Improving International Biodiversity Research Collaboration with Cyberinfrastructure

A Report of the Workshop

"Cyberinfrastructure for International Biodiversity Research Collaboration"

Held in City of Knowledge, Panama, on January 10-13, 2006

James H. Beach
Biodiversity Research Center
University of Kansas
Lawrence, Kansas

Julio E. Ibarra
Center for Internet Augmented Research and Assessment
Florida International University
Miami, Florida

18 December 2007

Funded by the U.S. National Science Foundation
Office of International Science and Engineering, Award #0549443.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Biodiversity scientists estimate that between 5 and 10 million species inhabit the planet; some think, with the inclusion of recent insect estimates and marine and soil microbes, that the number could be as high as 30 million. With the unprecedented pressure on wild populations from human population growth and its collateral impacts, such as climate change, systematists and taxonomists today have concluded that to even identify and minimally describe a modest fraction of those species before they are extinct, traditional biological inventory techniques must be completely transformed in order to greatly accelerate progress (Causey et al., 2004; Wheeler et al., 2004)

The enormity and finality of the task are daunting and sobering. And yet, researchers must chronicle the diversity of life on earth to understand how to maintain critical levels of biological complexity to sustain human life and the natural systems upon which we completely depend.

Except for one species, organisms do not recognize national boundaries. For that reason, biological diversity and ecological research have always been global enterprises; international research collaboration and training within systematics and ecology has been a longstanding practice. Now, the emergence of international data network communications, specifically high-bandwidth, optical fiber links, provides an opportunity to accelerate the pace and efficiency of biodiversity research within this tradition of international collaboration. Some of the most important grand challenges in this science domain this century will be to rapidly develop the human capital for urgent biodiversity research, to invent new methods of efficient, international collaboration, and to deliver as much knowledge as possible about the contents and health of earth's remaining wild areas. International cyberinfrastructure (CI), particularly high-bandwidth research and education networks will be a foundation for transforming traditional methods in order to meet these pressing, global objectives.

An NSF-funded workshop (NSF/OISE #0549443) held in Panama in January, 2006 brought together biodiversity scientists, network engineers, research policy-makers, and funding agency representatives from throughout the western hemisphere to examine the ways in which international collaborations among biodiversity researchers could be improved through CI. Numerous challenges were identified. Among them were the lack of adequate access to CI in some countries in Central and South America; a reluctance to share data due to a perceived atmosphere of competition; and the absence of adequate standard practices for the discovery, retrieval, and description of existing datasets. Also prominent was the challenge of how to find an international CI collaboration mechanism to build research communities based on shared data and computational resources. Human networking, bringing people together to establish common ground and a shared research vision, and most of all mutual trust, were all seen as challenges to be addressed before the deployment of CI-based collaborative efforts can take place. Moving traditional field-based collaborations, which operate on a time-scale of field seasons and deliver joint publications months or years later, to modern, CI-based data collecting, analysis, integration, and synthesis, which can operate on scales of seconds and result in essentially immediate communication, will require fundamental changes in the way biodiversity scientists value collaboration, interact, and trust one another.

Recommendations posed by workshop participants included increasing funding in countries where CI is lacking. By bringing together five funding agencies, the workshop provided a forum for discussion of potential cost-sharing. Relationships between limited funding and the absence of trust for collaborators were discussed, as competition for scarce research funding creates disincentives for data sharing and shared CI. To build trust, workshop participants suggested that domain workshops be held in which researchers could work shoulder-to-shoulder to identify shared value and common interests. They also suggested that CI-based collaborative communities should nucleate as small groups and add members slowly as accomplishments and trust build. Participants agreed that barriers to trust will fade if students are trained to collaborate, and that increased trust can lead to greater research efficiency and success and can be practically accomplished with scientists who are distantly remote. More detailed descriptions of participants' challenges and recommendations are described below.

Also discussed are three follow-on activities that will have resulted as a consequence of the workshop. One of the activities is a NSF-sponsored "Pan-American Advanced Studies Institute" or "PASI". The PASI's goal is to give students and young faculty from North, South, and Central America the opportunity to learn about CI tools and their application to biodiversity and ecological research through two weeks of lectures, hands-on technical laboratory exercises, and group discussion. The "Cyberinfrastructure for International, Collaborative Biodiversity and Ecological Informatics" PASI will be held in June 2008, in Costa Rica" Another significant follow-on activity is a one-year, international, collaborative analysis by environmental scientists and network engineers from the U.S. and Central America who will evaluate the research and education requirements for an environmental sensor network in Costa Rica. Third, the "Pan-American Sensors for Environmental Observatories (PASEO)" workshop held in Bahia Blanca, Argentina in June 2007 was an indirect byproduct of the Panama biodiversity workshop.

A clearinghouse for ideas and anlysis, the Panama workshop documented the challenges of scientists throughout the Americas for collaborative research through networked applications utilizing distributed expertise, data, and computational resources. Meeting participants agreed that international collaboration through CI will be essential to address the large, pressing environmental research questions of our day.

To increase the use and effectiveness of CI for international biodiversity and ecological research collaborations, we recommend an ongoing mechanism for building trust, developing a shared CI-based research vision, and improving graduate education in the use of network-based CI tools. A series of domain-focused workshops designed to build community would address those objectives and accelerate the effective utilization of international, network-based methods for biodiversity and ecological research collaboration.

INTRODUCTION

Global Environmental Research: Increasingly Urgent and Complex

Approximately 1.8 million species have been documented as a result of 300 years of exploration on the planet. The number is only a fraction of the total, which most scientists estimate to be between 5 and 10 million; estimates as large as 30 to 80 million have been suggested. Of the 1.8 million known species, more than 15,000 are threatened with extinction and many more may go extinct before they are identified or described. A grand challenge for 21st century scientists is to gain knowledge of this biological diversity before it is lost. That knowledge is critical for understanding and sustaining the natural systems that support the quality and future stability of human life.

In addition to the desperate need for more data from the field, environmental biologists also are faced with the challenge of effectively analyzing digital data which are widely-distributed in heterogeneous globally-distributed data archives. Traditional research practice in ecology, systematics, biogeography, and phylogenetics, has always been global in scope and context, but with paper-based methods, collaborations have been mediated through paper-based reprints and publications. With the elimination of many constraints of time and space (and resource discovery) by instantaneous data communication capabilities, international collaborators must now surmount new, social, technological and conceptual barriers inherent in the media to be able to accelerate their research agendas to contribute the analysis and synthesis society urgently requires for understanding the value and role of biological biodiversity in sustaining natural systems and life on earth.

New network-based research techniques, including rapid, molecular field assessments, labeling and managing organism identities with DNA bar codes, and development of cross-disciplinary, data discovery, retrieval, analysis, modeling, publication, and social networking tools will all be needed to leverage international CI. (Wilson 2000, 2003; Godfray 2002; and Herbert 2003).

