of hardware and software from vendors is a thorny issue. The problem with hardware vendors is that research applications typically are a small market, so some people building high-end computing systems have difficulty finding or acquiring appropriate technology for their needs. Also, hardware vendors have not been part of business agreements for shared infrastructures, whereby the cost of the equipment could be paid for over time as it is used, but upfront costs may be too much for new VOs or service providers. On the software side, software vendors are not used to licensing across multiple institutions or accommodating grid systems. For example, certain proprietary software is often common in the laboratories or computing centers of universities, but if VO members access that same software as an integrated part of a VO service, they would be required to pay for its use as if they were not licensed. Open-source software solves this problem to an extent, but the support that proprietary packages offer is often superior and thus a draw for scientists.

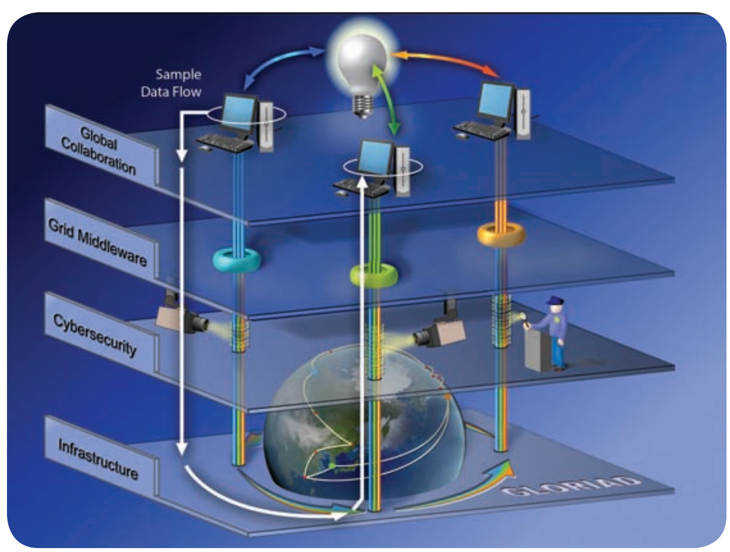


Exhibit 39. The Global Ring Network for Advanced Applications Development (GLORIAD) provides a network of high-speed computing capability to scientists and engineers in the Northern hemisphere. Through the use of grid middleware, the GLORIAD network allows applications and data from disparate sources to be worked on collaboratively by researchers across international boundaries.

The second issue is who will administer these complex VOs. Whether VOs continue to be one-offs or are built from a common infrastructure, a talented pool of professionals will be necessary to build and operate these systems. An important factor is to cross-train people in information technology, CI, and computer science alongside the domains. This training will enable them to appreciate all sides and develop science-oriented tools and resources that can be available to others. This requirement highlights the need for disciplines like bioinformatics, geoinformatics, ecoinformatics, and so forth, in which the scientists are explicitly exposed to and trained in technology concepts.

Unfortunately, each of the domains of information technology has gotten extremely complex, necessitating siloed specialists in networks, storage, databases, hardware, and applications. Somehow, new VO infrastructures need to be simple and easy enough to manage and operate by basic technicians and not only by doctors or highly skilled university researchers. Even if informatics disciplines continue to grow, qualified personnel will remain scarce. If we build 5-minute VOs, people who are not trained as system administrators will find themselves in charge. These tools need to bridge that gap.

5.3. *LOGISTICAL ISSUES FOR NSF TO CONSIDER*

Researchers expressed concern about the logistics associated with assembling a multidisciplinary cast to perform research on VOs. Assuming that studies of VOs best happen alongside the development of new VOs, proposals will need to identify collaborators in a domain as well as those parties who can study both technological and social issues. If all proposals have to integrate technology development and social science, it will limit who can apply. Potential applicants need to recognize the expertise necessary for the VO and identify potential partners in all areas. One workshop participant recalled a recent proposal that was rejected because the team had failed to realize the level of expertise necessary for the technical needs of the VO. Thus, a mechanism is needed for groups to learn from each other as they plan future research—perhaps a VO to study VOs?

Likewise, social scientists are eager to conduct field studies of VOs, but it takes a long time to develop relationships such that solid data can be collected. NSF support would be a huge help in establishing credentials and possibly even matchmaking. In the same way, partnerships between the disparate parties who come together to build a VO could benefit from having a "dating phase before a marriage" (as one workshop participant put it). These "social alpha" and "social beta" versions might help guarantee that investments in VOs are more successful.

Many of the existing and emerging VOs are coupled with existing communities of practice. To make participation viable in an ongoing way for other communities, the research should benefit them whenever possible (i.e., language, findings, metrics) and should be conveyed as soon as possible. An action research approach may be useful to cyclically harvest this understanding during a project.

6. RECOMMENDATIONS

Based on existing knowledge about and understanding of VOs and the anticipated research and development challenges, we offer the following recommendations for moving forward:

