

SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences

Fran Berman, San Diego Supercomputer Center and UC San Diego

Workshop Concept

Cyberinfrastructure is the coordinated aggregate of software, hardware and other technologies, as well as human expertise, required to support current and future discoveries in science and engineering. The challenge of Cyberinfrastructure is to integrate relevant and often disparate resources to provide a useful, usable, and enabling framework for research and discovery characterized by broad access and “end-to-end” coordination.

Today, most Cyberinfrastructure efforts are focused on the development and integration of Cyberinfrastructure technologies and resources. Fewer efforts have focused on the immense repercussions of the social dynamics and organizational, policy, management and administration decisions inherent in developing and deploying Cyberinfrastructure. Such choices, and the **social, cultural, and behavioral impacts of how we develop, manage, and evolve Cyberinfrastructure** will be critical to its success.

In addition to developing a useful and usable Cyberinfrastructure, attention must be focused on the way in which **successful Cyberinfrastructure will change our social and cultural environment**, including the way it will change the way we do social science itself. At scale, Cyberinfrastructure will support an increasing number and variety of users and usage scenarios, and will accelerate the evolution of a cyber-culture. Attention must be paid *now* to ensure Cyberinfrastructure is of most benefit as it evolves.

In order for Cyberinfrastructure to achieve its potential as an enabling framework for modern science and engineering, it must be thoughtfully designed, expertly developed, deployed at a stable, professional level, and its impact must be understood. A strategic and forward-looking approach to the development of a body of research, experimentation and infrastructure *starting now* that addresses the socio-cultural, economic, and policy challenges of both building and using Cyberinfrastructure will be critical to ensure best success. **Collaborations between social scientists and computer scientists, technologists, and engineers will be critical to meet the challenges of building and using successful Cyberinfrastructure.** Such individuals have a wealth of experience and context to offer, but have yet to realize the potential of partnering together in this critical area.

The first step to developing a strategic approach and priority areas for collaboration between social scientists, computer scientists, engineers, and technologists is to identify the key opportunities and challenges. The SBE/CISE Workshop on Cyberinfrastructure for the Social Sciences will seek to identify areas in which both the SBE and CISE communities can partner together to begin to address the most fundamental and strategic problems. The Workshop has been designed to engage an

outstanding group of participants from SBE, CISE and elsewhere in substantive discussions that will provide a targeted path forward, and key recommendations, useful to NSF and its constituencies as they move forward to build, deploy, and use Cyberinfrastructure.

Workshop Goals

Our goals are to provide substantive, useful and usable feedback to NSF on programs, and activities for which the SBE and CISE communities can partner together to build, deploy, and use Cyberinfrastructure. We seek to accomplish two goals at the workshop:

- 1) *The Workshop Report should lay out a Cyberinfrastructure research, experimentation, and infrastructure path forward for the SBE and CISE community and provide a framework for projects and efforts in this area.*
- 2) *The Workshop should provide a venue for community building within the SBE and CISE communities, and in particular a venue for a multi-disciplinary synergistic community which leverages the perspectives and research of both SBE and CISE constituencies.*

Workshop Framework

The workshop will combine invited talks, breakouts, and report-outs. Workshop participants will be assigned to one of 6 sessions, and will be asked to work together to produce session recommendations that meet the Workshop goals. It is our hope that the discussion and interactions facilitated at the workshop will produce not only useful recommendations for programs, and efforts which NSF can sponsor, but the beginnings of new interdisciplinary collaborations and teams that can work together to build an integrated SBE/CISE community focused on Cyberinfrastructure challenges.

Output of the Workshop

The “output” of the workshop will be a Final Report which provides an overview of the opportunities, challenges, and recommendations from each session and recommends a set of research, experimentation and infrastructure challenges which can serve to address these problems. The report should provide a concrete description of the Cyberinfrastructure socio-cultural, economic and policy landscape, and a set of key areas to pursue which will foster critical efforts and activities which address these challenges. The report will be made available publicly via the web and will also be printed in hardcopy for use by NSF and key individuals.

Workshop Sessions

Workshop participants have been assigned to one of 6 sessions, and each session is led by a “SBE”-oriented researcher and a “CISE”-oriented researcher. In preparation for the workshop, participants were asked to respond to a set of questions targeted to their session in a “White Paper”. The White Papers for each session will be made available via the web and will be available at the Workshop.

On the pages are the introductory text and questions for each session’s White Papers along with contact information for the Session Co-leads.