Nascent Cyberinfrastructure (CI) in the Americas

While the methods and time course of biodiversity research investigation are changing, communications network technology is simultaneously undergoing transformation. North and South America are now connected by dozens of wide-bandwidth fiber-optic cables that transport research data. These networks enable wide-bandwidth digital communications between the United States' networks, including Internet2, National Lambda Rail (NLR), and the National Research and Education Networks (NRENs) of South and Central America, the Caribbean, and Mexico.

Constructed with a goal of revolutionizing the way research and education is conducted, high-speed networks can change the way scientists think about problems (Robertson, 2003). With CI improvements to communication quality and technology, teams of scientists and technologists are increasingly able to investigate larger and more complex questions. Researchers in physics and astronomy have made enormous strides in their work by taking advantage of international networks. For example, more than 2,000 geographically dispersed physicists are using the networks to collaborate on a project involving the Large Hadron Collider, a massive particle accelerator situated on the Franco-Swiss border. The team is investigating some of the universe's greatest mysteries, such as whether there are dimensions to the Universe beyond 3-dimensional space and 1-dimensional time.

International Biodiversity Collaboration in a Networked World

While many researchers are advancing their knowledge through network-enabled international collaborations, biologists are just beginning to capture the power of the Internet for collaborative analysis of biodiversity data (Bisby, 2000, 2002; Edwards et al., 2000; Krishtalka, 2002; and Soberon and Peterson, 2004). This could be due to a lack of awareness of the possibilities held by networked grid computing and/or to a lack of access to the networks. Although the network infrastructure exists, many South and Central American universities and research organizations have not connected. Such gaps may slow the progress of scientists, who often wait hours to download or send information.

Because the data is distributed across the globe, biodiversity informatics, or the sharing of information regarding species and communities, will benefit tremendously from increased research and education network connectivity. As biologists tackle larger and more complex research questions, they will increasingly rely on CI to achieve their goals. Much of biodiversity informatics is already performed in a virtual environment – large datasets with hundreds of gigabytes of data, such as whole genome sequences, are available online – yet with full use of high-speed international research and education networks, the field can be transformed into one in which virtual voucher specimens can be accessed from any museum at the touch of a button, for example.

Networked computational grids are already creating a major paradigm shift for the field of biology. The Global Biodiversity Information Facility (GBIF) is one organization that is leading the way. Backed by numerous governments, GBIF's mandate is to develop standards and tools for global biodiversity data so that people from all countries can benefit from the use of this information. The Pacific Rim Application and Grid Middleware Assembly (PRAGMA), which aims to improve the use of grid technologies by communities of scientific collaborators, is another leader in bringing together people to establish common vision and trust needed for collaborative international research.

Promoting Awareness of the Need for CI in Biodiversity Research

Central America is one of the most biologically diverse regions in the world and one of the most underrepresented in terms of access to international CI. Our NSF-funded workshop held in Ciudad de Saber, Panama brought together environmental biologists, national research network technologists, policy-makers, and funding agency representatives from throughout the Western hemisphere to examine the ways in which international collaborations among biodiversity researchers, especially those in Central America, could be improved through international CI.

Numerous challenges were identified by participants. Among them were the lack of adequate CI in some Central American countries; a reluctance to share data due to a perceived atmosphere of competition; the need for metadata, or information about how, when, and where data were collected; and the need for a mechanism to carry out workshop recommendations. Recommendations posed by workshop participants included increasing funding in countries where access to CI is poor; bringing together funding agencies in order to establish cost-sharing; and building trust among researchers. A detailed description of participants' concerns and recommendations follows.

WORKSHOP FINDINGS

The Panama meeting was organized into plenary sessions and break-out groups. We had three plenary sessions with four break-out discussion groups, which met after each plenary. In Session 1, each of the four groups discussed opportunities for collaboration in biodiversity and ecological research with regard to informatics and CI. In Session 2, they discussed funding for biodiversity and ecological research with regard to informatics and CI. And in Session 3, participants discussed organizational structures for biodiversity and ecological research collaborations. Panel discussions also were held. Issues covered included the CI tools that are available or emerging to support collaborative biodiversity research among Central American nations, the U.S., and other countries; best practices and success stories for applying CI to collaborative research among teams of distributed researchers; and issues and challenges faced when applying CI to environmental research and other science domains.

CI Challenges

The foundation for a new age of scientific exploration has been laid. High-speed networks now connect numerous countries throughout the world, enabling scientists to collaborate in ways that were never before possible. In recent years, several initiatives have been undertaken to improve the communications infrastructure between the United States and South and Central America. Thus far, 12 countries in this region have connected to the network. But even with the potential to connect hundreds of biologists from across the Americas, these networks remain underutilized. Workshop participants cited a number of reasons for why international research and education networks have not met their full potential. Gaps in the network, under-appreciation of network capabilities, and data management issues all were identified as major CI challenges because of technical, social, and economic factors.

Network gaps were one of the most pressing challenges raised by Central American scientists. They noted that although the international networks land in their countries, many universities and research organizations have not connected to them. For example, Panama is a major network hub connecting the eastern and western hemispheres. Yet, few Panamanian research institutions are plugged into this hub. The lack of access may slow the progress of scientists, who sometimes wait hours to download or send information, while their colleagues in the United States, Europe, and elsewhere, who have access to high-speed networks, forge ahead with their discoveries. As the processes of science become transformed by CI, there is a concern that the Central American research and education communities will be challenged to share in this transformation, because of these gaps in the infrastructure.

Increasing scientists' access to the networks will do no good if researchers don't appreciate the potential that CI can bring to their research. While physicists, astronomers, and other scientists have significantly advanced their fields through CI, many biologists fail to see the utility of the networks. Nevertheless, some biological sub-fields are turning to network technology to aid in data acquisition and sharing. For example, several museums and other organizations that curate voucher collections and ecological datasets have embraced structured databases and web data exchange technologies as a means of providing remote researchers with access to information. In addition, environmental sensor networks are becoming increasingly relevant to ecological and biodiversity research and with the international deployment of high bandwidth research and education networks, they are likely to become a significant part of multi-national collaboration efforts.

Groups like GBIF and the Inter-American Biodiversity Information Network (IABIN) – which supports a decentralized, Internet-based, western hemisphere network that provides access to biodiversity information resources – help to educate biodiversity researchers about the usefulness of networks and help to raise awareness about potential gaps in network access. These organizations also are charged with developing logical and organized data management practices, which was another issue discussed by workshop participants. Currently, such standardized data management practices are not widely used, which may be due to the high costs associated with their employment.

Challenges

- Gaps in the network
- Under-appreciation of network capabilities
- Data management issues

Lack of Awareness

An effective lack of access isn't the only thing preventing scientists from using the international networks for regional collaborations. A lack of awareness about the potential for CI to transform research is equally problematic. Biologists have only just begun to capture the power of the Internet for collaborative analysis of biodiversity data (Bisby, 2000, 2002; Edwards et al., 2000; Krishtalka, 2002; and Soberon and Peterson, 2004). While traditional biodiversity inventory and analysis often consisted of identifying species using a dichotomous key, incrementally amassing voucher specimens for museums and herbaria, and manually collecting basic ecological data, much of today's biodiversity research employs molecular tools for species identification, virtual collections of voucher specimens, including tissues and DNA samples, and predictive computer models that simulate and forecast organisms' responses to a variety of ecological conditions. As biodiversity research becomes more data- and model-driven, requirements for computational and network resources will naturally increase, but computational thinking, particularly based on the use of international network infrastructure, is still in incipient stages. The threshold where network computing and workflows become an integral part of biodiversity analysis and synthesis has yet to be crossed by many.

