

The background of the cover features abstract, glowing light trails in shades of blue and white, resembling data streams or fiber optic paths, set against a dark background.

# **Workshop on Cyberinfrastructure for the Social and Behavioral Sciences**

**Final Report**

**National Science Foundation**



Suggested citation:

F. Berman and H. Brady, *Workshop on Cyberinfrastructure for the Social and Behavioral Sciences: Final Report*. (2005) <http://www.sdsc.edu/sbe/>.



# Workshop on Cyberinfrastructure for the Social and Behavioral Sciences: Final Report

**National Science Foundation**

Directorate for Social, Behavioral, and Economic Sciences (SBE)

Directorate for Computer and Information Science and Engineering (CISE)

**Francine Berman**

San Diego Supercomputer Center,  
University of California, San Diego

**Henry E. Brady**

Political Science Department,  
Survey Research Center,  
and Goldman School of Public Policy,  
University of California, Berkeley

May 12, 2005

Any views, findings, conclusions, or recommendations expressed in this report are those of the participants, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

## Workshop Organizing Committee

Ruzena Bajcsy	Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
Francine Berman	SDSC and Department of Computer Science and Engineering, University of California, San Diego
Henry E. Brady	Political Science Department, Survey Research Center, and Goldman School of Public Policy, University of California, Berkeley
John Haltiwanger	Department of Economics, University of Maryland
NSF Liaisons	
Julia Lane	SBE
Miriam Heller	CISE

## Participants

The Organizing Committee would like to thank the outstanding group of participants who devoted considerable time and thought to make the workshop a success. These participants include an excellent group of Session Co-Chairs, an outstanding group of speakers, a stellar group of participants, and an excellent and hardworking group of Workshop staff. We thank you all.

### Workshop Session Co-Chairs

Ruzena Bajcsy, U.C. Berkeley  
Francine Berman, SDSC and U.C. San Diego  
Henry E. Brady, U.C. Berkeley  
Jane Fountain, Harvard University  
John Haltiwanger, University of Maryland  
Jeff Mackie-Mason, University of Michigan  
Stephen Fienberg, Carnegie Mellon University  
Philip Rubin, Yale University  
Shankar Sastry, U.C. Berkeley  
Allan Snavely, SDSC  
Rich Wolski, U.C. Santa Barbara  
Stephen Wright, University of Wisconsin

### Workshop Speakers

Arden Bement, NSF  
Dan Atkins, University of Michigan  
Nikolaos Kastrinos, Commission of the European Union

### Workshop Staff

Nina Anderson, SDSC  
Pamela Fletcher-Rice, SDSC  
Nancy Jensen, SDSC  
Jon Meyer, SDSC

### Workshop Participants

Margot Anderson, University of Wisconsin  
Guy Almes, National Science Foundation  
Marc Armstrong, University of Iowa  
Dan Atkins, University of Michigan, Ann Arbor  
Ruzena Bajcsy, University of California, Berkeley  
Roberta Balstad, Center for International Earth Science Information Network (CIESIN)  
Peg Barratt, National Science Foundation  
Arden Bement, National Science Foundation  
Fran Berman, San Diego Supercomputer Center  
Bennett Bertenthal, University of Chicago  
Bruce Bimber, University of California, Santa Barbara  
Marjory Blumenthal, Georgetown University  
Henry E. Brady, University of California, Berkeley  
Larry Brandt, National Science Foundation  
John Brevik, University of California, Santa Barbara  
Lawrence Burton, National Science Foundation  
Claudia Carello, University of Connecticut  
Lynda Carlson, National Science Foundation  
Jeffrey Chase, Duke University  
David Clark, MIT Computer Science & Artificial Intelligence Lab  
Nosh Contractor, Science of Networks in Communities (SONIC), NCSA

Deborah Crawford, National Science Foundation  
 Anthony Cresswell, Center for Technology in Government, SUNY Albany  
 David Croson, Temple University  
 George Duncan, Carnegie Mellon University  
 Catherine Eckel, Virginia Tech  
 Barbara Entwisle, University of North Carolina  
 Joan Feigenbaum, Yale University  
 Stuart Feldman, IBM  
 Stephen Fienberg, Carnegie Mellon University  
 Tom Finholt, University of Michigan, Ann Arbor  
 Darlene Fisher, National Science Foundation  
 Jane Fountain, National Center for Digital Government  
 Kevin Franklin, University of California, Irvine  
 Peter Freeman, National Science Foundation  
 Jerry Goldman, Northwestern University  
 Austan Goolsbee, University of Chicago  
 Sucharita Gopal, Boston University  
 Shane Greenstein, Northwestern University  
 George Gumerman, Sr., School of American Research  
 Harle Lee, National Science Foundation  
 John Haltiwanger, University of Maryland  
 Miriam Heller, National Science Foundation  
 Lorin Hitt, University of Pennsylvania  
 Suzi Iacono, National Science Foundation  
 Donald Janelle, University of California, Santa Barbara  
 Daniel Jurafsky, Stanford University  
 Nikolaos Kastrinos, Directorate General for Research, European Technology Assessment Network  
 Sangtae Kim, National Science Foundation  
 John King, University of Michigan, Ann Arbor  
 Ken Klingenstein, Internet 2  
 David Kotz, Dartmouth College  
 Julia Lane, National Science Foundation  
 Eddie Lazear, Stanford University  
 John Ledyard, Cal Tech  
 Rick Lempert, National Science Foundation  
 Mark Liberman, University of Pennsylvania  
 Denise Lievesley, UNESCO Institute for Statistics  
 David Lightfoot, National Science Foundation  
 Tom Longstaff, Carnegie Mellon University  
 Arthur Lupia, University of Michigan, Ann Arbor

Jeff Mackie-Mason, University of Michigan, Ann Arbor

Brian MacWhinney, Carnegie Mellon University

Joan Maling, National Science Foundation

Steve Meacham, National Science Foundation

Walter Mebane, Cornell University

Jacqueline Meszaros, National Science Foundation

Jose Muñoz, National Science Foundation

Priscilla Nelson, National Science Foundation

Dan Newlon, National Science Foundation

David Parkes, Harvard University

Celia Pearce, University of California, Irvine

Feniosky Pena-Mora, University of Illinois

Gail Pesyna, Alfred P. Sloan Foundation

Raghu Ramakrishnan, University of Wisconsin, Madison

Karthik Ramani, Purdue University

Phil Rubin, Haskins Laboratories

Shankar Sastry, University of California, Berkeley

Walter Scacchi, University of California, Irvine

Kathryn Shaw, Stanford University

Matthew Slaughter, Dartmouth College

Allan Snavely, San Diego Supercomputer Center

Robert Strom, Kauffman Foundation

Ben Teitelbaum, Internet 2

Mark Urban, Health and Human Services, Office of Disability

Guy Van Orden, National Science Foundation

David S. Wall, University of Leeds

Dan Wallach, Rice University

Wanda Ward, National Science Foundation

Ken Whang, National Science Foundation

Rich Wolski, University of California, Santa Barbara

Steve Wright, University of Wisconsin







# Executive Summary

## BACKGROUND

The report of the Blue Ribbon Advisory Panel on Cyberinfrastructure (the “Atkins Report”) found that “a new age has dawned in scientific and engineering research” in which *Cyberinfrastructure* will play a crucial role. Cyberinfrastructure has the potential to be a fundamental enabler of innovations and new discoveries, and it is just as critical for the advancement of the social, behavioral, and economic (SBE) sciences as it is for engineering and the physical, natural, biological, and computer sciences. By participating in the development of Cyberinfrastructure, the SBE sciences can take a giant step forward.

It is equally true that SBE scientists are uniquely situated to work with computer scientists supported by NSF’s Directorate for Computer and Information Science Engineering (CISE) as well as other researchers to develop more effective Cyberinfrastructure. In addition to benefiting from and helping to design successful Cyberinfrastructure for the broad NSF science and engineering community, the SBE sciences can also help assess the effects of Cyberinfrastructure on science, engineering, technology, and society so that its potential can be realized and its benefits maximized.

## PROCESS

The National Science Foundation funded the SBE/CISE Workshop on “Cyberinfrastructure for the Social and Behavioral Sciences” in recognition of NSF’s role in enabling, promoting, and supporting science and engineering research and education. The workshop was intended to help identify the SBE sciences’ needs for infrastructure, their potential for helping CISE develop this infrastructure for engineering and all the sciences, and their capacity for assessing the societal impacts of Cyberinfrastructure. Over eighty leading CISE and SBE scientists were brought together at Airlie House in Virginia on March 15 and 16 in 2005 to discuss six areas:

1. Cyberinfrastructure Tools for the Social and Behavioral Sciences
2. Cyberinfrastructure-mediated Interaction
3. Organization of Cyberinfrastructure and Cyberinfrastructure-enabled Organizations
4. Malevolence and Cyberinfrastructure
5. Economics of Cyberinfrastructure
6. Impact of Cyberinfrastructure on Jobs and Income

Before, during, and after the Airlie House Conference, each working group produced reports on Cyberinfrastructure and the social sciences. (All Workshop reports can be found at <http://www.sdsc.edu/sbe/>). Based upon these materials as well as Workshop presentations (which can also be found on the Website), the organizers of the Conference have produced this report.

## FINDINGS

The workshop was designed to be a forum for identifying and understanding the ways that Cyberinfrastructure could facilitate social science research and the ways that the social and behavioral sciences could contribute to the development of better Cyberinfrastructure for the sciences and for society. With those goals in mind, the workshop was structured to maximize discussion and interaction. In varying degrees and from different perspectives, each of the Workshop sessions came to the following conclusions:

1. **Cyberinfrastructure can make it possible for the SBE sciences to make a giant step forward** – Cyberinfrastructure can help the social and behavioral sciences by enabling the development of more realistic models of complex social phenomena, the production and analysis of larger datasets (such as surveys, censuses, textual corpora, videotapes, cognitive neuroimaging records, and administrative data) that more completely record human behavior, the integration and coordination of disparate datasets to enable deeper investigation, and the collection of better data through experiments and simulations on the Internet.