1. Encourage cross-disciplinary studies involving both technologists and social scientists working with domain-centered VOs. To enable such studies, support a matchmaking process so that participants from multiple domains can more easily find collaborative partners. Additionally, NSF could usefully create a clearinghouse for information about existing technologies to prevent the needless duplication that can result from the lack of technological experience and awareness on new projects. Furthermore, require participants in such studies to feed lessons learned back into the project so that the VO benefits.
2. Combine knowledge from multiple studies to put forward a framework that can inform further VO research and practice. Facets of the framework might include the following: what constitutes a VO, where VOs come from, how VOs are governed and structured, how VOs operate, how technology supports operation, how needs (social and technological) shift over time, and how VOs grow most effectively.
3. Develop a checklist of necessary VOs features—technological, social, organizational, and so on—to ensure that new VOs start off on the right track. VOs need to be designed deliberately from the beginning [96]. To fully understand the system of virtualization, all inputs need to be quantified. Research supported by NSF should result in models and research systems that represent a basic level of effectiveness for distributed collaboration.
4. Design instrumentation, metrics, and evaluation as part of a VO from the beginning rather than adding measurements systems postmortem. Doing so should motivate the inclusion of social scientists in these projects. It would also increase the return on money invested in VO projects, producing both domain-centered and VO-centered research outcomes.
5. Support human capital development around VOs. To build a pipeline of qualified experts who understand and can support VOs requires education initiatives such as workshops, summer institutes, and even computational science concentrations at colleges and universities. Such efforts would build more of an institution and a professional community around VOs. (Consider, for example, the OGF [http://www.gridforum.org/], which sponsors education events, distributes publications and presentations online, and supports shared-interest communities as part of their mission of "providing an open forum for grid innovation and developing open standards for grid software interoperability.")
6. Investigate whether technological and organizational factors that support effective virtualization can be standardized or provided as commoditized infrastructure. Commoditized, on-demand computational and storage systems may offer more practical and economical solutions for certain types of VO.
7. Offer awards for supporting community services at all levels, including the development of new scientific applications, operation of technology infrastructures, and ongoing maintenance of these services. Such funding would make VOs more viable for the long term.
8. Identify incentives and offer rewards for metacontributors to VOs—that is, the people who build or reorganize features to make it easier for others. Because there is currently no scientific merit system for rewarding those who build a VO—despite the effort that it takes—there is a chicken-and-egg problem of how to entice people to join a VO if they think it will not be productive.
9. Support the development of hardened common tools and protocols for sharing knowledge and data. These development projects need to involve both technology experts and social scientists or human-computer interaction specialists because technological and organizational issues are inseparable and effective solutions must integrate them both.
10. Create proposal funding models that support the use and reuse of VO infrastructures. This type of funding would offer a mechanism and incentive for developing features that certain infrastructure packages do not have.
11. Encourage universities to support VOs with substantial, complementary investments. NSF is not mandated nor adequately funded to support all the VOs they help to initiate on a permanent basis. University support would help build localized, stable communities of computational science expertise.
12. Establish cross-directorate funding opportunities that could more appropriately evaluate and support projects that unite social scientists, computer scientists, and domain scientists. As most funding currently stands, awards are typically siloed within a single directorate. This makes it more difficult for VO participants to apply in unfamiliar domains because panel members are not accustomed to evaluating the merits of proposal content outside their area of expertise.

7. REFERENCES

1. Courant, P., Fraser, S., Goodchild, M., Hedstrom, M., Henry, C., Kaufman, P., McGann, J., Rosenzweig, R., Unsworth, J., and Zuckerman, B. "Our cultural commonwealth: The final report of the American council of learned societies commission on cyberinfrastructure for the humanities & social sciences." American Council of Learned Institutions, New York, NY (2006).
2. Altintas, I., Berkley, C., Jaeger, E., Jones, M., Ludäscher, B., and Mock, S. "Kepler: An extensible system for design and execution of scientific workflows." In *16th International Conference on Scientific and Statistical Database Management (SSDBM'04)*. IEEE Computer Society, Santorini Island, Greece, 2004, 423.
3. Ancona, D., and Chong, C.-L. "Entrainment: Pace, cycle, and rhythm in organizational behavior." In Staw, B.M., and Cummings, L.L. eds. *Research in Organizational Behavior.* JAI Press, Greenwich, CT, 1996, 251–284.
4. Anderson, D.P., Cobb, J., Korpella, E., Lebofsky, M., and Werthimer, D. "SETI@home: An experiment in public-resource computing." *Communications of the ACM* 45, 11 (2002), 56–61.
5. Atkins, D.E. "Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure." National Science Foundation, Arlington, VA (2003), http://www.nsf.gov/cise/sci/reports/atkins.pdf.
6. Austin, J.R. "Transactive memory in organizational groups: The effects of content, consensus, specialization, and accuracy on group performance." *Journal of Applied Psychology* 88, 5 (2003), 866–878.
7. Avery, P., and Foster, I. "The GriPhyN project: Towards petascale virtual data grids." Technical report GriPhyN-2001-15, Grid Physics Network (2001), www.griphyn.org.
8. Avery, P., Foster, I., Gardner, R., Newman, H., and Szalay, A. "An international virtual-data Grid laboratory for data intensive science." GriPhyN-2001-2 (2001), www.griphyn.org.
9. Bainbridge, W.S., "The scientific research potential of virtual worlds." *Science* 317 (2007), 472–476.
10. Barjak, F., Wiegand, G., Lane, J., Kertcher, Z., Procter, R., Poschen, M., Robinson, S., and Mentrup, A. "Accelerating transition to virtual research organisation in social science (AVROSS)." Information Society and Media Directorate General, Commission of the European Communities (December 21, 2007).
11. Baru, C., Moore, R., Rajasekar, A., and Wan, M. "The SDSC storage resource broker." In *8th Annual IBM Centers for Advanced Studies Conference.* Toronto, Canada, 1998.
12. Bernholdt, D., Bharathi, S., Brown, D., Chanchio, K., Chen, M., Chervenak, A., Cinquini, L., Drach, B., Foster, I., Fox, P., Garcia, J., Kesselman, C., Markel, R., Middleton, D., Nefedova, V., Pouchard, L., Shoshani, A., Sim, A., Strand, G., and Williams, D. "The earth system Grid: Supporting the next generation of climate modeling research." *Proceedings of the IEEE* 93, 3 (2005), 485–495.
13. Birnholtz, J. "When do researchers collaborate? Toward a model of collaboration propensity." *Journal of the American Society for Information Science and Technology* (in press).
14. Borgatti, S.P., and Cross, R. "A relational view of information seeking and learning in social networks." *Management Science* 49, 4 (2003), 432–445.
15. Branscomb, L.M., and Florida, R. "Challenges to technology policy in a changing world." *Chemtech* 28, 6 (1998), 13–22.
16. Bullen, C., and Bennett, J. "Groupware in practice: An interpretation of work experiences." In Baecker, R. ed. *Readings in groupware and computer-supported cooperative work.* Morgan Kaufmann, San Mateo, CA, 1993, 69–84.
17. Cairncross, F. *The death of distance*. Harvard Business School Press, Cambridge, MA, 1997.
18. Cannon-Bowers, J.A., Salas, E., and Converse, S. "Shared mental models in expert team decision making." In Castellan, N.J., Jr. ed. *Individual and group decision making.* Lawrence Erlbaum Associates, Hillsdale, NJ, 1993, 221–246.
19. Cataldo, M., Wagstrom, P., Herbsleb, J.D., and Carley, K. "Identification of coordination requirements: Implications for the design of collaboration and awareness tools." In *ACM Conference on Computer-Supported Cooperative Work.* Banff, Canada, 2006, 353–362.
20. Catlett, C. et. al. "TeraGrid: Analysis of organization, system architecture, and middleware enabling new types of applications." In *High performance computing and grids in action*, 2007.
21. Chen, K.K., and O'Mahony, S. "The selective synthesis of competing logics." In *Academy of Management Proceedings* (Academy of Management, Atlanta, GA, 2006), L1–L6.
22. Coombs, R., and Georghiou, L. "Research and development: A new 'Industrial ecology'." *Science* 296, 5567 (2002), 471.
23. Cummings, J., and Kiesler, S. "Collaborative research across disciplinary and organizational