In addition to incorporating more advanced technologies into their work, biodiversity researchers also are broadening the spatial scopes of their studies to landscape, regional, and continental scales. As local studies suggest broader and more general ecological processes are at work, the need for validation of data across multiple places increases. Yet, few studies can afford to collect additional data at tens or hundreds of sites. While the process doesn't scale, the capability to collaborate via high-speed networks breaks through the costs of travel logistics. Many scientists are unaware of the data-sharing capabilities that such networks afford.

Additionally, many biologists are only partially aware of the research interests and parallel research efforts being undertaken by colleagues across the Western Hemisphere. They may not realize the extent of overlap that their research has with these people, and therefore, may fail to develop international collaborations. Because they have traditionally collaborated with researchers in the U.S. and Europe, scientists from Central American countries have particularly limited collaborations with one another. This may be due to a lack of incentives and/or lack of affinity to collaborate. For example, many of the Costa Rican biologists that attended the conference were unaware of parallel efforts being undertaken in Colombia. Such ignorance of the work taking place in neighboring countries can lead to duplication of effort and incompatibility of datasets and computational methods. This can result in greater discontinuity as a result of gaps and the lack of coherence in aspects of work, which further complicates scientific collaborations (Watson-Manheim, 2002).

Challenges

- Under-appreciation of network capabilities
- Ignorance of research taking place by other researchers and in other countries

Limited Trust

An inability to establish trust relationships through entirely online interactions was frequently mentioned by workshop participants as a problem that inhibits initializing collaborations. Trust has been shown to be a key factor in the initiation and maintenance of virtual organizations (Kasper-Fuehrer, 2003-04) and collaborative work (Kouzes, 1996; Olson, 2001). It allows participants to feel comfortable sharing their ideas and concerns and enables work to be accomplished that would be impossible for one person alone to achieve. Participants in a collaborative relationship also must be given the opportunity to meet each of the individuals who will comprise their collaborative group. They must then verify that their own interests are aligned with the interests of the group (Olson et al., 2001). Yet opportunities to meet others scientists are often limited and require funding to travel.

As researchers are forced to compete for small and inconsistently available pools of money, they are also less likely to share data for fear of losing their competitive edge. The sparseness or lack of social networks also was mentioned as a trust inhibitor since it is difficult to build trust without having face-to-face encounters (Olson et al., 2001).

Challenges
• Lack of opportunities to meet other scientists • Limited funding promotes competition and lack of trust

Lack of Incentives

Most biologists are motivated by the desire to acquire knowledge about the natural world, but they cannot pursue their interests without funding. They also may require recognition for their work to feel motivated. While both of these incentives encourage researchers to continue to make great discoveries, incentives for collaboration on such research endeavors have, in the past, been limited, and incentives for using CI to advance research goals are even weaker (Foster, 2005). The fact that international network-based collaborations require much more “activation energy” than working with researchers down the hall, across campus or across the country (Cummings and Keisler, 2005) exacerbates this problem. Workshop participants identified a lack of incentives as a barrier to both collaboration and the adoption of CI-enabled community research efforts.

Another disincentive for pursuing projects that involve international collaborators working with new technologies may be a lack of institutional support. Institutional cultures can sometimes act as barriers preventing scientists from pursuing opportunities that involve international collaborations and new technologies. Promotion and tenure committees, for example, may not see such activities as valuable and may deny promotions to scientists who engage in this type of work. To increase the number of scientists engaging in international collaborations, these disincentives must be removed.

As previously mentioned, international collaborations can be particularly difficult to establish and maintain. For example, the amount of work necessary to obtain the legal permits needed to conduct research abroad is often inhibitive. In addition, each country has its own unique issues that must be sorted out locally. Researchers, therefore, may require incentives to overcome these barriers if they are to develop successful collaborations.

Challenges

- Building international, networked-based collaborations requires "activation energy"
- Institutional cultures can act as disincentives to collaborations and the use of new technologies.
- Barriers toward collaborating and adopting CI due to lack of incentives

Limited Funding

Workshop participants identified the lack of adequate funding in Central American countries as one of the most important inhibitors to the use of international networks. Such limited funding may be due to the competing costs of basic research operations, a lack of skill on the part of researchers and institutions to acquire funding, and/or the fact that many Central American institutions do not have appropriate mechanisms in place to allow the funds that are obtained to be sustained beyond budget and electoral cycles.

From DNA sequencers and subzero freezers to technicians' salaries and travel expenses, basic biodiversity research operations can consume large amounts of money. In many Central American countries, the flow of cash is often too slow to meet these basic needs, let alone pay for the infrastructure and training needed to collaborate on an international scale using high-speed networks. With such limited funding, Central American research institutions may opt to support local rather than international collaborations. Yet, international collaborations will undoubtedly define the future of biodiversity research. The issue of limited funding, therefore, must soon be addressed if Central American scientists are to participate in the global push for biodiversity conservation.

While funding options may be limited, it could be a lack of skill on the part of scientists and institutions in acquiring funding that prevents them from accessing funds. Workshop participants reported that individual scientists often do not engage in social networking. In this regard, science is not unlike business. Funding agencies and private donors may be willing to support projects, but if they are never approached by scientists, they may be unaware of the need. Institutions also were said to lack such entrepreneurial skills.

Sustaining funding once it has been obtained was also raised by workshop participants as an issue. It was noted that, from a particular grant, little or no money is allocated to building sustainability. In addition, the administrative skills for dealing with money are absent, and therefore, funds are often mismanaged. Finally, an absence of policies that enable programs to be institutionalized was cited as a problem that inhibits the acquisition of research funds.

Challenges

- Lack of fundraising skills
- Sustaining funding
- Lack of policies to institutionalize programs

NEXT STEPS

Enabling Solutions with CI

International collaboration is the key to overcoming problems associated with biodiversity data collection and analysis; it is also an important way to improve the education of young scientists. But, to connect researchers and students from around the world so they can jointly answer some of the world's greatest biological questions, major investments in CI must be made. Fortunately, much of the needed infrastructure has already been established. Yet, one of the most pressing issues facing network-enhanced research is underutilization of the networks. While it is expected that integration of CI into everyday research activities will take time, biologists, in particular, have been slow to adopt the new technologies. Gaps in the network, under-appreciation of network capabilities, and data management issues all were identified by workshop participants as major challenges to CI adoption.

Network gaps are the result of a failure on the part of universities and other research organizations to establish connections to network hubs. This may be because institutions lack the resources with which to connect or it may be that authorities do not feel compelled to connect if they have not received requests by faculty and staff to do so. The most important way to resolve connection issues is to increase researchers' awareness and appreciation of CI. Once that is achieved, the researchers will be in a position to apply pressure to their institutions to provide the connections. Increased funding may also help these organizations connect to the network.