ORGANIZATION OF CI AND CI-ENABLED ORGANIZATIONS

Session

Session Co-leads: *Fran Berman*, San Diego Supercomputer Center and UC San Diego (berman@sdsc.edu), and *Jane Fountain*, National Center for Digital Government, Harvard (jane_fountain@harvard.edu)

Cyberinfrastructure (CI) is the coordinated aggregation of technologies and other resources to enable science, engineering, and societal efforts. The development of Cyberinfrastructure, and its application, is often multi-disciplinary, team-oriented, domain-driven and collaborative. In this session, we focus on two aspects of Cyberinfrastructure: The development of successful organizational frameworks for Cyberinfrastructure itself, and the impact of successful Cyberinfrastructure on organizations.

Themes and questions to focus our session are outlined here. We ask you to prepare a one page statement to address questions within these themes and to use an additional page to raise other pertinent issues. These White Papers will form the basis for our discussion and help us craft a set of recommendations to NSF to contribute to a research program in the area of Cyberinfrastructure for the Social Sciences.

Key Themes and Questions:

Theme 1 -- Models and Frameworks:

- How should we develop organizational, management, incentive, and other key structures for CI to enable its success, support and sustainability?
- How are these measured? By whom and toward what ends? How should non-commensurable CI goals be balanced/weighted?
- What are appropriate decision making and conflict resolution models for CI?

Theme 2 -- CI Tools for Managing Organizations:

- What CI tools and technologies are likely to be most useful in effectively managing organizations?
- What are appropriate levels of usability, reliability, and predictability for CI?

Theme 3 -- Processes of Transformation:

Successful CI will enable new modes for organizational arrangements including but not limited to communication, integration and efficiency.

- How will successful CI change organizations? Are there sustainable successes or examples of transformation in CI-enabled organizations from which to draw generalizable lessons?

- What normative questions are important to research with respect to CI-enabled organizations? For example: How much privacy should organizational participants have? Should there be limits to transparency?

CYBERINFRASTRUCTURE TOOLS FOR THE SOCIAL SCIENCES ***Session***

Session Co-Leads: *Henry E. Brady*, UC Berkeley, (hbrady@csm.berkeley.edu), and *TBD*

Social science cyberinfrastructure is a complex world involving virtual organizations and models, multi-disciplinary research and collaboration, mediated interactions, computational science, and data-rich environments, but it is largely an abstract world that only becomes real through cyberinfrastructure tools. Cyberinfrastructure tools create the new capabilities and worlds that promise to transform social science. These tools are the interface between the user and pervasive IT technologies and applications; for the user, these tools are the face of cyberinfrastructure.

Social science needs tools that facilitate its basic tasks of comparing, measuring, and searching for patterns. Social scientists compare political systems, elections, economic statistics, budgets, social structure, religions, and many other phenomena across countries; they compare political views, economic behavior, marriage, families, and criminal behavior across individuals. Comparison is one of the most powerful things social scientists do. In making these comparisons, they must often work across different languages, cultures, and recording systems. National statistical systems, for example, are often organized differently across societies. Information is recorded in different languages immersed in different cultures. Social scientists measure things by counting, asking people questions, and coding data. Despite the seeming simplicity of these operations, however, there are many different ways that things are counted, that people answer questions, and that data can be coded. Finally, social scientists search for conceptual and causal patterns in data.

By linking data over time, across place, and across many different kinds of datasets, social scientists increase the possibilities for comparison and measurement. They also increase the possibilities for finding interesting patterns. Linkage is one of the most powerful things social scientists can do to facilitate comparison, measurement and pattern-seeking. But they face non-uniformities, changes, and incompatibilities in data over time, across place, and across datasets. Cyberinfrastructure tools can facilitate linking and analysis through comparison, measurement, and pattern-finding. In order to do this, social scientists need the help of computer scientists to develop better techniques for archiving, for creating meta-data and indexing, for the retrieval, use, and analysis of data, and for collaboration:

- **Timely and Broad Ranging Archiving Efforts** -- Because social science data are typically generated by non-social scientists and because they are often fugitive and ephemeral, social scientists need archives that are constantly on the alert for these data and which archive them in a timely manner. These archives must work with those that generate data (such as bureaucracies, organizations, and individuals) to document the data and to insure access for researchers.

- **Detailed Meta-Data and Indexing Systems that Deal with the Complexity of the Entities** – Social science data require detailed meta-data systems that consider many characteristics of entities including their content and context. One example of a meta-data system is the Data Documentation Initiative (DDI) which is an international effort to establish a standard for technical documentation describing social science data.
- **Software that Transform Data into Useful Measures and that Links Datasets Together** – Social scientists need better software that can code data in new ways such as fine-tuned linguistic or psychological analysis of data. There are already programs that do automated content analysis, but social scientists need an ongoing effort to improve these programs and to develop new ones with additional capabilities. There are also programs that transform statistical data from one format to another, but social scientists need programs that go much further to make it possible to link very diverse kinds of data into datasets that can be analyzed with existing techniques.
- **Sensitive Policies Regarding the Use and Archiving of Data** – Social scientists need to invent a legal and ethical framework that will allow researchers to use the data in archives while affording absolute assurances that confidentiality will be preserved.