boundaries." *Social Studies of Science* 35, 5 (2005), 703–722.
24. Cummings, J.N., and Kiesler, S. "Coordination costs and project outcomes in multi-university collaborations." *Research Policy* 36, 10 (2007), 1620–1634.
25. Deelman, E., Callaghan, S., Field, E., Francoeur, H., Graves, R., Gupta, N., Gupta, V., Jordan, T.H., Kesselman, C., Maechling, P., Mehringer, J., Mehta, G., Okaya, D., Vahi, K., and Zhao, L. "Managing large-scale workflow execution from resource provisioning to provenance tracking: The CyberShake example." Paper presented at *e-Science 2006* (Amsterdam, December 4–6, 2006), http://www.isi.edu/~deelman/deelman_Ecybershake.pdf.
26. DeSanctis, G., and Monge, P. "Communication processes for virtual organizations." *Organization Science* 10, 6 (1999), 693–703.
27. Drabek, T.E., and McEntire, D.A. "Emergent phenomena and the sociology of disaster: Lessons, trends and opportunities from the research literature." *Disaster Prevention and Management* 12, 2 (2003), 97–112.
28. e-Infrastructure Reflection Group (e-IRG). "e-Infrastructures roadmap." e-IRGSP, The Hague, The Netherlands (2005), http://www.eirg.org/roadmap/eIRG-roadmap.pdf.
29. Ellisman, M., and Peltier, S. "Medical data federation: The biomedical informatics research network." In *The Grid: Blueprint for a new computing infrastructure (2nd edition)*. Morgan Kaufmann, San Francisco, CA, 2004.
30. Espinosa, J.A., Kraut, R.E., Slaughter, S.A., Lerch, J.F., Herbsleb, J.D., and Mockus, A. "Shared mental models, familiarity, and coordination: A multi-method study of distributed software teams." In *International Conference in Information Systems*. Barcelona, Spain, 2002, 425–433.
31. European Strategy Forum on Research Infrastructures (ESFRI). "European roadmap for research infrastructures: Report 2006." Office for Official Publications of the European Communities, Luxembourg (2006), ftp://ftp.cordis.europa.eu/pub/esfri/docs/esfri-roadmap-report-26092006_en.pdf.
32. Finholt, T., and Birnholtz, J. "If we build it, will they come? The cultural challenges of cyberinfrastructure." In Bainbridge, W., and Roco, M. eds. *Managing nano-bio-info-cogno innovations*. Springer, Berlin, 2006.
33. Foster, I. "Globus Toolkit version 4: Software for service-oriented systems." *Journal of Computer Science & Technology* 2, 4 (2006), 513–520.
34. Foster, I. "What is the Grid? A three point checklist." *GRIDToday* (July 20, 2002).
35. Foster, I., and Kesselman, C. "Scaling system-level science: Scientific exploration and IT implications." *IEEE Computer* November (2006), 32–39.
36. Foster, I., Kesselman, C., and Tuecke, S. "The anatomy of the Grid: Enabling scalable virtual organizations." *International Journal of Supercomputer Applications* 15, 3 (2001), 200–222.
37. Foster, I., et. al. "The Grid2003 production Grid: Principles and practice." Paper presented at *IEEE International Symposium on High Performance Distributed Computing* (Honolulu, HI, 2004), www.ivdgl.org/grid2003.
38. Friedman, T. *The world is flat: A brief history of the twenty-first century*. Picador, New York, 2005.
39. George, A.L., and Bennett, A. *Case studies and theory development in the social sciences*. MIT Press, Cambridge, MA, 2005.
40. Gerberding, J.L. "Faster…but fast enough?: Responding to the epidemic of severe acute respiratory syndrome." *New England Journal of Medicine* 348 (2003), 2030–2031.
41. Gersick, C.J.G. "Time and transition in work teams: Toward a new model of group development." *Academy of Management Journal* 31, 1 (1988), 9–41.
42. Gibson, C.B., and Gibbs, J.L. "Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation." *Administrative Science Quarterly* 51 (2006), 451–495.
43. Griffith, T.L., and Sawyer, J.E. "Supporting technologies and organizational practices for the transfer of knowledge in virtual environments." *Group Decision and Negotiation* 15 (2006), 407–423.
44. Griffith, T.L., Sawyer, J.E., and Neale, M.A. "Virtualness and knowledge in teams: Managing the love triangle of organizations, individuals, and information technology." *MIS Quarterly* 27, 2 (2003), 265–287.
45. Handy, C. "Trust and the virtual organization." *Harvard Business Review* 73, 3 (1995), 40–50.
46. Herbsleb, J.D., Atkins, D.L., Boyer, D.G., Handel, M., and Finholt, T.A. "Introducing instant messaging and chat in the workplace." In *Proceedings of the ACM Conference on Computer-Human Interaction* (ACM Press, Minneapolis, MN, 2002), 171–178.
47. Hey, T., and Trefethen, A. "E-science, cyberinfrastructure and Grid middleware services." *Science* 308, 5723 (2005), 817–821.
48. Hinds, P., and Kiesler, S., eds. *Distributed work*. MIT Press, Cambridge, MA, 2002.