Even with adequate funding and awareness, many biologists still fail to see the utility of high-speed networks. These researchers might find it useful to learn about the successes of other groups and participate in workshops and meetings in which they have the opportunity to meet people with whom they might collaborate. Once they are fully educated about the benefits of CI, many of these biologists will likely seek out new opportunities to incorporate it into their work.

In order for data-sharing of this magnitude to work, storage and collection of data and metadata assignments must be standardized. Workshop participants suggested that a workshop should be held in which training on data quality control is conducted. They also suggested that an evaluation mechanism should be established that rates the quality of datasets, such as in eBay or Amazon (Shmueli et al., 2006). Tools might also be developed that search for inconsistencies in datasets. In addition, data-sharing might be improved if community-based middleware initiatives are stimulated. For example, a middleware layer would assist collaborative research involving environmental sensor networks. In this case, middleware would include software for common approaches to managing real-time data streams for caching sensor data and for Web services from analysis engines. Workshop participants suggested the need for pulling together (or creating) a suite of interoperable open-source CI tools and combining them with deployment test beds and training courses.

Beyond networks, other types of CI tools may be useful to Central American researchers that have already been developed using NSF funding. These include federated database technologies, which integrate multiple separate databases into one database (Ledlie, 2005); workflows, which combine separate application components into one large-scale analysis (Ludaescher, 2005; Deelman, 2006); semantic mediation, which integrates data from different sources (Villa, 2004; Nambiar et al., 2003; Michener et al., 2007), and sensor network technologies. The Global Lake Ecological Observatory Network (GLEON) is one example of an organization that employs sensor network technologies. An international group of limnologists, engineers, and information technology experts who study lake ecology, GLEON currently has no research sites in Central America, but would likely benefit greatly from such collaboration. An indirect byproduct of the Panama biodiversity workshop was another joint workshop held in June 2007 in Bahia Blanca, Argentina entitled Pan-American Sensors for Environmental Observatories (PASEO) on the use of sensors in the environment that had been in the early planning stages prior to Panama. This workshop led to a supplement request that was funded in August to support a team of GLEON researchers and their students to work closely with Argentinean and other Southern Cone countries beginning in 2008. Success with this pilot project may lead to a further expansion involving more of the Central and Latin American countries. Finally, the Collaborative Large-scale Engineering Analysis Network for Environmental Research (CLEANER), which has as its goal to enable researchers to study landscapes that are stressed by human activities, is another CI-enabled research effort that could be expanded to include Central American researchers.

While GLEON and CLEANER are two NSF-funded efforts that could benefit from expanding collaborations to include Central American researchers, another NSF-funded project is already doing so. One of the major direct outcomes of the Panama workshop, the Small Grant for Exploratory Research (SGER) was awarded to create a working group of environmental scientists and technologists from the U.S. and Central America to evaluate the research and education requirements for end-user software for the planned sensor network at the La Selva Biological Field Station of the Organization for Tropical Studies (OTS). Environmental sensor networks are becoming increasingly relevant to ecological and biodiversity research and with the international deployment of high bandwidth research and education network backbones, they are likely to become a significant part of multi-national collaboration efforts. This SGER creates an excellent opportunity to collaborate with community-based middleware and interoperable open-source CI tools initiatives to extend their capabilities for the sensor network at La Selva. The primary outcome of the work will be a requirements document for research, education, and outreach software applications of the sensor network.

In addition to NSF-funded projects, collaboration with members of industry through companies' university relations divisions can be highly beneficial to international research. Hewlett Packard, IBM, and Microsoft are examples of companies that could provide physical infrastructure, data management tools and services, and network capabilities at discounted rates. Likewise, companies, such as Onset, Decagon, and Campbell Scientific might be willing to provide low-cost field equipment and multi-sensor technologies to support international collaborative research.

Solutions

- Increase researchers' awareness of CI
- Increase data quality control
- Expand programs in the United States and elsewhere to include Central America
- Bridge physical network gaps

Increasing Awareness

The rate of adoption and effective use of CI by biologists has thus far been rather slow due, in part, to the fact that many researchers are unaware of the benefits that CI can bring to their research and abilities to collaborate. To increase the rate of adoption and use of CI, a comparative study to understand the readiness of Central American researchers to use technology innovations that are already in use in the U.S. should first be conducted, followed by education and outreach programs that spread the word about CI opportunities. But since CI is much more specialized and demanding than traditional information technology products, such as email and the Internet, a sustained effort in training and support will be required.

As mentioned in the previous section, a workshop could be held in which biologists are introduced to the possibilities that CI can bring to their research. Such a workshop could include case studies that highlight the successes of CI-enabled collaborative projects, such as the biological sensor network at La Selva. The workshop could also serve as a meeting place for scientists to forge new relationships with people who are studying similar aspects of biodiversity.

Students, especially, should be targeted through outreach programs. CI practitioners can be paired with students and their faculty advisors to foster the development of CI-enabled collaborations. Global CyberBridges[1], a program at Florida International University, is an excellent example of how students from a variety of scientific disciplines can be trained in the use of CI. The program trains and mentors graduate students, both in the United States and abroad, to acquire CI skills from CI staff and share them with their faculty advisors. For example, one of the students is learning how to use CI to design a method that allows scientists to determine a protein's function by rapidly searching databases for similar proteins with known functions. So far, 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 São Paulo.

Another effort that aims to disseminate scientific and engineering knowledge and stimulate collaborative learning and cooperation among the research communities of the Americas is the Pan-American Advanced Studies Institute (PASI), funded by an NSF program of the same name. One of the immediate outcomes of the Panama workshop, will be a two-week PASI in which students and young faculty from North, South, and

[1] http://www.cyberbridges.net/

Central America can learn about CI tools and their application to biodiversity and ecological research by attending a lecture series and getting hands-on laboratory experience. This PASI will provide a first step toward introducing the capabilities for enhancing research that CI can bring to researchers in Central America. The PASI will be held in June 2008, in Costa Rica, on the theme "Cyberinfrastructure for International, Collaborative Biodiversity and Ecological Informatics."

Solutions

- Hold workshops for scientists to meet and initiate collaborative research
- Increase training opportunities for students and young faculty
- Comparative study to assess readiness
- Targeted research, education and outreach programs

Increasing Trust

Establishing trust among participants is vital to the success of any collaboration. But as researchers in Central America are forced to compete for small pools of money, they are less likely to share data for fear of losing their competitive edge. In recent years, especially in the United States, collaboration has become the *sine qua non* of some funding agencies that wish to support research on large-scale, multidisciplinary questions. Workshop participants identified a number of ways in which to enhance trust among researchers, including increasing opportunities for scientists to meet and connect, and teaching them to collaborate at an early stage in their career.