What is revolutionary is that Cyberinfrastructure provides the ability to do these things at unprecedented scale and intensity using distributed networks and powerful tools just at a time when social and behavioral scientists face the possibility of becoming overwhelmed by the massive amount of data available and the challenges of comprehending and safeguarding it.

2. **SBE scientists can help CISE researchers design a functional and effective Cyberinfrastructure which achieves its full potential** – Cyberinfrastructure requires unprecedented organization, coordination, and integration and will have immense impact on the social dynamics, technological resources, and communication and interaction paradigms for both science and society. SBE leaders are needed to help guide the design, development, and deployment of a functional Cyberinfrastructure: Organizational researchers and political scientists can help develop appropriate management, decision-making and governance structures for Web-enabled research communities and the Cyberinfrastructure providers that support them. Economists can design incentive-compatible resource allocation methods for the sharing of multiple and diverse resources. Behavioral scientists can help develop better modes of human-computer interaction. Sociologists can analyze the implications for knowledge production of social networks developed on the Web. Psychologists and linguists can help computer scientists develop computer programs that understand, utilize, and translate natural languages. Working

together, SBE scientists and computer scientists can develop better statistical and analytical methods for dealing with data, and they can understand and control the malevolent behaviors that threaten to limit the achievement of the potential of Cyberinfrastructure.

- 3. Together, SBE and CISE researchers can assess the impacts of Cyberinfrastructure on society and find ways to maximize the benefits of Cyberinfrastructure** – Just as the Internet has forever changed the way we live and work, Cyberinfrastructure has the potential to accelerate innovation and discovery within the science and engineering community. However, it is critical to understand the way Cyberinfrastructure will impact the community and to use this information to improve Cyberinfrastructure.

It is already an accepted part of the mission of the SBE sciences to assess societal impact, but it is particularly important to assess the impacts of Cyberinfrastructure for engineering and the sciences. Social and behavioral scientists can be especially helpful in understanding changes in social interactions, changes in jobs and income, the impact of policy, and new conceptions of privacy and trust in the networked world. By increasing our understanding of these changes, SBE and CISE researchers can work with NSF communities to maximize the societal benefits from Cyberinfrastructure.

A major theme of the Airlie House Conference was that no single academic discipline or point of view is sufficient to comprehend all the implications of Cyberinfrastructure. On the technology side, the possibilities are exciting and daunting. On the human side, new challenges will arise from the unpredictable (and sometimes malevolent) uses to which Cyberinfrastructure tools and technologies are put by their users. To deal with these possibilities, both SBE and CISE can and must play key roles as research collaborators and expert consultants.

Moreover, true collaborative research is needed between SBE and CISE researchers. In order to achieve this, both intellectual and material interfaces must be shared. For example, it is not sufficient for SBE researchers to be told about Cyberinfrastructure possibilities if they do not possess the technical expertise to understand their ramifications. Many SBE researchers lack the technical know-how to participate without significant support from Cyberinfrastructure experts. Similarly, CISE researchers often lack sufficient domain-specific knowledge to appreciate the complexity of the technical problems that truly need to be solved by SBE researchers. The level of knowledge required by both sides will require true collaboration between the two research communities to make a joint research initiative successful. SBE researchers must become familiar with emerging Cyberinfrastructure technologies and CISE researchers must learn about the social sciences.

## RECOMMENDATIONS AND CHALLENGES

Recommendations from the Airlie House conference are listed throughout the following report. Section VI of the report provides a summary of the recommendations organized into the following eight overarching areas:

**Summary Recommendation 1:** Develop and deploy enabling data-oriented Cyberinfrastructure targeted to the social and behavioral sciences.

**Summary Recommendation 2:** Develop and deploy targeted toolkits, virtual, and computational environments for facilitating social and behavioral science research.

**Summary Recommendation 3:** Instrument and design technologies to gather and provide key data for social scientists. Conversely, utilize human and computer interaction data to instrument and design Cyberinfrastructure technologies.

**Summary Recommendation 4:** Ensure that confidentiality, privacy, and other social and policy considerations are included as part of the architecture of Cyberinfrastructure.

**Summary Recommendation 5:** Involve social and behavioral scientists in the design of organizational frameworks, incentive structures, collaborative environments, decision-making protocols, and other social aspects of Cyberinfrastructure.

**Summary Recommendation 6:** Develop adequate funding models for Cyberinfrastructure that will enable social and behavioral science research.

**Summary Recommendation 7:** Develop explicit venues for funding inter-disciplinary SBE and CISE research on the social impacts of Cyberinfrastructure.

**Summary Recommendation 8:** Develop the community for Cyberinfrastructure and social sciences through targeted funding programs, meetings, workshops, conferences, and other activities.

These recommendations focus on what needs to be done. In every case, the NSF can and should play a major role by developing projects, programs, and priorities that implement these recommendations.

Finally, throughout the report, we have included “Moonshot Challenges”<sup>1</sup> to represent major undertakings that would have an immense impact on science and society. The eight “Moonshot Challenges” included in this report (and scattered through the text) are:

**Moonshot Challenge:** Taking Society’s Temperature

**Moonshot Challenge:** Better Communication through Cyberinfrastructure

**Moonshot Challenge:** Cyberinfrastructure to Promote U.S. Leadership and Competitiveness

<sup>1</sup> The Organizers would like to thank Dan Atkins for the term “Moonshot Challenge.”

***Moonshot Challenge:*** Re-designing the Internet

***Moonshot Challenge:*** Better Economic and Political Institutions through Experimentation

***Moonshot Challenge:*** Better Information for Better Policy

***Moonshot Challenge:*** Cyberinfrastructure to Guard against Natural and Societal Threats

***Moonshot Challenge:*** Cyberinfrastructure to Promote Effectiveness and Productivity

The recommendations and challenges in this report make it clear that the social and behavioral sciences can be significantly advanced through Cyberinfrastructure, and that social and behavioral scientists are a critical part of the team needed to develop a functional, effective, and successful Cyberinfrastructure. Furthermore, NSF has a major role to play through its CISE and SBE Directorates. We hope that this report, and the outcomes of the Workshop, help move the SBE/CISE community forward to achieve these important goals.

***A Note about the Report:***

This report is the result of a true collaboration between the four workshop organizers--two of whom are computer scientists and two of whom are social scientists, between the CISE and SBE Directorates at NSF, between the 80 academics and NSF staff who attended the Airlie House conference who were drawn equally from the computer science and social science communities, and between the two authors, Francine Berman (a computer scientist) and Henry E. Brady (a social scientist). We hope that this collaborative effort, at many levels, sets the stage for future collaborations between the computer science and social science communities.



# Table of Contents

Workshop Organizing Committee .....	2
Participants .....	3
Executive Summary .....	7
<b>I. Introduction</b> .....	<b>15</b>
<b>II. Developing Cyberinfrastructure to Support and Enable the SBE Sciences</b> .....	<b>19</b>
A. What Cyberinfrastructure Can Do for the Social and Behavioral Sciences.....	19
B. Data Collection.....	21
C. Algorithms, Tools, and Computing Power for Analyzing Data and Models .....	25
D. Tools for Data Comparison and Measurement .....	26
E. Methods of Data Storage and Archiving .....	30
F. Communication and Collaboration.....	32
G. Challenges and Opportunities.....	34
<b>III. SBE Helping CISE Design Cyberinfrastructure</b> .....	<b>35</b>
A. What the Social and Behavioral Sciences Can Do for Cyberinfrastructure.....	35
B. Better Interfaces .....	35
C. Better Organizations and Institutions .....	36
D. Deterring and Controlling Malevolence .....	39
E. Better Resource Allocation and Incentive Systems.....	41
F. Challenges and Opportunities.....	43
<b>IV. Assessing the Societal Impact of Cyberinfrastructure</b> .....	<b>45</b>
A. How Social and Behavioral Sciences Can Assess the Impact of Cyberinfrastructure.....	45
B. Changes in Social Interactions.....	45
C. Changes in Jobs and Income.....	47
D. Privacy and Trust in the Networked World .....	49
E. Challenges and Opportunities.....	51
<b>V. Potential Roles of SBE and CISE</b> .....	<b>53</b>
A. Cyberinfrastructure for the SBE Sciences.....	53
B. SBE Helping CISE Design Scientific Infrastructure.....	56
C. CISE and SBE Assessing the Societal Impact of Cyberinfrastructure .....	57
D. Interdisciplinary, Not Just Multidisciplinary Research.....	57
<b>VI. Summary Recommendations</b> .....	<b>59</b>





# Introduction I

Cyberinfrastructure (CI) enables and supports scientific research through online digital instruments, emerging sensor and observing technologies, high-powered computers, extensive data storage capabilities, visualization facilities, and networks for communication and collaboration. The report of the Blue Ribbon Advisory Panel on Cyberinfrastructure (the “Atkins Report”) signals that the sum of these changes constitutes “a new age” which “has crossed thresholds that now make possible a comprehensive ‘Cyberinfrastructure’ on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research in new ways and with increased efficacy.” Science and engineering are being transformed by Cyberinfrastructure.

This is just as true of the social, behavioral, and economic (SBE) sciences as of the physical, natural, engineering, and biological sciences. **With the development of needed, appropriate, and usable Cyberinfrastructure, the SBE sciences can take a giant step forward.** Cyberinfrastructure can enable the development of more realistic models of complex social phenomena, the production and analysis of larger datasets (such as surveys, censuses, textual corpora, videotapes, cognitive neuroimaging records, and administrative data) that more completely record human behavior, and the collection of better data through experiments and simulations on the Internet. Moreover, the revolutionary potential of Cyberinfrastructure is the ability to do these things at a much greater scale and intensity using distributed networks and powerful tools just at a time when social and behavioral scientists face the possibility of becoming overwhelmed by the massive amount of data available and the challenges of comprehending and safeguarding it.