49. Hollan, J., and Stornetta, S. "Beyond being there." In *Proceedings of the SIGCHI conference on human factors in computing systems* (ACM Press, Monterey, CA, 1992), 119–125.
50. Jarvenpaa, S.L., and Leidner, D.E. "Communication and trust in global virtual teams." *Organization Science* 10, 6 (1999), 791–815.
51. Jirotka, M., Procter, R., Rodden, T., and Bowker, G. "Special issue: Collaboration in e-Research." *CSCW* (2006).
52. Johansen, R. *Groupware: Computer support for business teams*. Free Press, New York, 1988.
53. Jordan, T.H., Maechling, P., and SCEC/CME Collaboration. "The SCEC community modeling environment: An information infrastructure for system-level earthquake science." *Seismological Research Letters* 74, 3 (2003), 324–328.
54. Karonis, N., Toonen, B., and Foster, I. "MPICH-G2: A Grid-enabled implementation of the message passing interface." *Journal of Parallel and Distributed Computing* 63, 5 (2003), 551–563.
55. Kesselman, C., Foster, I., and Prudhomme, T. "Distributed telepresence: The neesgrid earthquake engineering collaboratory." In *The Grid: Blueprint for a new computing infrastructure (2nd edition)*. Morgan Kaufmann, San Francisco, CA, 2004.
56. Kiesler, S., and Cummings, J. "What do we know about proximity and distance in work groups?" In Hinds, P., and Kiesler, S. eds. *Distributed work.* MIT Press, Cambridge, MA, 2002, 57–80.
57. Klein, K.J., and Kozlowski, S.W.J., eds. *Multilevel theory, research, and methods in organizations*. Jossey-Bass, San Francisco, CA, 2000.
58. Klimeck, G., McLennan, M., Mannino, M., Korkusinski, M., Heitzinger, C., Kennell, R., and Clark, S. "Nemo 3-D and nanoHUB: Bridging research and education." Paper presented at *Sixth IEEE Conference on Nanotechnology* (2006).
59. Knorr Cetina, K. *Epistemic cultures: How the sciences make knowledge*. Harvard University Press, Cambridge, MA, 1999.
60. Lave, J., and Wenger, E. *Situated learning: Legitimate peripheral participation*. Cambridge University Press, Cambridge, England, 1991.
61. Lawrence, K.A. "Walking the tightrope: The balancing acts of a large e-Research project." *Computer Supported Cooperative Work* 15 (2006), 385–411.
62. Lawrence, K.A., Finholt, T.A., and Kim, I. "Warm fronts and high pressure systems: Overcoming geographic dispersion in a meteorological cyberinfrastructure project." In *40th Hawaii International Conference on System Sciences (HICSS-40)*. IEEE Computer Society (Xplore Digital Library), Hawaii, 2007.
63. Lee, C.P., Dourish, P., and Mark, G. "The human infrastructure of cyberinfrastructure." In *Computer Supported Cooperative Work Conference (CSCW)*. Banff, Canada, 2006.
64. Lester, B. "Sustainable science: Greening the meeting." *Science* 318, 5847 (2007), 36–38.
65. Lewis, K. "Knowledge and performance in knowledge-worker teams: A longitudinal study of transactive memory systems." *Management Science* 50, 11 (2004), 1519–1533.
66. Lund, B., Hammond, T., Flack, M., and Hannay, T. "Social bookmarking tools (ii): A case study—connotea." *D-Lib Magazine* 11, 4 (2005).
67. Majchrzak, A., Jarvenpaa, S.L., and Hollingshead, A.B. "Coordinating expertise among emergent groups responding to disasters." *Organization Science* 18, 1 (2007), 147–161.
68. Markus, M.L. "The governance of free/open source software projects: Monolithic, multidimensional, or configurational?" *Journal of Management and Governance* 11, 2 (2007), 151–163.
69. Markus, M.L., Steinfield, C.W., Wigand, R.T., and Minton, G. "Industry-wide IS standardization as collective action: The case of the US residential mortgage industry." *MIS Quarterly* 30, Special Issue (2006), 439–465.
70. Maznevski, M.L., and Chudoba, K.M. "Bridging space over time: Global virtual team dynamics and effectiveness." *Organization Science* 11 (2000), 473–492.
71. McGrath, J., and Hollingshead, A.B. *Groups interacting with technology: Ideas, evidence, issues, and an agenda*. Sage, Thousand Oaks, CA, 1994.
72. Mortensen, M.A., and Hinds, P. "Fuzzy teams: Boundary disagreement in distributed and collocated teams." In Hinds, P., and Kiesler, S. eds. *Distributed work.* MIT Press, Cambridge, MA, 2002, 283–308.
73. National Science Board. *Science and engineering indicators–2004*. National Science Foundation, Arlington, VA, 2004.
74. Newman, M.E.J., Watts, D.J., and Strogatz, S.H. "Random graph models of social networks." In *Proceedings of the National Academy of Sciences* 99 (2002), 2566–2572.
75. NRC. "National collaboratories: Applying information technology for scientific research." National