Building trust is nearly impossible without face-to-face encounters. While telephone and e-mail conversations are useful, the most meaningful connections are usually forged when people meet in person. Such one-on-one opportunities can be increased by holding meetings and workshops in which scientists from different countries can come together. The Pacific Rim Application and Grid Middleware Assembly (PRAGMA[2]), which aims to improve the use of grid technologies by communities of scientific collaborators, is one organization that makes effective use of annual workshops. The group's mandate is to "focus on how practically to create, support, and sustain international science and technology collaborations" by creating "teams of application and middleware researchers focused on a specific research challenge that requires integration of application and grid technologies." This concept of holding workshops and meetings in which teams of researchers can be formed is essential to building trust. The NSF should play a role in providing funding for such meetings.

Such productive communities of collaborators don't appear overnight. Time and patience are key ingredients to building such communities. Workshop participants suggested that collaborative communities should start out small and add new members slowly in order to establish an atmosphere of collaboration that is comfortable for everyone. This slow pace of community-building is expected to help individuals build trusting relationships.

[2] http://www.pragma-grid.net/

A large trust relationship in which the members share data can then subsist by implementing a "data commons," in which each scientist has the right to access data only if they themselves contribute data.

Participants also suggested that collaboration is a skill that can be taught to scientists at an early stage in their careers. If students and early-career scientists are able to establish healthy collaborative relationships at the beginning of their careers, the issue of trust will eventually fade. The Pacific Rim Undergraduate Experiences (PRIME) program at the University of California, San Diego is a good example of how students can be trained to work together. The program provides undergraduates with an opportunity to contribute to the development of global CI and emphasizes collaboration. The NSF-sponsored East Asia and Pacific Summer Institutes for U.S. Graduate Students (EAPSI) is another example of how students can be trained to collaborate on research endeavors. The goals of EAPSI are to "introduce students to East Asia and Pacific science and engineering in the context of a research laboratory, and initiate personal relationships that will better enable them to collaborate with foreign counterparts in the future." A counterpart to EAPSI does not exist for Latin America. The NSF and other funding agencies at the workshop should consider creating an EAPSI-like program for Latin America. The NSF Office of International Science and Engineering (OISE) offers funding programs to support U.S. undergraduate or graduate students, and dissertation research for one doctoral student to conduct research abroad in collaboration with foreign investigators.

Meetings, workshops, and training opportunities for students all require investments of money. A moderate level of funding over several years will be necessary to build human capacity and relationships for collaboration. It would make sense to identify existing, incipient international research networks in which some trust and vision is already shared and there is already some sense of common cause and community. Some possible communities might include:

> 1) Biodiversity collections institutions and their researchers who already share formalisms for research practice and who have complementary data collections and computational requirements. Shared CI exists in this community and it would be well positioned to advance to a higher level of network interaction and collaboration.
>
> 2) A second community could be anchored by the existing U.S. Long-Term Ecological Research Network (LTER), a counter-part in Mexico (Red Mex-LTER[3]), and other countries associated with the International LTER[4] (ILTER). The ILTER aims to "foster and promote collaboration and coordination among ecological researchers and research networks at local, regional and global scales; improve comparability of long-term ecological data from sites around the world, and facilitate exchange and preservation of this data; deliver scientific information to scientists, policymakers, and the public and develop best ecosystem management practices to meet the needs of decision-makers at multiple levels;

---

3 http://www.mexlter.org.mx/

4 http://www.ilternet.edu/

and facilitate education of the next generation of long-term scientists." The LTER community through visionary funding by NSF has evolved a robust, networked research informatics infrastructure, which is well positioned to bridge into Central and South American collaborations.

3) A third area might be characterized as the environmental sensor network community. In the U.S. this involves ecologists, in particular environmental physiologists, who are piecing together sites and organizational CI infrastructure to instrument natural communities and organisms with sensors to detect and quantify various macro- and micro-environmental dimensions. Networks of sensors are, of course, inherently CI-based and with common research objectives and common technological deployment (hardware and software) challenges, sensor network researchers would be an ideal candidate for internationalization. Effective remote access to streaming data sensors and whole sensor networks would open doors for international collaboration.

Solutions

- Bridge gaps in the social collaboration network
- Under-appreciation of network capabilities
- Data management issues

Increasing Incentives

International network-based collaborations require much more 'activation energy' than working with researchers down the hall, across campus, or across the country. The initial work that must be done may deter some researchers from pursuing collaborative relationships on an international scale. Beyond requiring scientists to make data public as a condition of funding, other incentives may be required to motivate these scientists to make the leap. Such incentives may include social and professional mechanisms that showcase and reward successful collaborations, case studies that prove success is possible, and increased funding. For example, Panama's Ministry of Science and Technology is increasing incentives by offering grant solicitations to Panamanians to propose solutions to research challenges that utilize the country's research and education network infrastructure.

Recognition is a strong incentive that motivates many researchers. Scientists traditionally receive recognition for their work by publishing in prominent journals, receiving media attention, and winning awards, fellowships, and tenure. But what recognition do they get for their efforts to collaborate on an international scale? Many researchers recognize that by collaborating with people from around the world and from other disciplines, they will become exposed to new ideas and new ways of solving problems. If such relationships lead to innovative ways of understanding the world, traditional forms of recognition will follow.

There are several examples of studies that have benefited tremendously from collaboration. The LTER network is one of them. The project, which employs more than 1,800 scientists and students who work together to investigate ecological processes, recently increased their scope to include more international work. The success of projects such as the ILTER provides a strong incentive for biodiversity researchers to undertake network-enhanced international collaborations. By watching these organizations, biodiversity researchers throughout the western hemisphere can learn about the things that work and don't work, and can apply these lessons learned to their own collaborative efforts.

Solutions
• Improve recognition for work • Share success stories

Increasing Funding

Limited funding for biodiversity research in Central American countries was identified by workshop participants as one of the most important inhibitors to the use of international networks. Competing costs of basic research operations, a lack of skill on the part of researchers and institutions to acquire funding, and the fact that many Central American institutions do not have programs in place to sustain funding were all cited as problems contributing to the lack of adequate funding. Potential solutions to each of these problems were given and include seeking out funding opportunities from public, private, and academic sectors, leveraging available resources, supplementing existing domestic grants to support international collaboration, and institutionalizing programs that enable funding and the products of funding to be sustained.

Although much funding is available to collaborate internationally, the amount is not enough to achieve such lofty goals as cataloguing all of Earth's biodiversity. For example, a representative from the European Union described a program between Germany and Costa Rica that funds faculty and student internships in Germany. U.S. research and education institutions also have exchange programs with Central American countries. The Smithsonian Tropical Research Institute (STRI) in Panama and the La Selva Biological Station in Costa Rica are two facilities that attract international researchers to conduct studies, often in collaboration with local scientists. Still, more funding must be devoted to these international collaborations.

Improving fundraising skills is one way to acquire additional funds. Social networking is particularly important for individual investigators since funding agencies and private donors might be willing to support projects but may be unaware of the need. It is also important for scientists and institutions to develop skills in obtaining funding from a variety of sources, including public, private and academic sources. Such skills can be acquired by implementing training programs, collaborating with researchers and institutions that have already attained fundraising skills, and by passing on fundraising knowledge to students and early-career scientists.