**It is equally true that the SBE sciences are uniquely situated to help computer scientists supported by NSF’s Directorate for Computer and Information Science and Engineering (CISE) and their community create better Cyberinfrastructure for all the sciences and engineering.** Behavioral scientists can help develop better modes of human-computer interaction. Sociologists can analyze the implications for knowledge production of social networks developed on the Web. Organizational theorists and political scientists can develop better management and governance structures for Web-enabled research communities and the Cyberinfrastructure providers that support them. Economists can design incentive compatible resource allocation methods. Psychologists and linguists can help computer scientists develop computer programs that understand, utilize and translate natural languages. Working together, SBE scientists and computer scientists can develop better statistical and analytical methods for dealing with data, and they can understand and control the

“The emerging vision is to use Cyberinfrastructure to build more ubiquitous, comprehensive digital environments that become interactive and functionally complete for research communities in terms of people, data, information, tools, and instruments and that operate at unprecedented levels of computational, storage, and data transfer capacity. Increasingly, new types of scientific organizations and support environments for science are essential, not optional, to the aspirations of research communities and to broadening participation in those communities. They can serve individuals, teams, and organizations in ways that revolutionize what they can do, how they can do it, and who participates. This vision also has profound broader implications for education, commerce, and social good.”

--Executive Summary, page 2,  
*Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure*, (Atkins Report), 2003.

malevolent behaviors that threaten to limit the capabilities of new Cyberinfrastructure.

In addition to benefiting from and helping to design successful Cyberinfrastructure for all the sciences and engineering, **the SBE sciences can also assess the effects of Cyberinfrastructure on all of society.** This task is already an accepted part of the mission of the SBE sciences, but it is also a natural outgrowth of efforts to develop better Cyberinfrastructure for the sciences and engineering. After all, the Internet is the result of technical innovations that were initially confined to research communities but which expanded to society at large; it seems likely that many future Cyberinfrastructure innovations for research will also find their way into mainstream society. In addition, the vast changes expected from society-wide Cyberinfrastructure must be studied and understood to better channel and control them.

The National Science Foundation funded the SBE/CISE Workshop on “Cyberinfrastructure for the Social and Behavioral Sciences” in recognition of the SBE sciences’ needs for infrastructure, their potential for helping CISE develop this infrastructure for all the sciences, and their capacity for assessing the impacts of society-wide Cyberinfrastructure. This workshop had three goals:

1. Provide focused recommendations and a path for Cyberinfrastructure research, experimentation, and infrastructure for the SBE/CISE community.
2. Provide a framework and initial ideas for projects and efforts in this area.
3. 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, expertise and research at the frontiers of both SBE and CISE.

To achieve these goals, over eighty leading CISE and SBE scientists were brought together at Airlie House in Virginia on March 15 and 16 in 2005 to discuss six areas:

1. Cyberinfrastructure Tools for the Social and Behavioral Sciences
2. Cyberinfrastructure-mediated Interaction
3. Organization of Cyberinfrastructure and Cyberinfrastructure-enabled Organizations
4. Malevolence and Cyberinfrastructure
5. Economics of Cyberinfrastructure
6. Impact of Cyberinfrastructure on Jobs and Income

In varying degrees, each of these sessions considered the three major challenges described above:

1. Developing Cyberinfrastructure to support and enable the SBE sciences
2. Helping CISE design Cyberinfrastructure to support and enable all the sciences
3. Forging a partnership between SBE and CISE to assess the impacts of Cyberinfrastructure on society.

The first two tasks clearly fit within the Cyberinfrastructure initiative proposed by the Blue Ribbon Advisory Panel on Cyberinfrastructure which considered Cyberinfrastructure needs for the sciences. We believe that the third is equally important because Cyberinfrastructure, much of it emanating from CISE and SBE research, will radically transform society in ways that must be studied and understood. To take two examples, the need to control the malevolent uses of the Web and the opportunity to revolutionize the workplace through the use of massive databases, powerful operations research tools, and complex query engines require a joint CISE/SBE collaboration. This collaboration goes beyond the goal of simply supporting the sciences with better Cyberinfrastructure. It goes to the heart of helping America make a successful transition into the 21<sup>st</sup> century. Moreover, it is in complete agreement with the Blue Ribbon Advisory Panel's observation that their vision "has profound broader implications for education, commerce, and social good." If the development of Cyberinfrastructure is, as we and the Blue Ribbon Panel believe, a Promethean enterprise, fraught with both great promise and great danger, we must study and understand its implications for society.



# Developing Cyberinfrastructure to Support and Enable the SBE Sciences

## II

### A. What Cyberinfrastructure Can Do for the Social and Behavioral Sciences

New methods for collecting and analyzing data have repeatedly advanced the social and behavioral sciences. Data collection and organizing methods such as the following have made it possible for the social and behavioral sciences to record more and more information about human social interactions, individual psychology, and human biology<sup>2</sup>:

- national income accounts
- psychological testing and measurement
- map making and geographic information systems (GIS)
- audio and video recording of language, social interactions, and culture
- field archaeology
- genetic sampling
- national censuses
- cognitive neuroscience techniques (fMRI, PET, MEG, ERP)
- survey sampling
- laboratory and field experiments

Not all these methods were invented by the social and behavioral sciences (although some of them were), but all of them have been adapted, applied, and further developed and enhanced by work in the SBE sciences. It is worth reflecting on what the world would look like without them. National income accounts make it possible to monitor and control modern economies. Psychological testing and measurement are widely used to assess educational, medical, and other outcomes. GIS now makes it possible to move people, goods, and services from one place to another with even greater effectiveness. National censuses are used every day by businesses and governments to make decisions. Survey sampling is used by marketing experts, government decision-makers, and others to gather information that is representative of populations. Experiments are used to test the effectiveness of programs and products. Field archaeology has taught us about our origins.

---

<sup>2</sup> The social and behavioral sciences also study the behavior of primates and even the behavior of animals such as rats and insects such as ants.

Social and behavioral scientists have also developed, borrowed, and adapted analytical techniques such as the following which have enabled breakthrough research and understanding about human interactions, psychology, and biology:

- statistical methods (biometrics, psychometrics, econometrics, spatial analysis)
- archaeometry
- game theory, linear programming, decision analysis
- content and textual analysis
- genetic analysis
- nonlinear analysis and methods
- linguistic annotation
- simulations and agent based modeling

Cyberinfrastructure provides a mechanism to use these methods and techniques more effectively, facilitating and improving both data collection and data analysis. This will significantly increase social and behavioral scientists' capacity for making comparisons, undertaking measurements, and searching for patterns. The revolutionary potential of Cyberinfrastructure for SBE and other domains is the ability to do these things at a much greater scale and intensity using distributed networks and powerful tools.

Other fields are being transformed by better Cyberinfrastructure. Environmental and atmospheric sensing methods combined with powerful computing make it possible to model and even predict weather and climate with increasing reliability. Genetic sequencing methods combined with computerized databases and analysis tools make it possible to understand human evolution and human disease. Better social and behavioral science Cyberinfrastructure will make it possible to:

- Model life-time decision-making by individuals with respect to work, marriage, saving, and retirement by following large numbers of people over time and devising models that take into account the full complexity of these decisions.
- Model research and development investments and returns with respect to workforce issues and scientific careers.
- Code the verbal and non-verbal cues in large numbers of video-taped interactions such as physician-patient interactions which have substantial consequences for proper medical diagnoses.



- Comprehend changes in metropolitan areas over time by simultaneously geo-coding and temporally-coding land-use, environmental, social interaction, institutional, and other data for a large area over a long time.
- Enhance our understanding of the brain activity underlying decision-making processes by the meta-analysis of fMRI data across many individuals and tasks.
- Develop and analyze databases of tens of thousands of legislative votes, speeches, and actions in order to better understand the functioning of government.
- Track change in human behavior at multiple time scales and from multiple perspectives.
- Understand the development and functioning of social networks on the Web by coding message frequency and content over time and space.
- Develop better institutional and technical methods to reduce malevolent behavior on the Web by understanding not only the Web's technical vulnerabilities but also the realistic and feasible threats from human agents.

These last two examples are especially important. Cyberinfrastructure itself is creating new modes of interaction and new human artifacts that need to be understood, managed, and sometimes controlled. Social and behavioral scientists have much to contribute to this endeavor, but they, in turn, need the tools that will make it possible to do so. Computer scientists can not only help social scientists do better work – social scientists can also help computer scientists develop the best ways to implement new Cyberinfrastructure. Both computer scientists and social and behavioral scientists have a stake in developing the best possible Cyberinfrastructure tools for studying human behavior.

Cyberinfrastructure can help solve these problems because it provides unprecedented potential for advances in data collection and integration, computing power for analyzing data, tools for data comparison, methods of data storage, and communication and collaboration.

## **B. Data Collection**

The Internet has led to a true revolution in communication. It supports rapid, location-independent, 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 computing, digital imaging and music, wearable computers, GPS devices, and innovative display technologies. 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 Internet now potentially connects every human and human-made institution and makes them accessible and available for data collection. For example, Time Sharing Experiments for the Social Sciences (TESS) uses Internet-based surveys to collect data on survey subjects in randomized experiments that are themselves submitted, reviewed, and implemented on the Web. The Iowa Electronic Markets use the Web to collect information on people's beliefs through the operation of real-money futures markets in which contract payoffs depend upon economic and political events such as elections. Internet based experiments and virtual organizations on the Web are just a few of the possibilities for using Cyberinfrastructure to develop new ways to learn about human behavior. These methods make use of Cyberinfrastructure's ability to connect people in new ways and at low cost, and they can increase the generalizability of our studies by using random selection or stratification methods that allow for a broader range of subjects and conditions.