Issues to be addressed in this session

What are some of the fundamental tasks that social science cyberinfrastructure should accomplish to provide for better linking and analysis?

- *What exists now?*
- *What can be developed in the future?*
- *What must be developed for useful cyberinfrastructure?*

What are the characteristics of effective tools?

- *What is the user interface? What is the machine interface?*
- *How do you measure effectiveness?*

How can the social science and computer science communities develop effective ways to deliver these tools since they are often either public goods or not commercially viable?

- *What kinds of services or tools are commercially viable and what are not?*
- *How can computer scientists work with social scientists to develop better tools?*
- *What are the appropriate models for collaboration between academia and the private sector?*

IMPACT OF CYBERINFRASTRUCTURE ON JOBS AND INCOME Session

Session Co-Leads: *John Haltiwanger*, University of Maryland, (haltiwan@econ.umd.edu), and *Stephen Wright*, University of Wisconsin (swright@cs.wisc.edu)

Cyberinfrastructure has had a profound influence on the organization and structure of firms and of the workplace. While rapid changes in technology have generally driven increased demand for skilled workers for the past several decades, the adoption of information technology and cyberinfrastructure has especially been associated with profound and accelerating changes in such areas as the boundaries of firms, the contractual relationships between firms and their workers, and the locus of firms and their workers. For information-technology-related products and services, global multinational companies have been restructuring the location of their activities in complex ways. Reorganization of activities within and between firms increasingly crosses national boundaries, and for this reason it poses challenges for labor relations, industrial policy, and national and international markets.

Key Questions:

- *How has CI changed the organization of the workplace?*
 - What are new stresses and capabilities that are transforming jobs (e.g. telecommuting, outsourcing, globalization)?
 - What are the challenges to making the workplace more cyber-friendly and the cyber-world more human-friendly?
 - Can a fully virtual firm operate successfully within the economy?
- *What are the implications of CI for national and international financial markets?*
 - How are liquidity, stability, and money supply affected?
 - How does CI contribute to international competitiveness/cooperation?
- *How does CI contribute to job wealth, income equality, the demand for skills, and knowledge creation?*
 - How do we measure the impact of CI investment?
 - Are there ways to make CI more accessible in the workplace?
 - What are the technological barriers to positive impact of CI on the workplace?

CYBERINFRASTRUCTURE-MEDIATED INTERACTION Session

Session Co-Leads: *Ruzena Bajcsy*, UC Berkeley, (bajcsy@eecs.berkeley.edu), and *Philip Rubin*, Yale University, (rubin@haskins.yale.edu)

The Internet has led to a true revolution in communication. It supports rapid and inexpensive text-based communication in the form of electronic mail and instant messaging. In addition, it provides for the sharing of visual and auditory information and, to a degree, even kinesthetic information, allowing for more meaningful and realistic communicative interactions. Other advances that rely on information technology are having profound effects on the communicative experience. These include mobile phones, PDAs, distributed and embedded sensors, ubiquitous and affective computing, digital imaging and music, wearable computers, GPS devices, innovative display technologies, etc. These changes have affected not only how we conduct science, but they also have had a profound effect on many aspects of our lives, including commerce, education, health care, and other behavioral, social and cultural activities. The impact of these advances can also be very broad, for example helping to alter national boundaries and, hence, contributing to globalization.

Below are some questions that we feel are pertinent to how cyberinfrastructure (CI) mediated communication has impacted human interaction. You should not necessarily limit yourself to only these questions. We ask each of you to prepare a two page statement that will address these questions and raise other issues. These statements will provide the basis for our discussion and eventually for the final product of our session during the workshop. We look forward to your statements. Please do not hesitate to contact either of us should you need any additional information. Thank you!

Questions

1. Effective CI-mediated communication in the “doing of science”

How is CI-mediated communication changing (or can it change) the relationships among researchers, effectiveness of researchers, etc. and what can be expected in the future?

What can we dream about from a technology perspective that would make a difference?

How has communication become different because of these changes?

2. How is CI-mediated communication affecting interpersonal relationships and social networks?

To what extent are relationships changing?

How are boundaries being blurred?

What are the prospects for changing these relationships through cyber-development?

3. How is communication across national and cultural boundaries transformed by the mediation of Cyberinfrastructure?

What are the implications for democratization? For increasing understanding and cooperation? For affecting stress/antagonism? For increasing economic growth? What cyber-developments will make it more natural for people to communicate across divides?

ECONOMICS OF CYBERINFRASTRUCTURE Session

Session Co-Leads: *Rich Wolski, UC Santa Barbara, (rich@cs.ucsb.edu), and Jeff Mackie-Mason, University of Michigan, (jmm@umich.edu)*

Framing Question 1: "What can economics contribute to CS research about effectively building and operating CI?"