Many resources needed by Central American scientists who wish to participate in international network-enabled biodiversity research already exist. To avoid reinventing the wheel, researchers must seek out opportunities to leverage these readily available resources. Foundations, non-profits, government agencies, and universities all contain knowledge, fundraising skills, human resources, and infrastructure that could potentially be shared.

As a leader that supports collaborative research via high-speed networks, the United States should play a role in helping Central American researchers and institutions obtain funding and participate in collaborative research projects, especially since the country's own funding agencies, including the NSF, have begun to place higher priority on funding research by international collaborators. Furthermore, future funding initiatives, whether by United States organizations or Central American organizations, should consider funding more cross-disciplinary network and computational biodiversity training events, workshops, and institutes. In recent years, tightened restrictions on using NSF funds for international collaborative activities have diminished opportunities for seeding new collaborations. This policy should be reexamined, in light of the interests of NSF, to promote regional- and global-scale collaboration.

The NSF PASI is a good starting point to improve collaborative efforts between the United States and Latin American collaborators. A PASI may serve as a mechanism for obtaining funding to hold a short course that involves lectures, demonstrations, research seminars, hands-on tutorials, and discussions at the advanced graduate and post-doctoral level. Outcomes of a PASI can be very useful for longer-term programs, such as the NSF Partnerships for International Research and Education (PIRE) and the International Research Experiences for Students (IRES), and the Cyberinfrastructure Training, Education, Advancement, and Mentoring for Our 21st Century Workforce (CI-TEAM) programs.

The PIRE is one NSF program that promotes the establishment of collaborations between U.S. researchers and international institutions. According to the program's Web site, "International partnerships are, and will be, increasingly indispensable in addressing many critical science and engineering problems. As science and engineering discoveries result more and more from international collaboration, U.S. researchers and educators must be able to operate effectively in teams comprised of partners from different nations and cultural backgrounds."

The IRES is another NSF program that promotes international collaboration. According to that program's Web site, "The United States needs to educate a globally-engaged science and engineering workforce capable of performing in an international research environment in order to remain at the forefront of world science and technology. To support this aim…"IRES supports groups of U.S. undergraduate or graduate students conducting research abroad in collaboration with foreign investigators." The NSF CI-TEAM also supports students as well as current scientists. This program focuses more heavily on funding individuals and groups who wish to use CI tools to expand their research. Global CyberBridges, which is discussed above, is one program that has benefited from NSF CI-TEAM funding. Proposals to the NSF can include an international component for evaluation by the Office of International Science and Engineering (OISE). This office will provide additional funding to projects that develop new intellectual collaborations with a foreign partner, in which the foreign partner provides some otherwise unrealized benefit in terms of expertise, facilities, or other resources. Active engagement of U.S. students and junior researchers at the foreign site is also required. The NSF OISE will fund planning visits for U.S. scientists to visit foreign countries for short visits. Such visits can enable U.S. investigators to link their NSF-funded projects to activities in Central America, then receive funds from OISE for a planning visit geared toward developing a proposal for a larger program.

Once funding is acquired, mechanisms should be in place to sustain the flow of money as well as the infrastructure and data that result from funding. This can be achieved by using a percentage of the funds to build sustainability, by using research results to generate more income and secure additional donations, and by focusing on cultivating a variety of funding sources. In addition, the communities and programs that result from funding should be institutionalized to maintain continuity.

Solutions
• Seek out funding opportunities from a variety of sources • Improve fundraising skills • Leverage available resources • Build sustainability

RELEVANT SUCCESSES AND OPPORTUNITIES

A common grid-based computational infrastructure is designed to avoid independent, efforts which cannot be leveraged. Yet, the organizations that make use of this infrastructure, while often maintaining common goals, are inherently different, and it behooves us to learn from these groups' successes and failures. In addition to improving the potential for this project to succeed, keeping abreast of the activities of others will allow us to synchronize efforts. Here we discuss some of the organizations whose goals are most closely aligned with ours.

The Pacific Rim Application and Grid Middleware Assembly (PRAGMA) aims to improve the use of grid technologies by communities of scientific collaborators. The organization's success is due to several factors. First, its participants share a common vision. While not all participants agree on everything all the time, there is a critical mass of individuals that agrees on the organization's goals and activities. Such buy-in is essential to the success of groups formed from international collaborators. Regular face-to-face meetings also contribute to PRAGMA's success. These meetings help to build trust among participants, especially when they are held in different locations around the world and include informal opportunities to network, such as during meals. PRAGMA also employs a staff that is dedicated to the program. These paid employees are better able to carry out the organization's operations than would be the program's participants if they were responsible for such day-to-day tasks. PRAGMA also prides itself on its ability to follow through with its commitments. By doing what they say they will do, PRAGMA builds trust among its own participants, funding sources, and other external partners. One of PRAGMA's final keys to success is their operating principles and procedures document, which defines who they are and how they go about their business. The document is comprehensive, yet flexible, and helps guide the organization as it conducts its business.

Another successful organization is the Taxonomic Databases Working Group (TDWG), which aims to improve international collaboration among users of biodiversity databases. The group's goal is to develop and foster the implementation of data description and exchange protocols and related technical standards, and to serve as an international forum for discussion by hosting annual meetings and producing publications. TDWG has persisted and prospered because it offers the opportunity for open exchange of technical solutions and approaches to biodiversity data access and integration challenges. TDWG is focused on a few dimensions of CI for biodiversity research, in particular handling species data such as taxonomies and classifications, specimen and museum data from voucher specimens, documenting the occurrence of organisms, and 'character' data, which is essentially phenotype information typically compiled at the level of species. These three classes of data comprise the basis for much biodiversity research, and TDWG has filled a need for a social and technical venue to address their standardization and use.

In addition, membership and consensus-building organizations like the Global Biodiversity Information Facility (GBIF) sponsor activities to build common data management, linkage, and computational architectures for the biodiversity research community. With an ample budget to hold frequent training workshops and technical discussion meetings focused on particular sub-problems that the entire global community of biodiversity data providers face, GBIF has been able to make considerable headway in building a shared vision and technology bases for biodiversity data access and integration.

The Inter-American Biodiversity Information Network (IABIN), which supports a decentralized, Internet-based, western hemisphere network that provides access to biodiversity information resources, is successful because it has established a global network of 34 countries with official representatives who actively provide access to local biodiversity information (Dr. Ivan Valdespino, IABIN Secretariat, personal communication). Also participating are major non-governmental organizations representing national and international communities, including The Nature Conservancy (TNC), World Wildlife Fund (WWF), O Boticario Foundation of Brazil, and others. IABIN has been successful in implementing five different thematic networks in the following areas: Species and Specimens, Protective Areas, Invasive Species, Pollinators, and Ecosystems. Three of these five are already serving data and are accessible by people using web portal technologies. A significant achievement for IABIN has been the effective use of decision-support tools for collaboration with its partners, such as the national Biological Information Infrastructure at the United States Geological Survey (USGS), the USGS Center for Earth Resources Observation and Science (EROS) and the Comisión Centro Americana de Ambiente y Desarrollo (CCAD), with funding from the World Bank. Currently, IABIN is funding data digitizing grants to promote data sharing among countries in the Americas and, for this, has awarded 27 grants to an equal number of institutions in the region for a total of $275K. Next year, the remainder of the $1.3M will be available for grants for data digitizing. Countries are already using this data for their conservation planning. For example, CCAD countries in Central America have incorporated tools for regional biodiversity monitoring programs, Uruguay has used the data to develop a national invasive species strategy, and Chile is using the data portal and tools developed by the national biodiversity information system.