The availability of these enabling technologies signals a paradigm shift. For example, innovative randomized experiments, such as simulated economies, politics, or social systems involving hundreds or even thousands of people, which could not be carried out in the laboratory (and which would suffer from artificiality), should be possible on the Web. These experiments can randomly assign treatments to subgroups so that reliable inferences can be made about the impact of changes in economic or social conditions. To date, economic, political, and social experimentation with human subjects is mostly performed in small single-room labs with 12-30 workstations. Building, administering and operating such labs at each researcher's location is expensive and inefficient. It is impractical to build labs large enough to perform experimental protocols involving the interactions of large numbers of humans, which better match the conditions the scientist is trying to test.

Creating the physical and especially software infrastructure to implement large scale virtual labs by federating facilities across multiple locations, will reduce research costs and enable important experimental investigations not possible today. It will also support greater diversity in subject pools and opportunities to have participants with domain expertise (e.g., business and engineering professionals who cannot all be brought to a single location for group experiments). The work required to make these virtual labs a reality includes creating standardized toolkits, high-level language, and transparent application program interfaces (APIs). There is also a need for human-computer interface design for economic communication: learning and expressing preferences, rules, and strategies.

The ultimate purpose of building geographically dispersed, networked experimental lab infrastructure is to pursue important research questions. The goals of large-scale human subject/agent experimentation and simulation include:

1. Eliminate bad system designs quickly (for example, avoiding the costly and damaging problems that some recent policy-created markets have encountered before field deployment).
2. Test policies and methods on subjects with domain-specific knowledge or experience.
3. Implement experiential learning methods in large (200+ student) classrooms.

The Web also makes it possible to collect machine readable versions of media, governmental, and statistical information which used to be buried in widely dispersed reference libraries, if available at all. Researchers can obtain national statistics, reports of voting in legislatures, texts of speeches, news reports, personal blogs, and many other types of information. As a result, researchers can put together large databases on almost any conceivable subject, which can be used to advance scientific knowledge. This is possible, however, only so long as tools are robust enough to handle the extra data flow, intelligent enough to ensure that the data are representative of important populations, and intuitive enough to enable humans to manage and make sense of the increased data. And it is only desirable to collect this data if mechanisms are in place to protect individual privacy.

One of the novel opportunities is to collect data on the increasingly large volumes of transactions and interactions (purchases, recruitment to organizations, information searches, political debates, or discussions) that flow through the Internet, creating detailed micro-activity datasets of a sort that have almost never been previously available. The opportunities are already visible in the Web services efforts of firms like Amazon, which now makes freely available all of its product listings with full descriptive characteristics, based on historical activity patterns of account holders. Recording and organizing such data for transactions (with the usual necessary privacy safeguards, of course) would be invaluable.

Data can also be gathered externally to the Web from devices such as embedded field-sensors (e.g., traffic-flow sensors), GPS enabled devices including cell phones, hand-held PDAs, sensors monitoring food consumption, microscopic “smart dust” scattered invisibly everywhere to sense light, vibration, chemicals and even patient movements, and remote-sensing satellite systems. As a result, almost any observational study or experiment that in the past might have been done with a limited set of human volunteers or via human observation can now be scaled up many-fold.

Such a deluge of potential data raises problems and challenges as well as opportunities. Social scientists face the challenge of collecting many different kinds of data (e.g., administrative, video, qualitative, survey, biomedical specimens, digital, analog) with many different units of observation (e.g., thoughts, speech-acts, interactions, individuals, couples, households, neighborhoods, friendships, person-months). The tremendous variety of data stretches the capabilities of even the most versatile data indexing systems and relational databases. But there is a general need to preserve and record data over the long-term to enable future research (for example, recording endangered languages for posterity).

These vastly heterogeneous data need to be collected, archived, and most importantly, reduced to manageable proportions and made useful for social scientists. There are very few institutions that perform these functions. Libraries, museums, centers, and other institutions are grappling with the issues of dealing with today’s deluge of data in a stable, reliable, and evolutionary way. Note that policy plays a substantive and important role as a portion of this data involves delicate and difficult ethical questions about individual privacy.

**Workshop Recommendations for Data Collection:**

- *New Kinds of Data Collections* – Cyberinfrastructure should be designed to make it possible for the social and behavioral sciences to collect new kinds of data in a broad range of human behavioral, environmental, and ecological settings. For example:
  - Large-scale “virtual labs” should be developed using networked infrastructure which can involve human beings and artificial agents in experiments and simulations of economies, organizations, political systems, and social systems.
  - Large-scale collections of transactional and administrative data should be constructed. Existing datasets should be combined into long-term time-series, into detailed cross-sections of information at many different scales, and into many other formats. This requires the development of toolkits for data integration, data validation, and data analysis scaled to the size and federated nature of these massive data sets. It also presumes that efforts will be made to find ways that social and behavioral science researchers can take advantage of the data contained in administrative and interaction systems developed for the Web and for organizational administration.
  - A new scale and a new generation of virtually immersive environments should be developed that will allow experiments that could not be done without Cyberinfrastructure. These environments allow the experimenter to create virtual worlds in which all stimuli but the experimentally manipulated stimuli are constant. This makes it possible to study human decision-making in complex or dangerous (e.g., driving, flying, reacting to surprising stimuli) situations.
- *New Kinds of Data Collection Methods* – Cyberinfrastructure should leverage emerging sensing and data collecting methods to provide new kinds of data for the social and behavioral sciences:

**Moonshot Challenge: Taking Society’s Temperature**

Today, polls and surveys are widely used to “take the temperature” of society on a broad spectrum of issues and topics. From the determination of the monthly growth in jobs to the assessment of people’s health, survey information is used to take stock of our society and to make important public policy and business decisions. For all kinds of surveys, the Internet offers unprecedented potential for gathering information.

However the need for quality control and the potential for bias from non-random samples make it hard to use the Internet fully and effectively. The challenges to developing an efficient environment for Internet sampling include inadequate access and connectivity among racial, ethnic, and lower income groups, accurate weighting of self-selected samples, and the problems of ensuring that Web surveys provide reliable and representative information. As with all other methods of data capture and storage, privacy and confidentiality pose serious challenges.

Gathering large-scale Internet samples of useful social information to “take society’s temperature” is a Moonshot, whose achievement will require a multi-disciplinary team of social and computer scientists and statisticians and targeted Cyberinfrastructure.

- Cell phones with GPS, PDAs, fMRI's, embedded sensors, satellite imaging, and other technologies should be designed to collect data for the social and behavioral sciences. To avoid overwhelming our available resources, data collection procedures should be designed and instrumented to cut off the collection of irrelevant data at its source, and they should be designed to make it possible to fuse and link data as easily as possible.
- Cyberinfrastructure developers should provide tools for managing ubiquitous, possibly mobile data collection devices, including human user interfaces (UIs) that are well designed for robustness and intuitive for social scientists to use.
- Methods of using the Web to collect survey information should be developed and tested with large-scale comparisons with in-person and telephone surveys as checks on the results from Internet surveys.

### C. Algorithms, Tools, and Computing Power for Analyzing Data and Models

Theoretical progress in many crucial areas is constrained by limits on our abilities to obtain analytic solutions to complex mathematical models. For example, economists have designed the Federal Communications Commission's spectrum auctions and the participation strategies for bidding firms. These auctions have raised tens of billions in revenue from licenses for public airwaves. However, despite significant theoretical efforts, it has not been possible to solve analytically for the optimal auction design; nor has it been possible to solve for optimal bidding strategies for participants. As a consequence, substantially greater revenue and the achievement of an even better allocation of the communications spectrum have been foregone.

One of the most promising methodologies for complementing our limited analytic capabilities is the application of computationally-intensive numerical analyses. Numeric methods have been developed to obtain arbitrarily precise approximations to analytic solutions, to solve otherwise intractable statistics problems with Markov Chain Monte Carlo methods, to use agent-based computational models to study complex systems and their dynamics, and to use massive Monte Carlo simulations to apply empirical game theoretic reasoning to difficult problems. All these methods are confounded by the curse of dimensionality, and in many cases by exponential run-times, thus they require the massive distributed computing power available from forthcoming Cyberinfrastructure.<sup>3</sup>

Social and behavioral scientists also need more raw computing power for the intensive signal processing required for speech analysis and recognition and the extraction of meaning and other linguistic characteristics. The need for significantly increased computational power arises as well in agent-based models involving thousands or even millions of simu-

<sup>3</sup> Alternatively, they require more work to develop accurate analytical approximations and algorithms which can be implemented and tested using Cyberinfrastructure.

lated decision-makers over many different scenarios, complex geographic models involving many levels of data and complicated interactions among factors over time, and complex and noisy fMRI data from many different subjects. Massive real-time computing resources have also made it possible to create new virtual environments and simulations involving people in laboratories or on the Web that mimic reality. The result is a significant increase in the ability to test theories in the social and behavioral sciences. Moreover, these methods can help computer scientists understand the implications of the systems they are designing by providing the opportunity to test them before implementing them.

Social and behavioral scientists face special problems in these areas including limited computer literacy and unfamiliarity with computationally intensive methods in some cases, lack of cultural support within their disciplines for those who are heavily involved with computers, and small computing budgets for the social and behavioral sciences.

#### ***Workshop Recommendations for Computing Power for Analyzing Data and Models:***

- *Support for More Computing Power* – More attention should be paid to providing computing cycles, data storage, and support to social and behavioral scientists who are heavy computer users of computationally intensive methods, but who are sometimes isolated from others with similar interests. Social and behavioral scientists should be fully involved in efforts to increase access to computing resources through grid computing and other methods.
- *Support for Better Algorithms and Methods* – Computing power must be complemented by powerful tools. For example, as noted by a companion workshop involving the CISE and operations research communities, tools for optimization and operations research will be essential for extracting meaning from data and for making projections about future outcomes. This workshop has proposed the development of an “Operations Cyberinfrastructure” that incorporates tools for gathering and distributing data, tools for modeling and solving very large optimization and equilibrium problems, and tools that target specific problems arising in supply network management and logistics. These tools must be able to handle nonlinearities, discrete and combinatorial aspects of models, and data uncertainty (stochastic optimization) as well as an enormously large scale.