Academy Press, Washington, DC, 1993, 7.
76. O'Leary, M.B., and Cummings, J.N. "The spatial, temporal, and configurational characteristics of geographic dispersion in teams." *MIS Quarterly* 31, 3 (2007), 433–452.
77. Oinn, T., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M.R., Wipat, A., and Li, P. "Taverna: A tool for the composition and enactment of bioinformatics workflows." *Bioinformatics* 20, 17 (2004), 3045–3054.
78. Olson, G.M., and Olson, J.S. "Distance matters." *Human-Computer Interaction* 15 (2000), 139–178.
79. Olson, G.M., Zimmerman, A., and Bos, N., eds. *Scientific collaboration on the Internet*. MIT Press, Cambridge, MA, in press.
80. Olson, J.S. "Bridging distance in collaborations: Lessons learned from a broad look at collaborations in science & engineering and in the corporate world." Paper presented at *National Science Foundation Collaborative Expedition Workshop No. 44—Pioneering Governance Mechanisms for Collaboration: Toward High-Performance Mission Delivery in a Networked World* (Arlington, VA, September 23, 2005), http://colab.cim3.net/file/work/Expedition_Workshop/2005-09-23_Pioneering_Governance_Mechanisms/Olson_2005_09_23.ppt and http://colab.cim3.net/file/work/Expedition_Workshop/2005-09-23_Pioneering_Governance_Mechanisms/Olson_Theory_of_Remote_Collaboration_2005_09_23.doc.
81. Olson, J.S., Bos, N., and Dahl, E. "Collaboratories at a glance." Paper presented at *Science of Collaboratories Comparative Investigation Workshop* (Ann Arbor, MI, September 25–27, 2002), http://www.scienceofcollaboratories.org/Workshops/WorkshopSept252002/index.php?Sept2002FinalSummary-Ca2G.
82. Olson, J.S., Hofer, E.C., Bos, N., Zimmerman, A., Olson, G.M., Cooney, D., and Faniel, I. "A theory of remote scientific collaboration (TORSC)." In Olson, G.M., Zimmerman, A., and Bos, N. eds. *Scientific collaboration on the Internet.* MIT Press, Cambridge, MA, in press.
83. Olson, J.S., Olson, G.M., and Meader, D.K. "What mix of video and audio is useful for small groups doing remote real-time design work?" In *SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, New York, NY, 1995, 362–368.
84. Orlikowski, W.J., and Iacono, C.S. "Research commentary: Desperately seeking the 'IT' In IT research—a call to theorizing the IT artifact." *Information Systems Research* 12, 2 (2001), 121–134.
85. Oster, S., Hastings, S.L., Langella, S., Ervin, D.W., Madduri, R., Kurc, T.M., Siebenlist, F., Foster, I., Shanbhag, K., Covitz, P.A., and Saltz, J.H. "caGrid 1.0: A Grid enterprise architecture for cancer research." In *Proceedings of the 2007 AMIA Annual Symposium* (Chicago, IL, 2007).
86. Oster, S., Langella, S., Hastings, S., Ervin, D., Madduri, R., Phillips, J., Kurc, T., Siebenlist, F., Covitz, P., Shanbhag, K., Foster, I., and Saltz, J. "caGrid 1.0: An enterprise Grid infrastructure for biomedical research." *Journal of the America Medical Informatics Association* (2008).
87. Porac, J.F., Rao, H., Weick, K.E., and Lawrence, K.A. "Distributed sensemaking across interorganizational networks: The diagnosis of the West Nile virus." Paper presented at *National Academy of Management Meeting* (Seattle, WA, August, 2003).
88. Pordes, R., Petravick, D., Kramer, B., Olson, D., Livny, M., Roy, A., Avery, P., Blackburn, K., Wenaus, T., Würthwein, F., Foster, I., Gardner, R., Wilde, M., Blatecky, A., McGee, J., and Quick, R. "The open science Grid." Paper presented at *Scientific Discovery through Advanced Computing (SciDAC) Conference* (2007).
89. Preece, J. *Online communities: Designing usability and supporting sociability*. John Wiley & Sons, Ltd, Chichester, England, 2000.
90. Ragin, C.C. *The comparative method: Moving beyond qualitative and quantitative strategies*. University of California Press, Berkeley, CA, 1987.
91. Ragin, C.C. *Fuzzy-set social science*. University of Chicago Press, Chicago, IL, 2000.
92. Ribes, D., and Finholt, T.A. "Tensions across the scales: Planning infrastructure for the long term." In *Proceedings of GROUP 2007 Conference* (ACM Press, Sanibel Island, FL, 2007).
93. Robinson, J.P., and Neustadtl, A. "Social psychology of internet." In Christensen, K., and Levinson, D. eds. *Encyclopedia of community: From the village to the virtual world.* Sage Publications, Thousand Oaks, CA, 2003, 779–785.
94. Rocco, E. "Trust breaks down in electronic contexts but can be repaired by some initial face-to-face contact." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM Press, Los Angeles, CA, 1998), 496–502.
95. Rocco, E., Finholt, T.A., Hofer, E.C., and Herbsleb, J.D. "Out of sight, short of trust." Paper presented at *Founding Conference of the European Academy of Management* (Barcelona, Spain, 2001).
96. Spencer, B.F., Jr., Butler, R., Ricker, K., Marcusiu, D., Finholt, T., Foster, I., and Kesselman, C.

"Cyberenvironment project management: Lessons learned." (National Science Foundation, Washington, DC, 2006), http://neesgrid.ncsa.uiuc.edu/documents/CPMLL.pdf.

97. Thain, D., Tannenbaum, T., and Livney, M. "Condor and the Grid." In Berman, F., Fox, G., and Hey, T. eds. *Grid computing: Making the global infrastructure a reality.* John Wiley & Sons Inc., 2002.
98. Tierney, K.J., Lindell, M.K., and Perry, R.W. *Facing the unexpected: Disaster preparedness and response in the United States*. Joseph Henry Press, Washington, DC, 2001.
99. Tuckman, B. "Developmental sequence in small groups." *Psychological Bulletin* 63 (1965), 384–399.
100. Turkle, S. *Life on the screen: Identity in the age of the Internet*. Simon and Schuster, New York, 1995.
101. United States General Accounting Office. "West Nile virus outbreak: Lessons for public health preparedness." U.S. General Accounting Office, Washington, DC (September, 2000), www.gao.gov.
102. van Ahn, L. "Games with a purpose." *IEEE Computer* June (2006), 96–98.
103. von Eschenbach, A.C., and Buetow, K. "Cancer informatics vision: caBIG™." *Cancer Informatics* 2 (2006), 22–24.
104. Vonortas, N.A. *Cooperation in research and development*. Kluwer, Boston, 1997.
105. Wulf, W.A. "The collaboratory opportunity." *Science* 261, 5123 (1993), 854–855.
106. Zammuto, R.F., Griffith, T.L., Majchrzak, A., Dougherty, D.J., and Faraj, S. Information technology and the changing fabric of organization." *Organization Science* 18, 5 (2007), 749–762.
107. Zhao, Y., Hategan, M., Clifford, B., Foster, I., von Laszewski, G., Nefedova, V., Raicu, I., Stef-Praun, T., and Wilde, M. "Swift: Fast, reliable, loosely coupled parallel computation." In *IEEE Congress on Services (Services 2007)*. IEEE Computer Society, Salt Lake City, UT, 2007, 199–206.
108. Zimmerman, A., and Nardi, B.A. "Whither or whether HCI: Requirements analysis for multi-sited, multi-user cyberinfrastructures." In *CHI 2006.* ACM Press, Montreal, Quebec, Canada, 2006, 1601–1606.