It seems clear that effectively building, maintaining, and operating CI requires a major research effort from computer science and engineering. Because CI is intended to be large-scale, distributed, heterogeneous, and federated its realization can benefit from the theories, perspectives methods of economics. Economics as a discipline has largely specialized in the design and operation of large-scale, complex and distributed systems in which self-interested parties both collaborate and compete. Thus the design, structure, and operational principles of CI are likely to be well informed by economic research and results.

Possible topic foci include

- **Market formulations and analysis techniques for CI resource allocation and control**
 - Economic mechanisms for decentralized resource management

- **Incentive-centered systems design**
 - Reputation/recommender systems
 - Public goods preference revelation
 - Markets for high-demand resources
 - Procurement mechanisms and their impact on innovation

- **Analysis of CI productivity**
 - Investment strategies
 - Maintenance cycles
 - Efficiency and other measures of success
 - Measurements of impact on scientific output

Framing Question 2: "How can computer science help identify ways in which to effectively use CI to answer economic questions?"

The primary recommendation of the Blue Ribbon Panel for NSF to heavily invest in CI is to create infrastructure that will "radically empower" science and engineering research and education. One of the areas identified that can benefit from CI is economic research, however, most economists have little or no experience in using CI in their research. While the Computer Science community has extensive experience in building distributed systems, the most successful CI-focused research in Computer Science uses

newly achieved application capabilities as its motivation. Thus, the primary responsibility for identifying ways in which CI can empower economics research lies with economic researchers informed by a collaboration with CI systems researchers.

Possible topic foci in this area include

- **Aggregation of computational and storage capacity to facilitate large-scale simulation**
 - Monte Carlo studies
 - Empirical game theoretic studies
 - Agent-based economic analyses of complex systems
- **Distributed data analysis spanning disparate data sources**
 - Distributed Data mining and statistical analysis of federated data sets
 - Resource intensive model formulation and testing
- **Networked infrastructure for experimental studies**
 - Distributed human-subject lab experiments, data acquisition, and monitoring
 - Coupled human-subject and simulation studies
- **Integrated access to high-performance resources for numerically intensive (e.g. those requiring a supercomputer) calculations**
 - Numerical solutions of structural dynamic models
 - Coupled numerical and database intensive modeling calculations

In both questions, it should be understood that economists and computer scientists should not be called upon as mere consultants to provide service to the other field. An NSF research program should always be addressing fundamental unsolved research questions. Thus, we need to identify questions on which economists can do interesting research in collaboration with computer scientists that helps to build and operate CI. And we need to identify questions on which computer scientists can do interesting research that improves or creates new tools and methods for doing economic research.

MALEVOLENT USES OF CYBERINFRASTRUCTURE Session

Session Co-Leads: *Stephen E. Fienberg*, Carnegie Mellon University, (fienberg@stat.cmu.edu), and *Shankar Sastry*, UC Berkeley, (sastry@eecs.berkeley.edu)

The Internet has led to a true revolution in communication. It provides online access to databases that only a short while ago were available to a privileged few. It supports rapid and inexpensive text-based communication in the form of electronic mail and instant messaging. In addition, it provides for the sharing of traditional information and databases as well as visual and auditory information and, to a degree, even kinesthetic information, allowing for more meaningful and realistic communicative interactions. Other advances that rely on information technology are having profound effects on the communicative experience. These include mobile phones, PDAs, distributed and embedded sensors, ubiquitous and affective computing, digital imaging and music, wearable computers, GPS devices, innovative display technologies, etc. These changes have affected not only how we conduct science, but they also have had a profound effect on many aspects of our lives, including commerce, education, health care, and other behavioral, social and cultural activities. The impact of these advances can also be very broad, for example helping to alter national boundaries and, hence, contributing to globalization.

But the very nature of the interconnected cyberworld offers a host of new opportunities for inimical behavior on the part of individuals and groups who are intent on abusing the information to which we now have access. Such malevolent behavior conflicts with the privacy and other rights of individuals and organizations whose information is shared, either in an open or a restricted fashion. Below are some questions that are intended to get you to reflect not only on the current issues surrounding malevolent behavior but also what we should be thinking about in the ever evolving cyber world.

Questions

- 1.** Malevolence means different things to different intellectual communities. Explain how you interpret the word as it relates to cyberinfrastructure and describe one or more critical issues in dealing with malevolence where we can benefit for collaborative research at the interface of computer science and social science/statistics.
- 2.** What features of the current cyberinfrastructure create the biggest opportunities for malevolent behavior and what type of research would allow us to begin to deal with the resulting problems.
- 3.** What cyber developments are likely to raise new privacy-protection issues and how can we as a society prepare for them?

4. Is the culture and structure of the cyber world at odds with the protection of privacy and if so what can we do about this?