#### **D. Tools for Data Comparison and Measurement**

Data comparison is a basic step in any scientific endeavor. Typically scientists are looking for similarities and differences in data that suggest causal regularities or meaningful patterns. Here are two simple examples. Google searches for Web pages with *similar* content based on search words and indicators about the importance of pages (such as “page-rank”), whereas SETI@home searches for signals, possibly from extraterrestrial intelligent life, which are *different* from random background noise in the universe. These two Cyberinfrastructure tools automate comparison processes that would be too painful to even consider doing by hand. They also incorporate notions of prediction, causality, and meaningfulness



which try to distinguish spurious association from true relevance. In the social and behavioral sciences, data comparison tasks may be further complicated by different formats from different domains. For example one might wish to compare information from quantitative political polls and video political advertising or one might want to link data collected at different scales such as a household-level survey with data at the census-tract, Congressional district, county, or state level.

Tools are needed to make it possible to link these diverse kinds of data and to connect detailed annotations with data in a systematic and useable form. Standard formats for data collections can help, but better metadata descriptions (XML) are also needed. Once these are in place, tools must be available for linking data together in ways that allow for statistical analysis. One of the problems faced by many social and behavioral scientists is that they are accustomed to analyzing rectangular datasets (cases by variables) that fit into the classic matrix format, but data increasingly come in much more complicated forms. Moreover, computer scientists tend to think of relational databases as the paradigm for data storage, but these databases do not always lend themselves to analysis using rectangular dataset methods.

Statisticians and computer scientists have developed techniques that transcend these limitations, but their transfer into the social and behavioral sciences has been slow. Social and behavioral scientists use hierarchical linear models for educational data in which students are nested in classrooms located in schools which are part of school districts, but they are just beginning to learn about methods for linking and analyzing text and picture segments, strings of related data, or data with different spatial and temporal granularities. Much more needs to be done to bring these methods into the social and behavioral science and to develop new methods that can apply to even more anarchic data structures. Among other things, these methods will not only help social and behavioral scientists, they will also help those designing better ways to browse and search the Web.

An example of the usefulness of federating and analyzing large data-sets constructed from multiple loosely coupled sources is a study of the impact of Enhanced 911 (E911) emergency response services on heart attack survival rates.<sup>4</sup> This research required the federation of telecom, ambulance and hospital databases: the authors combined over 100,000 ambulance records with over 1.7 million hospital records and the results of a telephone survey, among other data. They were limited by the scale of the data problem to just one county in Pennsylvania, but their results answer important questions about information technology and health resources.

To be most useful, federated databases must be documented and maintained. To do so, technological challenges must be overcome, but more than technological developments are needed. Data must be documented in ways that make it useful to others. Privacy, data integrity, and accountability requirements must be addressed by researchers, custodians, and security personnel. For example, if public policy were to result from the analysis of federated data, free access to the public data portions might be mandated. At the same time, since the analysis might need to be reproduced in different legal contexts, the federated data would need to be protected and auditable. Finally, the data must be maintained

<sup>4</sup> Susan Athey and Scott Stern, "The impact of information technology on emergency health care outcomes", *RAND Journal of Economics*, 33(3): 399-432 (2002).



in a way that will ensure that they are available as long as they are needed. Thus, the federation of economic data requires both human and technological assets to be developed, maintained, and evolved, and in so doing, the resulting capability necessarily becomes a piece of shared Cyberinfrastructure.

Developing better ways to code and annotate human speech and human action is a common challenge for both computer scientists and social and behavioral scientists. Advances in this area would help both communities enormously. There is, for example, a pressing need for browser-based annotation tools for storing collaborative commentary and argumentation (Weblogs) that are attached to particular kinds of data, especially video, audio, and transcript segments, but, in the longer-run, ways must be found to replace at least some human coding and commentary with automatic preprocessing and filtering to assist researchers in coping with volume. Such automation is currently a challenge (mostly work to be done) but some first-pass semantic parsing can be done now; examples include automated segmentation in text by detection of topics and similar meanings, (i.e., first pass semantic text parsing) and methods for automatically indexing and linking related data. These methods can not only help social and behavioral scientists but they can also aid computer scientists trying to improve human-computer interaction and those trying to develop better ways of searching for objects on the Web. In the long-run, video and audio coding for gestures, speech, emotions, and interactions could greatly advance these fields.

Creating measures from data is another basic challenge for both communities. There is an opportunity for social and behavioral scientists to help Cyberinfrastructure design here as well as improving scientific research. Cyberinfrastructure is badly in need of metrics for usability, satisfaction, and utility if Cyberinfrastructure is to be designed to meet the needs of human beings. Similarly, for social and behavioral scientists, tools are needed to develop better measures of populations, speech patterns, the economy, and human behaviors.

***Workshop Recommendations for Tools for Data Comparison and Measurement:***

To deal with the volume of data that can now be acquired, further automation is needed to assist in the first phases of information extraction. For example, rich data formats such as audio and video pose challenges to automatic information extraction (Google searches for images only return pages with similar text titles, not necessarily with similar visual content) so additional effort in tools development is needed here. Similarly, Internet-accessible tools for acquiring and processing geographic information need enhancement for the visualization and analysis of space-time data and for the modeling of behavioral, social and economic processes. Work in the following domains is needed:

- *Data Integration and Linkage* – Data linkage and the integration of text, video, audio, and map information pose challenges in the social and behavioral sciences, as there is often a need to integrate many types of diverse data into a single database. This poses technological challenges for computer scientists as well.
  - Tools for automatically indexing, linking, and querying distributed repositories of social science data are needed which utilize metadata standards for describing the temporal and geographical footprints of the data.

- For example, the data required for studying the impact of the changing Cyber-infrastructure on the workplace for both firms and workers will require rich integration and analysis of data from multiple modes. Data at both the individual and firm level, as well as national and international levels, must be integrated. Complex qualitative and quantitative data from a variety of modes must be combined. Data on traditional firm and worker outcomes like outputs, employment, capital assets, materials use, profits, wages, and prices should be integrated with multi-modal data (captured by video interviews, monitoring, and sensors) on the nature of organizational teams, networks, and human resource practices.
- *Annotation and Processing* – Data sources such as video, satellite images, audio, and texts must be annotated and processed before they can be analyzed, and they often require special analysis methods.
  - Tools for annotating audio and video, whether in automated or semi-automated fashion, should be improved.
  - Tools for data comparison should be extended to deal with audio and video. One should, for example, be able to query a video database on child-parent interactions for certain stylized behaviors.
  - Social and behavioral scientists must become more familiar with statistical methods for analyzing non-rectangular datasets and nonlinear phenomena; these methods must be extended to the specific problems faced by the SBE sciences.
  - Tools that enhance Geographic Information Systems for the seamless integration of temporal data, for pattern recognition, and for space-time-process modeling should be developed to deal with a broad set of interdisciplinary problems.
- *Coding, Managing, and Extracting Meaning from Video and Other Forms of Data* – Video, audio, textual, and related kinds of data present special opportunities for the SBE sciences.
  - A conference bringing together social, behavioral, and computer scientists should consider the possibilities and opportunities for improving the coding, managing, and sharing of audio, video, and textual data. Special attention should be paid to methods for extracting meaning from these forms of data by coding speech, gestures, emotions, and facial expressions.
  - Infrastructure should be developed that supports multi-modal, multi-media repositories and virtual worlds, including tools for annotation, analysis, visualization, and extraction of meaning. Key developments will revolve around the continuing evolution of devices as extensions of human memory and of human reasoning capacity.

### E. Methods of Data Storage and Archiving

Storage is cheaper than ever and follows a sort of “Moore’s Law” whereby capacity for the dollar doubles approximately every 2 years. Thus there is an opportunity to capture and preserve unprecedented volumes of social and behavioral science data. In some cases the opportunity is fleeting, as with endangered languages. Also there are challenges of converting and preserving analog archives such as the National Broadcast Archives that contain hundreds of million hours of spoken word. This is an instance of a more general problem. How can we ensure long-term preservation of important data across technological changes from computer tape, to disk, to CD, and onwards? One could argue that today we have more data but also that it is more ephemeral than in the past. The Egyptians carved their important data in stone that has lasted thousands of years while all data recorded on floppy disks ten years ago is in immediate danger of being lost due to lack of devices that can read and write it and the relative fragility of the media. These issues with longevity of technology and long-term preservation are being addressed by organizations such as the Inter-University Consortium for Political and Social Research (ICPSR at <http://www.icpsr.umich.edu>) and projects such as the SDSC’s Chronopolis project (digital preservation across space and time), but more needs to be done.

Storage is only part of the problem. Efforts must also be made to create the institutional structures and norms for preserving and protecting information. Researchers must be encouraged, if not required, to archive their data using commonly accepted contribution and indexing standards. Institutions, akin to libraries, must be established to collect, preserve, and disseminate these data, but at the moment, there is nothing like a nationwide “library” infrastructure available that can complement the ICPSR. The ICPSR in Ann Arbor, Michigan provides a central repository for data that are made widely available through the Web to member institutions, but the ICPSR must be complemented by institutions around the country that work with local researchers, governments, and other groups to collect, document, and preserve data. Too often, data collections are maintained by individuals, and they depend upon the long-term commitment and effort of these individuals. Methods must be developed to provide for institutional support and continuity that go beyond merely depositing data in a central location.