8. APPENDIX A: ABBREVIATIONS AND ACRONYMS

ADHO	Alliance of Digital Humanities Organizations
Amazon EC2	Amazon Elastic Compute Cloud
Amazon S3	Amazon Simple Storage Service
API	Application Programming Interface
BIRN	Biomedical Informatics Research Network
caBIG™	cancer Biomedical Informatics Grid
Calit2	California Institute for Telecommunications and Information Technology
CDC	Centers for Disease Control and Prevention
CDI	Cyber-enabled Discovery and Innovation
CERN	European Organization for Nuclear Research
CI	Cyberinfrastructure
CluE	The Cluster Exploratory
CSCW	Computer-Supported Cooperative Work
DOE	U.S. Department of Energy
ESG	Earth System Grid
ETPs	European Technology Platforms
ETSI	European Telecommunications Standards Institute
EVO	Enabling Virtual Organizations
FLOPS	FLoating-point Operations Per Second
GEON	The Geosciences Network
GLORIAD	Global Ring Network for Advanced Application Development
GNU	GNU's Not Unix
HASTAC	Humanities, Arts, Science, and Technology Advanced Collaboratory
HIPAA	Health Insurance Portability and Accountability Act
HPC	High-Performance Computing
IRB	Institutional Review Board
Knowbot	Knowledge Based Object Technology.
LEAD	Linked Environments for Atmospheric Discovery
LHC	Large Hadron Collider
LPP	Legitimate Peripheral Participation
MREFC	Major Research Equipment and Facilities Construction
NASA	National Aeronautics and Space Administration
NCI	National Cancer Institute
NCSA	National Center for Supercomputing Applications
NEES	Network for Earthquake Engineering Simulation
NEON	National Ecological Observatory Network
OGF	Open Grid Forum
OOI	Ocean Observatories Initiative
OSG	Open Science Grid
OptIPuter	*Opt*ical networking *I*nternet *P*rotocol comp*uter*
PRAGMA	Pacific Rim Applications and Grid Middleware Assembly
SARS	Severe Acute Respiratory Syndrome
SCEC	Southern California Earthquake Center
SDSS	Sloan Digital Sky Survey
SL	Second Life
SoC	Science of Collaboratories
SRB	Storage Resource Broker
SQL	Structured Query Language
SVG	Scaleable Vector Graphics
TORSC	Theory of Remote Scientific Collaboration
USGS	U.S. Geological Survey
VO	Virtual Organization
WATERS Network	WATer and Environmental Research Systems Network
W3C	World Wide Web Consortium

9. APPENDIX B: DESIGNING, BUILDING, AND EVALUATING VIRTUAL ORGANIZATIONS SEPTEMBER 18-19, 2007 WORKSHOP PARTICIPANTS

Name	Institution
Alvarez, Heidi	University of Illinois, Urbana-Champaign
Argote, Linda	Florida International University
Atkins, Dan	Carnegie Mellon University
Avery, Paul	National Science Foundation
Baru, Chaitan	University of Florida
Bender, John	San Diego Supercomputer Center
Beroza, Greg	Stanford University
Birnholtz, Jeremy	Stanford University
Butler, Brian	Cornell University
Clark, Allison	University of Pittsburgh
Cogburn, Derrick	Syracuse University
Contractor, Nosh	University of Illinois, Urbana-Champaign
Crowston, Kevin	Syracuse University
Cummings, Jonathon	Duke University
Davis, Rebecca	National Institute for Technology and Liberal Education
Delaney, John	University of Washington
Espinosa, Alberto	American University
Fernandez-Gonzalez, Mercedes	Atos Origin
Finholt, Tom	University of Michigan
Foster, Ian	Argonne National Laboratory and University of Chicago
Fulk, Janet	University of Southern California
Giles, Roscoe	Boston University
Grethe, Jeff	University of California, San Diego
Griffith, Terri	Santa Clara University
Halverson, Cristine	IBM
Herbsleb, Jim	Carnegie Mellon University
Hollingshead, Andrea	University of Southern California
Jarvenpaa, Sirkka	University of Texas
Kesselman, Carl	University of Southern California/ISI
Klimeck, Gerhard	Purdue University
Lane, Julia	University of Chicago
Lee, Charlotte	University of California, Irvine
Markus, M. Lynne	Bentley College
Minsker, Barbara	University of Illinois, Urbana-Champaign
Myers, Jim	National Center for Supercomputing Applications (NCSA)
O'Leary, Michael	Boston College
Orlikowski, Wanda	Massachusetts Institute of Technology
Reed, Dan	Renaissance Computing Institute (RENCI)
Rhoten, Diana	National Science Foundation
Siebenlist, Frank	Argonne National Laboratory and University of Chicago
Strong, Paul	eBay
Wladawsky-Berger, Irving	IBM and MIT

10. APPENDIX C: DIFFERENCES BETWEEN TRADITIONAL AND COMPLEX SYSTEMS

At the Building Effective Virtual Organizations workshop in January 2008, William B. Rouse of Georgia Tech's Tennenbaum Institute gave a keynote talk titled "Complexity and Organizations: Challenges for Virtual Organizations." In this talk, he highlighted how the shift from traditional to networked organizations implies corresponding shifts in how we value, design, and manage these new systems. VOs are a specific instance of complex networks and should consider the significant shifts necessitated by the network model. The table below contrasts a few of these shifts.