The Global Lake Ecological Observatory Network (GLEON) is an international group of limnologists, engineers, and information technology experts who are collaborating to build a network of lake ecology observatories with a goal of understanding important processes such as carbon cycling and the effects of climate change on lake dynamics. The group plans to install instrumented platforms on lakes around the world to collect data and upload it to web-accessible databases. Three management approaches have contributed to GLEON's momentum (Dr. Tim Kratz, personal communication). First, GLEON quickly became an active ongoing initiative because of its professional structure, with low barriers to participation to all biology and CI researchers interested in systems involving high-frequency aquatic monitoring data. Second, GLEON found that non-federal funding was essential to support the level of international collaboration that they needed. NSF needs to re-evaluate the negative research impacts non-support of international collaborators for any activities. Funding from the Moore Foundation was critical to fill this gap and contribute to the project's ongoing momentum and success. The third factor contributing to GLEON's success is its focus on incremental collaborative projects which are attainable with their existing financial resources.

The Protein Data Bank (PDB) is a worldwide depository for 3-dimensional protein data. The bank uses a web portal to provide information to users. Here, researchers can search the database, acquire summaries of data, and post questions and answers on a discussion forum. The PDB is run by an international advisory committee comprising experts in a variety of fields. This group is successful because they have significant funding, a large number of developers, a common data type (protein structure), and a well-focused, partnership with NSF/National Institutes of Health (NIH) and the National Institute of Standards and Technology (NIST). In addition, PDB has numerous international connections and an enlightened leadership.

The LTER network, which employs more than 1,800 scientists and students who work together to investigate ecological processes, recently increased their scope to include more international work. LTER has achieved great success due to long-term support around a common vision of field station-based science and informatics; recognition that problems in ecology are not local, but usually occur at many sites and scale to regional and continental scale in many cases; a philosophy that values long-term observation and experimentation; and a well-organized community of information managers within the

LTER that advances a common information technology agenda in support of ecological research and cross-site comparisons and collaboration. NSF played a large role in helping LTER scientists see the value of collaboration and shared CI through strategic funding of collaborative informatics activities in the 1980s and 1990s. Due to their experience with managing large amounts of data and making them available to the public, the LTER and ILTER would be excellent partners with which to collaborate on a workshop that focuses on data quality control. Likewise, the National Oceanic and Atmospheric Association (NOAA), National Aeronautics and Space Administration (NASA), and USGS would be good partners given their expertise with biological databases and metadata.

SYNTHESIS

Biological diversity transcends national boundaries and the study of biodiversity cuts across myriad research disciplines. Biodiversity informatics provides communication and computational infrastructure for this global, interdisciplinary research enterprise. This report describes the impetus for and results from a workshop held in Panama in January 2006, in which opportunities for collaborative research using high-speed international networks were discussed.

Findings from the workshop revealed four areas where there are barriers impeding the effective use of communications networks and CI to enhance U.S.-Central America scientific collaborations. Those four areas identified were lack of awareness, limited trust, lack of incentives and lack of funding. Specific challenges in each of these four areas were described. Recommendations to reduce or eliminate these barriers were described for each of the corresponding areas.

In general, to increase the use and effectiveness of CI for biodiversity research, we recommend holding additional workshops that will be more narrowly focused and include a greater number of relevant participants, such as scientists who conduct biodiversity research, as a next step. These workshops would allow participants to resolve some of the more specific barriers to the widespread use of CI among international collaborators.

WORKSHOP PARTICIPANTS

Gabrielle Allen
Associate Professor, Computer Science
Louisiana State University

Heidi Alvarez
Director, Center for Internet Augmented Research and Assessment (CIARA)
Florida International University

Peter Arzberger
Chairman
Pacific Rim Application and Grid Middleware Assembly (PRAGMA)

Jim Beach
Assistant Director for Informatics, Biodiversity Research Center
University of Kansas

Rodrigo Bernal
Professor
Instituto de Ciencias Naturales of Universidad Nacional de Colombia

Jesús M. Castagnetto
Professor, Department of Chemisty
Universidad Peruana Cayetano Heredia

Marta Cehelsky
Senior Adviser for Science and Technology, Department of Sustainable Development
InterAmerican Development Bank

Edwin Castellanos
Director of the Center for Environmental Studies
Universidad del Valle de Guatemala

José Castro
Director, Computing Research Center
Costa Rican Institute of Technology (ITCR)

Julio Escobar
National Secretary
Science, Technology and Innovation (SENACYT)

Tony Fountain
Director of CLEOS, San Diego Supercomputer Center
University of California

César Garita
Director, School of Computer Science
Costa Rican Institute of Technology

Amado Gonzalez
Director, Access Grid activities
Florida International University

Eric Graham
Staff Biologist, Center for Embedded Networked Sensing,
University of California

Saul Hahn
Head of the Division of Science and Technology
Organization of American States (OAS)

Dewayne Hendricks
CEO
Dandin Group, Inc.

Don Henshaw
Chairman
LTER Network Information System Advisory Committee

Leonard P. Hirsch
Senior Policy Advisor
Smithsonian Institution

Mandë Holford
Postdoctoral Fellow
University of Utah

Steven G. Huter
Research Associate
University of Oregon

Raúl Hazas Izquierdo
Head of the Telematics Directorate
CICESE

Julio Ibarra
Executive Director of the Center for Internet Augmented Research and Assessment (CIARA)
Florida International University

Steve Kelling
Director of Information Sciences
Cornell Lab

Suzanne Lao
Database Manager, Center for Tropical Forest Science
Smithsonian Tropical Research Institute

Lee Liming
Manager, Distributed Systems Laboratory
University of Chicago

Genevieve Lucet
Head of the Computer Services for Research program
Universidad Nacional Autonoma de Mexico

Gabriel Macaya
Professor and Researcher, Center for Cell and Molecular Biology (CIBCM).
University of Costa Rica

Erick Mata
Associate Director of INBio
Costa Rica Institute of Technology

Steven Paton
Data Manager and Physical Monitoring Project manager, Environmental Sciences Program
Smithsonian Tropical Research Institute

Juan A. Sánchez
Assistant Professor, Marine Molecular Biology laboratory (BIOMMAR)
University of the Andes

Alejandro F. Flamenco Sandoval
Doctoral Student, Centro de Investigaciones en Ecosistemas
Universidad Nacional Autonoma de Mexico

Mark Schildhauer
Director of Computing
National Center for Ecological Analysis and Synthesis (NCEAS)

Ana Sittenfeld
Director, Office of International Affairs and External Cooperation (OAICE)
University of Costa Rica

Jorge Soberon

Senior Scientist, Biodiversity Research Center
University of Kansas

Harold Stolberg
National Science Foundation

Sylvia Spengler
National Science Foundation

Florencio I. Utreras
Executive Director
Latin American Cooperation of Research Networks

Iván A. Valdespino Q.
Secretariat Director
Inter-American Biodiversity Information Network (IABIN)

ACKNOWLEDGMENTS

The organizers would like to thank all of the participants of the workshop who contributed to the findings of this report. Their presentations, panel summaries, and contributions from the breakout sessions provided the necessary information to inform this report.