Standards must also be developed for protecting privacy and for ensuring confidentiality. There is now a considerable amount of research in social statistics on confidentiality and privacy protection for individual level data, but only some aspects of this confidentiality and disclosure limitation research transfer easily to more extensive data-sets such as those with information about everyone in an organization, those that link audio, video, and other multi-media data, or those that involve a highly integrated structure linking people and their interactions and actions. Computer scientists have worked on privacy-preserving multi-party computation which offers one way to think about protecting privacy, but much more needs to be done to bring computer scientists and social scientists together to think about these issues.

Privacy protection was a common theme at the Airlie House conference, and many believe that the problem of finding ways to meet the legitimate privacy and confidentiality concerns of human subjects is the Achilles heel of the current data explosion. One overarching

approach is to limit access to data, but this can severely restrict the possibilities for data analysis that can solve pressing human problems such as improving education, improving health care, or reducing crime. Mechanisms such as the Census Research Data Centers provide very strong protections of data by allowing researchers limited access to non-public information under very strict conditions, but they set-up many obstacles for potential users. Artificial datasets that mimic the actual data while preserving the multivariate relationships in the original data might be useful, but we do not yet know whether researchers will find these datasets to be informative. Methods for automated disclosure analysis that would allow a researcher to submit an analysis plan which would be automatically checked to ensure that disclosure risks are minimal might also be helpful, but these methods are still in their infancy.

More work needs to be done in all these areas, but if we are to preserve the possibilities for socially useful data analysis from micro-data, we must also develop new approaches that utilize trust relationships and legal sanctions that enhance access instead of denying it. The anonymity and anarchy of the Web do not lend themselves to these kinds of approaches so that new forms of Cyberinfrastructure must be developed specifically for researchers that encourage responsibility and that sanction irresponsibility. In the end, short of making all data inaccessible, the best protections for data are institutions that enhance trust, responsibility, and confidentiality.

Finally, SBE Cyberinfrastructure is only as valuable as the information it contains, coordinates, and analyzes, whether this information is collected by cutting edge digital methods or face-to-face interviews. The path to cutting edge cyber-capabilities and the ability to gather and utilize data is being hampered not just by privacy, confidentiality, and other policy and technical concerns, but also by fiscal concerns about the sheer expense of gathering relevant data. Any program for building better Cyberinfrastructure for the SBE sciences should recognize the costs of information gathering, ingestion, and data curation, and more broadly it should budget for the true costs of infrastructure development, provision, and use.

#### ***Workshop Recommendations for Data Storage and Archiving:***

There is an opportunity to make storage a commodity in the sense that anyone can purchase (cheaply) as much as they want or need. But there is a need to think long term in developing the institutional structures to make data preservation possible. Cyberinfrastructure for archiving should provide:

- *Collecting and Archiving* – There should be seamless, transparent, networked, and automated backup and archiving from PDA, laptop, cell-phone, supercomputer, and other data sources. There should be robust, automated conversion process from one format to another such as from analog to digital or from one statistical analysis package to another. There should be efforts to collect ephemeral or endangered data such as perishable video tapes or recordings of dying languages.

- *Preserving* – An institutional and technological framework should be developed jointly by computer scientists and social and behavioral scientists that will collect, preserve, and protect data.
  - There should be a thousand-year plan for transferring and preserving important data across generations of technology.
  - Existing standard methods for recording and preserving data should be regularly utilized by researchers and extended to new forms of data.
  - Institutions must be developed (most likely located at universities) that will, like libraries, collect, preserve, and disseminate data. Funding and organizational continuity are essential for these institutions.
- *Safeguarding* – Technological and institutional innovations must go hand-in-hand to protect the privacy of data providers.
  - There must be more exploration of technological solutions such as encryption, artificial databases, and automated disclosure. Privacy protection methodologies emanating from the machine learning and statistical communities should also be developed further.
  - But there must also be a commitment to developing new institutional mechanisms that enhance trust, responsibility, and confidentiality. The goal should be to maximize access to data for researchers by creating mechanisms that rely upon trust, responsibility, and ultimately sanctions for bad behavior instead of simply restricting access to data.
  - Costly methods such as Census Research Data Centers which protect data by rigidly controlling access must be funded adequately to make it possible for researchers to use them.

## F. Communication and Collaboration

Cyberinfrastructure allows for new kinds of *interactions* which cut across many different activities and scales. Research at the frontiers of the social and behavioral sciences now often requires global collaborations, terabyte data infrastructures, grid computing and software tools for automating research and enabling collaboration. Cyberinfrastructure enables a multidisciplinary group of researchers to address problems that far exceed the measurement capabilities, computing power and data storage capabilities of individual labs.

Scientists have much to gain and much to learn from emerging practices in cyberculture, which provide models for a wide variety of new research and dissemination strategies and methods. Computer games, such as the extremely popular *The Sims* provide many opportunities for situated learning and embodied interaction with simulations of complex

systems. New models for distributed collaboration can also be found in the modern domain of massively multi-player online role-playing games. We cannot fully leverage these trends in network culture until we understand them better.

In the near term, Cyberinfrastructure is changing search strategies for scientific information, scientific publishing, and scientific collaboration. Journal databases such as JSTOR (<http://www.jstor.org/>) have searchable archives of hundreds of scientific journals which currently sustain four million searches per month. LexisNexis® provides searchable databases of legal documents, newspapers, public records and business information. The OYEZ project is a multimedia relational database of Supreme Court arguments accessible via the Web (<http://www.oyez.org/oyez/frontpage>) and History and Politics Out Loud (<http://www.hpol.org/>) is a searchable archive of politically important audio materials. The TalkBank program (<http://talkbank.org/>) is fostering fundamental research in the study of human and animal communication by constructing sample databases and tools for analyzing these data. TalkBank allows researchers to share and comment upon primary materials via networked computers.

Almost all journals now offer on-line editions, and some of them are integrating shared data and other links into these editions. Thus a published paper can now include links to code that will generate its statistical analyses, tables and graphs from the original data, or run the models that it proposes.

#### ***Workshop Recommendations for Collaboration and Communication:***

- *Support for Innovative Social Science Cyberinfrastructure Collaborations* – Innovative social science collaborations such as the OYEZ project, TalkBank, VoteWorld, and others should be provided with start-up resources and ongoing support.
- *Meta-Social and Behavioral Science Portal* – Although search engines provide one way to find data and projects on the Web, meta-portals could provide additional services by registering specific types of projects and requiring certain levels of documentation and availability. Meta-portals should be used as a way to bring research projects and communities together.

#### **Moonshot Challenge: Better Communication through Cyberinfrastructure**

Today's Cyber-communities provide an unprecedented opportunity for "leveling the playing field" among participants and minimizing socio-economic, technical, international and linguistic differences. Cyberinfrastructure can provide a fundamental enabler for facilitating person-to-person communication as well. In today's technology-enabled world, it is possible to gather visual, spoken, haptic, facial, gestural and physiological source data, as well as information about emotional state, movement, and discourse structure to extract meaning from and enhance communication and human interaction.

Integrating multiple, diverse and multi-modal sources of information while applying appropriate confidentiality and privacy safeguards represents a Moonshot for Cyberinfrastructure. New modes of instrumentation and analysis, new theoretical approaches, and the development of appropriate frameworks for enabling communication, collaboration, and Cyberinfrastructure-mediated interaction present an immense challenge. They also promise great benefits to both science and society by enabling next-generation communication and collaboration paradigms which can bring people together in new ways and facilitate new forms of communication and collaboration.



## G. Challenges and Opportunities

The development of Cyberinfrastructure tools for the social and behavioral sciences present major challenges. There must be incentives to produce tools and standards for assessing their utility. While this report describes what must be developed, the social, behavioral, and computer science communities face substantial problems developing tools that may not be commercially viable. Yet researchers need (at least) the level of support and robustness normally associated with commercial tools if they are not to spend all of their time chasing bugs and developing kludge solutions. We have described how computer and social scientists can collaborate to make better tools. But a culture shift is also required in government support to tool *builders* and *maintainers*. It is important that tool providers have stable support and be rewarded for providing a public good, not just for spectacular demos and/or publications out of research “professor-ware”. It is also important for universities to understand the value of faculty members pursuing the development of new tools. Finally, all Cyberinfrastructure tools must be designed for human users with usability metrics as a basic part of their design and assessment.

### *Workshop Recommendations for Sustainable Tools Efforts:*

- *Valuing Tool Development* – Social and behavioral sciences communities, including tenure committees, should value domain-relevant tools design and development as scholarly contributions to the same degree as books and papers. Interdisciplinary collaborations should be fostered and rewarded; it should be as good for a tenure case to be an author or co-author of a very good tool as it is to be an author or co-author on a very good research paper.
- *Supporting Tools Development* – Funding models for tools efforts should evolve from the one to three year funding model currently in place to a five to 10 year model. To ensure proper review and progress, government agencies should own the sponsored tool project. If the PIs are not making progress they should be removed but the tools effort itself should be sustained. Funding agencies and universities should set-aside resources for maintaining Cyberinfrastructure as well as creating it. These funding agencies should also help develop standards for the usability of these tools.

# SBE Helping CISE Design Cyberinfrastructure III

## A. What the Social and Behavioral Sciences Can Do for Cyberinfrastructure

Cyberinfrastructure is much more than just new technology – it also consists of human interfaces, organizations and institutions, and resource allocation and incentive systems. Cyberinfrastructure also offers opportunities to those who want to engage in anti-social and illegal activities. Unless it is designed correctly, it can become a playground for malevolence.

The social and behavioral sciences can help in developing interfaces, institutions, and resource allocation systems for Cyberinfrastructure. Psychology, linguistics, cognitive science, and anthropology can help to develop useful interfaces for Cyberinfrastructure that take into account human perception, cognition, language, and values. Political science, sociology, geography, organizational studies, and law can help to design workable organizations and institutions in conjunction with economics and the decision sciences which can propose innovative resource allocation and incentive systems. Finally, the social sciences have a wealth of knowledge about malevolent behavior that can be brought to bear in developing ways to deter and control it.