	Traditional system	**Complex network**
Roles	Management	Leadership
Methods	Command and control	Incentives and inhibitions
Measurement	Activities	Outcomes
Focus	Efficiency	Agility
Relationships	Contractual	Personal Commitments
Network (how things get done)	Hierarchy	Heterarchy
Design	Organizational design	Self-organization

11. APPENDIX D: "VOLUNTEER" PROJECTS AND POTENTIAL IDEAS FOR IMMEDIATE RESEARCH

During the September workshop, participants identified a variety of potential areas and projects for future research on virtual organization (VOs). Specific research challenges that could bear greatly on the near future include biodiversity and environmental conservation, preparation for natural disasters such as hurricanes and earthquakes, and population growth. Participants also saw a need for vastly improving education outreach in hard-to-reach places through the use of VOs.

Some participants identified their specific projects as venues for VO research. These projects are described briefly in this section.

Center for High Energy Physics Research and Education Outreach (CHEPREO; http://www.chepreo.org/): This interregional grid-enabled center at Florida International University, one of the largest minority schools in the United States, is a physical organization as well as a VO with participation from Florida State University, the University of Florida, the California Institute of Technology, and the Brazilian high-energy physics community. This hybrid case study includes an integrated program of research, cyberinfrastructure, and education and outreach.

The Global CyberBridges project (GCB; http://www.cyberbridges.net/): This initiative is designed to address the problem of inadequate adoption and use of cyberinfrastructure (CI) by bridging the divide between the information technology and science communities. It is a U.S. implementation of multinational efforts to improve technology training for scientists so that they understand the potential of CI, and likewise CI professionals can work more effectively with domain scientists. In addition to the U.S. partners, the project has committed participation from the Computer Network Information Center of the Chinese Academy of Sciences, the City University of Hong Kong, and the University of Sao Paulo's School of the Future of Brazil.

The Southern California Earthquake Center (SCEC; http://www.scec.org/): This collaboratory (also described in the main report) spans academia, government, and private industry. Together, these partners are trying to improve understanding of earthquakes and their effects. A primary science goal of the SCEC collaboration is to transform seismic hazard analysis into a physics-based science through high-performance computing and communications. The underlying physics is multiscale and highly nonlinear, creating a tremendous need for computation to address these questions, thereby requiring input from a wide range of disciplines. In particular, the SCEC community is working toward improvements of understanding the effects of earthquakes on structures by augmenting data with simulations. This has the advantage of incorporating existing knowledge about earthquakes directly into the process. Specific significant issues include high-level questions about how the ground shakes, as well as how to respond effectively to earthquake disasters, how to involve diverse contributors to the seismic hazard analysis community, and how to preserve the knowledge for use across disciplinary boundaries. (See sidebar 1.)

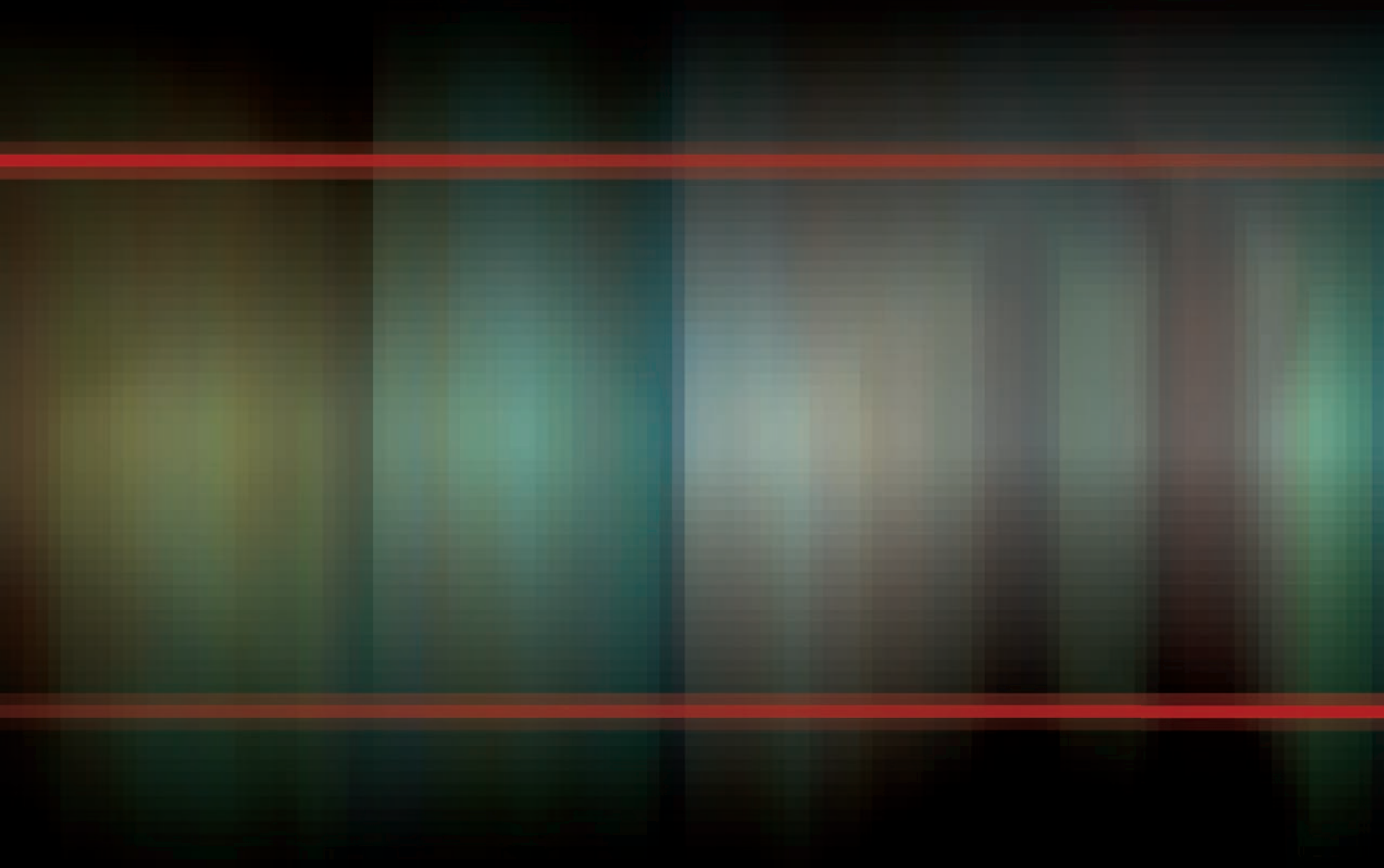
WATer and Environmental Research Systems Network (WATERS Network; http://www.watersnet.org): This project for water research (described in the main report) brings together environmental engineers and hydrologic scientists to implement an integrated, real-time, distributed observation system that will address deficiencies in our current scientific understanding of the dynamics and spatial variability of water processes. By incorporating networked sensors, assimilation of high-frequency data, and interdisciplinary experimentation, the network will identify how water quantity, quality, and related earth system processes are affected by natural and human-induced changes to the environment. (See sidebar 16)