We would like to thank Dr. Julio Escobar, the National Secretary of Science, Technology and Innovation (SENACYT) of Panama, for hosting the workshop in Panama City, and Ms. Gisele Didier of SENACYT for providing coordination and organization in Panama.

We would like to thank our partners, the U.S. National Science Foundation (NSF), SENACYT of Panama, CONACYT of Mexico and CR-USA of Costa Rica for providing support for their nation's scientists, practitioners, and policy-makers to participate. Special thanks go to the Smithsonian Tropical Research Institute for making it possible for the workshop participants to experience the Barro Colorado Island biological research facility and to the Global Biodiversity Information Facility (GBIF) for their sponsorship.

Financial support for the workshop was provided by the National Science Foundation (NSF OISE award #0549443) and by the Global Biodiversity Information Facility (GBIF) in support of this workshop.

The authors would like to acknowledge the exceptional contribution and patience of the Editor of this report, Ms. Sara LaJeunesse. We offer Sara our heartfelt appreciation and gratitude.

REFERENCES

Altintas, I., A. Birnbaum, K. Baldridge, W. Sudholt, M. Miller, C. Amoreira, Y. Potier, and B. Ludaescher. 2005. A Framework for the Design and Reuse of Grid Workflows, Intl. Workshop on Scientific Applications on Grid Computing (SAG'04), LNCS 3458, Springer, 2005.

Bisby, F. A. 2000. The quiet revolution: biodiversity informatics and the Internet. *Science* 289: 2309–2312.

Bisby, F.A., J. Shimura, M. Ruggiero, J. Edwards, and C. Haeuser. 2002. Taxonomy, at the click of a mouse. *Nature* 418: 367.

Causey, D., D. H. Janzen, A. T. Peterson, D. Vieglais, L. Krishtalka, J. H. Beach, and E.O. Wiley. 2004. Museum Collections and Taxonomy. *Science* 305: 1106-1107.

Cummings, J. N. and S. Kiesler. 2005. Collaborative Research Across Disiciplinary and Organizational Boundaries. *Social Studies of Science* 35(5): 703-722.

Deelman, E.,Y. Gil, et al. 2006. Report of the 2006 NSF Workshop on the Challenges of Scientific Workflows. 11 pp. Available at: http://grids.ucs.indiana.edu/ptliupages/publications/IEEEComputer-gil.pdf

Edwards, J. L., M. A. Lane, and E. S. Nielsen. 2000. Interoperability of biodiversity databases: biodiversity information on every desktop. *Science* 289: 2312–2314.

Foster, I. 2005. Service-Oriented Science. *Science* 308: 814-817.

Godfray, C. 2002. Challenges for taxonomy. *Nature* 417: 17–19.

Hebert, P., A. S. Cywinska, L. Ball, and J. R. de Waard, J. R. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond.* B 270: 313–321.

Kasper-Fuehrer, E. C. and N. M. Ashkanasy. 2003-04. The Interorganizational Virtual Organization: Defining a Weberian Ideal. *International Studies of Management and Organization* 33(4): 34-64

Kouzes, R. T., J. D. Myers, and W. A. Wulf. 1996. Collaboratories: Doing Science on the Internet. *IEEE Computer* 29(8): 40-46.

Krishtalka, L., A. T. Peterson, D. A. Vieglais, J. H. Beach, and E. O. Wiley. 2002. The Green Internet: a tool for conservation science. In *Conservation in the Internet age: strategic threats and opportunities* (ed. J. N. Levitt), pp. 143–164. Washington, DC: Island Press.

Ledlie, J., C. Ng, D. A. Holland, et al. 2005. Provenance-Aware Sensor Data Storage. 21st International Conference on Data Engineering Workshops (ICDEW'05), IEEE. http://doi.ieeecomputersociety.org/10.1109/ICDE.2005.270

Lin, K., B. Ludaescher, B. Broadaric, et al. 2003. Semantic mediation services in geologic data integration: A case study from the GEON grid. Geological Society of America Abstracts with Programs, Vol. 35, No. 6, September 2003, p. 365.

Ludaescher, B., I. Altintas, I., C. Berkley, C., et al. 2005. Scientific Workflow Management and the Kepler System. *Concurrency and Computation: Practice and Experience (Special Issue: Workflow in Grid Systems)* 18(10):1039-1065.

Michener, W. K., J. H. Beach, M. B. Jones, B. Ludaescher, D. D. Pennington, R. S. Pereira, A. Rajasekar, and M. Schildhauer. 2007. A knowledge environment for the biodiversity and ecological sciences. *J. Intell. Inform. Systems* 29(1): 111-126.

Nambiar, U., B. Ludaescher, K. Lin, C. Baru. 2006. The GEON portal: accelerating knowledge discovery in the geosciences. Proceedings of the 8th annual ACM international workshop on Web information and data management, pp. 83-90.

Olson, G. M. and J. S. Olson. 2001. Distance Matters. *Human Computer Interaction* 15: 139-179.

Robertson, D. S. 2003. Phase Change: the Computer Revolution in Science and Mathematics. Oxford, Oxford University Press.

Shmueli, G., W. Jank, A. Aris, C. Plaisant, and B. Shneiderman. 2006. Exploring auction databases through interactive visualization, *Decision Support Systems* 42(3): 1521-1538.

Soberón, J. and A. T. Peterson. 2004. Biodiversity informatics: Managing and applying primary biodiversity data. *Philosophical Transactions of the Royal Society of London B*, 359: 689-698.

Villa, F. 2007. A semantic framework and software design to enable the transparent integration, reorganization and discovery of natural systems knowledge. *Journal of Intelligent Information Systems* 29(1): 79-96.

Watson-Manheim, M. B., K. M. Chudoba, and K. Crowston. 2002. Discontinuities and Continuities: A New Way to Understand Virtual Work. *Information Technology and People* 15(3): 191-209.

Wheeler, Q.D., P. H. Raven, and E. O. Wilson. 2004. Taxonomy: Impediment or expedient? *Science* 303: 285.

Wilson, E.O. 2000. On the future of conservation biology. *Conservation Biology* 14: 1-3.

Wilson, E.O. 2003. The encyclopedia of life. *Trends in Ecology and Evolution* 18: 77-80.