## B. Better Interfaces

Web browsers have revolutionized the way we interface with our computers and with others, but new intelligent user interfaces will go much further by reading our gestures, gauging our emotions, tracking our workload, and understanding natural language. Virtual environments will allow us to immerse ourselves in distant or artificial environments, to sense, touch, see, and feel the environment, and even to interact with it to change it. The social and behavioral sciences can help to develop these interfaces in ways that will be most useful to human beings. The SBE community can help us understand the “factory-day” of a scientist so that we can better learn what is needed and what works. SBE research can study the ways scientists’ multi-task, the ways they split their attention, and the ways that they get things done. Perhaps the most important contribution of linguists and psychologists will be to develop ways that computers can understand human written and spoken language and ways to search through large bodies of written, spoken, or video-taped information and to code it in useful ways. This, coupled with other advances may suggest breakthroughs in computer design.

### *Workshop Recommendations for Better Human-Computer Interfaces:*

- *Studying How Humans Encounter Cyberinfrastructure* – Social and behavioral scientists should work with Cyberinfrastructure developers to understand how humans interact with Cyberinfrastructure:



- Research should be done on the “factory day” of working scientists to understand how they multi-task, how they split their attention, and how they get things done.
- Working together, SBE and CISE scientists should embed data collection in Cyberinfrastructure (with proper privacy protections) to measure and report how humans encounter Cyberinfrastructure. Measures of human satisfaction, usability, utility, and productivity should be formulated and automatically gathered for continuous feedback and improved Cyberinfrastructure design.
- Working together, SBE and CISE scientists should study how people respond at the machine-user interface to the things that they experience on the machine. Among other things, we should learn more about how people think about their interactions and the circumstances under which they trust the communications, documents, and other information they get on the Web.
- *Designing Better Interfaces* – Social and behavioral scientists should continue to lead the way to user interfaces that are intuitive for humans to use and that communicate the function of cyber-systems. Multi-disciplinary teams of SBE scientists (e.g. experts on human computer interface), computer scientists, and others should work together to design quality user interfaces that are intuitive for humans to use, easy to install, promote through design and usability analysis a reduction in the potential for human error, and can capture field and other data accurately.

### C. Better Organizations and Institutions

For the last few decades, the conduct of scientific and engineering research has been undergoing an organizational shift. Many of the most exciting scientific discoveries today are the result of multi-disciplinary, team-oriented, science-driven and technology-enabled collaborations. Cyberinfrastructure with its need for integration, end-to-end performance, and coordination of its many components and constituencies, will require a functional organizational framework, including the development of appropriate management, oversight, incentive, decision-making, and other key organizational structures, to be successful.

Organizational analysis and research are critical to the successful development of Cyberinfrastructure as infrastructure. Cyberinfrastructure does not – and cannot – exist without organizational, administrative, management and governance infrastructure. These are necessary to ensure cooperation and coordination when technologies fail to mesh, to allocate resources equitably within shared systems, and to manage and resolve inevitable and potentially damaging disputes and conflicts. Although some elements of Cyberinfrastructure will be self-organizing, many critical elements will require more formal organization. Social and behavioral scientists can describe, explain, and help design the critical governance and management structures that are an essential part of Cyberinfrastructure.

These organizational designs, structures, and processes will have to be continually updated as Cyberinfrastructure is developed and used. Organizations are undergoing fundamental change as Cyberinfrastructure enables radically new types of information collection,

storage, communication and analysis. It is imperative to understand, design and manage new organizational forms and arrangements resulting from Cyberinfrastructure.

In addition to taking on key roles as designers of scientific Cyberinfrastructure, organizational researchers will be important in spreading and implementing new Cyberinfrastructure throughout all kinds of organizations. Today, information and computational tools and technologies facilitate all aspects of the modern organization. Cyberinfrastructure has the potential to provide a new level of organizational capability for enterprise systems, workflow control, supply chain and provisioning, archiving and records management, and other organizational components. Cyberinfrastructure provides immense potential for facilitating more capable and more successful organizations of all types.

Yet implementing Cyberinfrastructure is not easy. The high failure rate of large information technology (IT) projects and their tendency to overrun budgets indicate that we must pay attention to implementation. Strong empirical evidence demonstrates that the sources of failures and cost overruns typically are organizational and managerial rather than technological. The higher level of complexity of Cyberinfrastructure demands a commensurably higher level of organizational and managerial knowledge and expertise.

Applying organizational research and practice to Cyberinfrastructure presents great opportunities and challenges. Current organizational research can certainly help achieve the vision of Cyberinfrastructure compellingly described in the Atkins Report. At the same time, Cyberinfrastructure presents new challenges for organizational research because it will create new kinds of organizations, including virtual organizations. Consequently, there is also an opportunity to develop a new body of theory and research on organizations.

The organizational fields affected are found at all levels: At the micro level, we can learn more about individual and social psychology in organizations, including motivation, job satisfaction, organizational commitment, creativity, and related phenomena. At the level of social behavior in organizations, we can learn more about small group behavior and decision making (including virtual teams, knowledge sharing among groups and teams, communication and coordination patterns). At the level of the individual organizations, we can learn about organizational design, leadership, command and control, innovation, adaptation and flexibility, and strategic planning. Finally, at the level of inter-organizational

#### **Moonshot Challenge: Cyberinfrastructure to Promote U.S. Leadership and Competitiveness**

The erosion of U.S. competitiveness and leadership in global science, engineering, and technology is emerging as a crisis for the United States. The U.S. is losing ground in a wide variety of areas – from an increasingly under-prepared student population and workforce, to diminishing resources for supporting research and education, to increasing outsourcing beyond U.S. borders.

The development of adequate Cyberinfrastructure can help foster the high levels of research and organizational creativity, innovation, flexibility, and dynamic adaptability needed to promote U.S. competitiveness and leadership. Cyberinfrastructure can provide cutting edge communications and information technologies needed to regain U.S. dominance in the academic sector, and to build the “organizations of the future” required for global competitiveness and leadership. Moonshot Challenges include the targeted application of Cyberinfrastructure to organizational and educational frameworks to provide a structure for accelerating U.S. leadership and competitiveness, as well as the development of an increasingly savvy domestic workforce of executives, managers, scientists, engineers, and leaders who can most effectively utilize the greater resources provided by Cyberinfrastructure to move the U.S. forward.

arrangements and networks of organizations, Cyberinfrastructure is a catalyst for research on network structures and behaviors and their relationship to outputs and outcomes such as productivity, efficiency, and innovation.

Cyberinfrastructure in the sciences and in the overall society provides a fertile field for organizational research using insights from economics, sociology, psychology and political science as well as newer, hybrid subfields specifically oriented toward new manufacturing designs, new forms of team-based organizations, and innovation. Research targeted to and spanning these areas is needed to extend the social, behavioral and economic sciences so that they account for changes in phenomena.

The challenges associated with organizational research and Cyberinfrastructure are numerous. Three particular challenges are noted here. First, Cyberinfrastructure is changing rapidly making research sites moving targets and demanding *dynamic research designs*. Second, there is the inevitable danger of applying theories and research questions developed for a pre-Cyberinfrastructure environment. The danger is that some theories, variables, and relationships are mis-specified when applied to newer forms of organization structured by Cyberinfrastructure, hence the need for *Cyberinfrastructure-targeted research and theories*. A third challenge lies in an overemphasis on the technological construction of Cyberinfrastructure while paying too little attention to social, behavioral and economic variables as both antecedent to and a consequence of Cyberinfrastructure, hence the need to develop *more comprehensive and holistic models for Cyberinfrastructure*.

#### ***Workshop Recommendations for Better Organizations and Institutions:***

- *Expand and Develop Research on Cyberinfrastructure and Organizations* – Research on Cyberinfrastructure enabled organizations should be encouraged and supported. This research should use methods ranging from ethnographic studies that develop “thick” description to formal modeling and simulations. The research program should employ methods from across the social and behavioral sciences. The research should address topics such as:
  - What organizational frameworks and processes are most useful for the design, implementation, and upgrading of Cyberinfrastructure? Why do Cyberinfrastructure projects succeed? Why do they fail?
  - What kinds of organizations, communities, and groups are most likely to be created, supported, and enabled by Cyberinfrastructure?
  - What kinds of problems does Cyberinfrastructure create for organizations?
  - How do we use Cyberinfrastructure to exploit cutting edge communications and information technologies to build the “organizations of the future” required for globally competitive, leadership organizations?

- *Produce Research and Design Work on Governance Structures for Cyberinfrastructure* – Social scientists should study and design governance structures for coordination and control of Cyberinfrastructure. These studies should determine:
  - What forms of governance have been used for Cyberinfrastructure?
  - What new modes and forms of governance, coordination and control are likely to become necessary as Cyberinfrastructure is extended to all areas of society?
  - How can we develop governance structures that allow minimal but effective control of Cyberinfrastructure-enabled organizations and that do so in a way that does not subvert or thwart flexibility and innovation?
  - What are the fiscal costs and social, technical, and organizational benefits of candidate governance structures for Cyberinfrastructure?

#### **D. Deterring and Controlling Malevolence**

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. Malevolence relates both to physical systems, involving hardware and software (e.g., electronic voting machines) and to sensitive information about individuals and organizations that could harm owners/users/subjects if intentionally misused. Malevolence ranges from strictly illegal (and perhaps extra-jurisdictional) to the anti-social. The latter may be harder to control in practice, because the boundary between acceptable and unacceptable is fuzzier and less well defined.

Much 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. The abuse often leads to harm to the Cyberinfrastructure, to the users of it, and to the individuals whose data is stored within it. Some malevolent behavior can be labeled as cybercrime. We have already lived through at least three generations of cybercrime:

1. Those within discrete computing systems (computer crimes within the mainframe).
2. Those occurring across networked computing systems (hacking across networks).
3. Those involving networked, distributed and increasingly automated--moving towards complete mediation-- technology (e.g., spam driven, 'phishing' into 'pharming').