Globus Toolkit (http://www.globus.org): This open-source software project is a VO, with developers at multiple institutions. It also provides infrastructure software used in many virtual organizations (see sidebar 15). Globus deployments incorporate opt-out "usage reporting" mechanisms that provide anonymizedreports on usage. The resulting data provide an opportunity to study the processes of CI adoption.

12. IMAGE CREDITS

Image	Credit
Cover	Paul Woodward, Laboratory for Computational Science and Engineering, University of Minnesota
Exhibit 1	*Top*, Air Traffic, Aaron Koblin
Exhibit 1	*Bottom*, Router Connections, Zina Deretsky, National Science Foundation, and Chris Harrison, Human-Computer Interaction Institute, Carnegie Mellon University
Exhibit 2	Visualization by Amit Chourasia, San Diego Supercomputer Center, UCSD; data by Robert Graves and the Southern California Earthquake Center
Exhibit 3	National Cancer Institute, Cancer Biomedical Informatics Grid (caBIG™) Initiative
Exhibit 4	Gary Strand, National Center for Atmospheric Research (NCAR)
Exhibit 5	Courtesy of CERN
Exhibit 6	Mark Lundstrom, Director, Network for Computational Nanotechnology, Purdue University
Exhibit 7	Image by Thomas Deerinck and Mark Ellisman from the work of the National Center for Microscopy and Imaging Research at UCSD
Exhibit 8	Courtesy of HASTAC
Exhibit 9	Fermilab
Exhibit 10	© ART+COM
Exhibit 11	Daniel E. Atkins, National Science Foundation
Exhibit 12	Jurgen Schulze, Calit2, University of California, San Diego
Exhibit 13	Nicolle Rager Fuller, National Science Foundation
Exhibit 14	Consortium of Universities for the Advancement of Hydrologic Science, Inc (CUAHSI)
Exhibit 15	© 2008 JupiterImages Corporation
Exhibit 16	*Top*, Penguin photo, © 2008 JupiterImages Corporation
Exhibit 16	*Bottom*, Linux Penguin, Larry Ewing lewing@isc.tamu.edu; General Image Manipulation Program (The GIMP)
Exhibit 17	Department of Computer Science, University of North Carolina, Chapel Hill
Exhibit 18	Second Life is a trademark of Linden Research, Inc. Certain materials have been reproduced with the permission of Linden Research, Inc. Copyright © 2001-2008 Linden Research, Inc. All rights reserved.
Exhibit 19	© 2004-2008 Blizzard Entertainment, Inc. All rights reserved
Exhibit 20	Courtesy of Indiana University, based on illustration by Nicolle Rager Fuller, National Science Foundation
Exhibit 21	Courtesy of Peter Ginter
Exhibit 22	Facebook © 2008
Exhibit 23	Courtesy of myExperiment www.myexperiment.org
Exhibit 24	National Center for High-performance Computing, Taiwan
Exhibit 25	© 2008 JupiterImages Corporation
Exhibit 26	John Sondek, University of North Carolina, Chapel Hill
Exhibit 27	Renaissance Computing Institute
Exhibit 28	The Globus Alliance
Exhibit 29	Ernest To, Center for Research in Water Resources (CRWR), University of Texas at Austin; Yong Liu, National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign; © 2008 Google
Exhibit 30	Richard Rogers and the Govcom.org Foundation, Amsterdam
Exhibit 31	© 2002 The Regents of the University of California
Exhibit 32	Jeffrey Heer, Computer Science Division, University of California at Berkeley

Exhibit 33	http://www.innocentive.com/
Exhibit 34	TORSC Team, University of Michigan
Exhibit 35	*Left*, Printing press, factory Library of Congress, LC-USZ62-131927
Exhibit 35	*Middle*, Phone © 2008 JupiterImages Corporation
Exhibit 35	*Right*, Mail © 2008 JupiterImages Corporation
Exhibit 36	Research and node layout by Kevin Boyack and Dick Klavans. Data from Thompson ISI; graphics & typography by W. Bradford Paley. Commissioned and partially supported by Katy Borner and the Places and Spaces: Mapping Science exhibition. Copyright (c) 2006 W. Bradford Paley, all rights reserved.
Exhibit 37	*Left*, Supercomputer © University Corporation for Atmospheric Research
Exhibit 37	*Right*, Laptop © 2008 JupiterImages Corporation
Exhibit 38	© 2008 JupiterImages Corporation
Exhibit 39	Zina Deretsky, National Science Foundation
Graphic Design	Adrian Apodaca, National Science Foundation



This work was supported by the National Science Foundation under Award Nos. 0751539 and 0816932. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.