The existence and spread of various types of malevolent behavior illustrates the need to protect Cyberinfrastructure and to clarify the rights and responsibilities associated with Cyberinfrastructure use. Social scientists can be especially helpful in developing an understanding of the motivations and capacities of those who might engage in malevolent behavior, in designing institutions and procedures that deter malevolent behavior and that

**Moonshot Challenge: Re-designing the Internet**

The Internet has radically changed the way we communicate, gather information, conduct business, learn, recreate, and generally conduct modern life. However the original design of the Internet, as a means of exchanging academically-oriented information and as a venue for technically sophisticated users who generally knew and trusted one another, did not foresee the great variety of today's uses and does not protect against a wide variety of opportunities for misuse.

Providing adequate support for today's myriad of uses and protecting the Internet from misuse and malevolent behavior will involve a fundamental re-design of system. Specific design challenges span technical, social, policy, and behavioral dimensions and include:

- Design tradeoffs between identity, accountability, anonymity and freedom of action
- Designing to enhance comprehensibility and usability of the system
- Choices about the recognition (or not) of national, jurisdictional and institutional boundaries in the design

This Moonshot Challenge will require research contributions from multidisciplinary teams of economists, computer scientists, social scientists, security experts, lawyers, humanists, and others to inform and directly shape a future architecture for an Internet. Such an architecture will have important implications for Cyberinfrastructure, both for science and for society. It will need to support new modes of gathering critical information, such as the structured gathering of data on preferences and needs, and enable a wide variety of current and future uses in an increasingly technology-enabled society.

produce trustworthy Cyberinfrastructure, and in assessing the success or failure of these approaches. Computer scientists and social scientists need to work together to find ways to protect the rights of individuals (including the right to privacy) while at the same time maintaining system and communication integrity. How to do this is the fundamental challenge. There is an ever-increasing threat to confidentiality that comes from access to the ever-increasing multiplicity of data sources that come with inexpensive storage options and expanding Cyberinfrastructure. Ironically there may be greater protection of statistical data that have limited capacity to harm than for personal and financial data that pose great threats to privacy. But this does not lessen the need for new approaches to data protection and access.

***Workshop Recommendations on Deterring and Controlling Malevolence:***

Many of the risks from using Cyberinfrastructure come from the combination of *vulnerabilities* such as “backdoors” in Windows or essentially zero-cost e-mail and the people or groups that *threaten* to exploit these vulnerabilities such as hackers or spammers.

Technological fixes typically attempt to eliminate vulnerabilities. Social, behavioral, and institutional fixes attempt to reduce threats by increasing the costs or reducing the benefits of such behavior. Both methods are needed to reduce malevolent activity:

- *Assess Threats* – Social and behavioral scientists should conduct research to understand the motivations and capabilities of those who might engage in malevolent behavior.
- *Assess Technological Approaches to Reducing Vulnerabilities* – Social and behavioral scientists should work closely with computer scientists to assess which vulnerabilities are most prone to malevolent exploitation and which technological solutions are most likely to work.
- *Designing Technological and Institutional Solutions Together* – Social and behavioral scientists should work with computer scientists to develop combinations of institutional and technological methods for controlling malevolent behavior. Collaborative research should propose Cyberinfrastructure architectures that enhance accountability while respecting the legal and social expectations of network participants about their privacy and their rights.

### **E. Better Resource Allocation and Incentive Systems**

Economics, decision sciences, economic sociology, and political economy provide tools to improve the allocation of Cyberinfrastructure resources and to design incentives for its optimal use. Ultimately, Cyberinfrastructure resources (machines, networks, storage devices, etc.) must be shared among and allocated to potentially competing users in a way that promotes both system and application efficiency. This need for effective sharing among self-interested participants in a distributed and federated system can be addressed using social science methods such as incentives, norms, and regulations. For example, at present, the use

#### **Moonshot Challenge: Better Economic and Political Institutions through Experimentation**

Reducing environmental pollution, allocating the telecommunications spectrum, managing energy supply and demand, and governing local communities require the development of more creative institutional forms that rely upon individual decision-making while maximizing efficiency and effectiveness. Economists, political scientists, and behavioral scientists have proposed many such innovative approaches such as markets for environmental pollution, auctions for the telecommunications spectrum, markets for energy, and voting systems which combine individual information and preferences in optimal ways.

These ideas need to be tested before they are implemented in order to avoid catastrophic failures (e.g., the poorly designed deregulation of the California energy market) and to get the best possible design. Currently, they are often tested in small-scale experimental simulations in academic laboratories using college students. The Web provides the opportunity to undertake innovative randomized experiments, such as simulated economies, polities, or social systems involving hundreds or even thousands of people, which could not be done in the laboratory (and which would suffer from their artificiality).

Creating the physical, and especially software, infrastructure to implement large scale virtual labs by federating facilities across multiple locations will reduce research costs and enable important experimental investigations not possible today. It will also support greater diversity in subject pools. The achievement of this Moonshot Challenge will weed out flawed ideas before they cost society billions of dollars and find new ways to increase efficiency, effectiveness and responsiveness.



of federated email resources by potentially anonymous users results in an almost impenetrable miasma of unwanted and unsolicited email or “spam.” The problem is that the incentive not to generate such mail (due to cost, legal repercussions, etc.) is either absent or ineffective. Without properly engineered incentives, the possibility of Cyberinfrastructure-enabled spam – or cyberspam – looms large as a potential impediment to success.

Incentive engineering will also be critical to effective “on-demand” utility computing in which additional capacity must be allocated to a critical problem with little or no warning. Forecasting of potentially violent weather conditions and coordinated disaster response are just two examples where resources must be effectively located and employed on-demand. Developing the necessary policies that enable this form of computing with the need for maintaining large amounts of unused capacity is a subject of social science research.

Economic principles must also inform the protocols and resource control strategies that make up the everyday fabric of Cyberinfrastructure. Using social science mechanisms such as incentives, norms, and regulations to control how and when resources are allocated to users, or more immediately, to their applications offers the possibility of ensuring desirable properties for Cyberinfrastructure. For example, careful auction design may be able to ensure incentive compatibility and budget balance while equilibrium theory can be applied to the problem of overall efficiency. While these areas of study are not new, the scope and dynamics of the Cyberinfrastructure setting presents significant new research challenges.

For example, social scientists can move into the largely unexplored frontier of technology-constrained mechanisms: that is, incentive-centered allocation mechanisms that respect computational limits and algorithmic compatibility.<sup>5</sup> Computer scientists can develop scheduling and assignment policies that are incentives-constrained. Both need to address problems in the design and implementation of fiscal policies to support incentive-based allocation mechanisms such as the creation and assignment of endowments, specifying media for value exchange, dynamics for entry/exit and system evolution. Some of the impacts we anticipate from this research include:

- Reduced free-riding in peer-to-peer systems (spam, PlanetLab).
- Improved disaster planning / management.
- Reduced program costs for supercomputing, genomic database, climate change modeling.

---

<sup>5</sup> For a nice (cross-disciplinary) example of the possibilities, see J. Feigenbaum, C. Papadimitriou, R. Sami and S. Shenker, “A BGP-based Mechanism for Lowest-Cost Routing”, *Distributed Computing* (forthcoming), which proposes incentive-compatible mechanisms for inter-domain Internet routing that are backward compatible with existing BGP protocols.

***Workshop Recommendations for Better Allocation and Incentive Systems:***

*Incentive-Centered Allocation Mechanisms for Cyberinfrastructure* – Research should be devoted to two goals:

- Developing problem-respecting feasible mechanisms for sharing limited facilities, especially in circumstances where there is “peak” demand or “urgent” demand.
- Designing protocols for information and communication technology systems that respect self-interested behavior by participants.

**F. Challenges and Opportunities**

Implementing Cyberinfrastructure for engineering and the sciences, and for society more generally, requires attention to human interfaces, organizations and institutions, resource allocation, and the possibility of malevolent behavior. The social and behavioral sciences can provide insights in all these areas. Moreover, Cyberinfrastructure itself provides a new and important subject of study for the behavioral and social sciences. Cyberinfrastructure creates new kinds of human interactions, new organizations, new kinds of resource allocation and incentive problems, and new forms of malevolent behavior. Cyberinfrastructure also provides a vehicle for documenting, measuring, and tracking these interactions as well. These new phenomena must be studied by social and behavioral scientists in the same way as traditional social behaviors.





# Assessing the Societal Impact of **IV** Cyberinfrastructure

## **A. How Social and Behavioral Sciences Can Assess the Impact of Cyberinfrastructure**

Computer scientists are not only developing better Cyberinfrastructure for the sciences and engineering, they are also developing better Cyberinfrastructure for society at large. Cyberinfrastructure innovations, such as the Internet, geo-coding, and Web browsers, often start as tools for scientists and become widely used by the entire society. Consequently, it is not a very big step from helping CISE design better scientific Cyberinfrastructure to assessing the impact of Cyberinfrastructure on society and offering improvements in its design. Although there are many dimensions of possible impacts, three seem especially important to us: changes in social interactions, changes in jobs and income, and new conceptions of privacy and trust in the networked world.

## **B. Changes in Social Interactions**

The emerging ubiquity of inexpensive, intelligent, networked devices and the development of the Cyberinfrastructure that supports them is a major change in society. One harbinger of this is the spread of cell phones. Cell phones are important because their cost and ease of use represent a model for how to ensure that the technology is used. In certain countries, such as Finland, Hungary and Japan, the vast majority of the population have access to such devices. In some moderate income countries such as Estonia, cell-phones have made it possible to get full connectivity without having to replace an antiquated Soviet-era phone system. In underdeveloped countries, cell phones can be the saving device for connecting between big centers and for communication to the outside world.

Why has that happened? There are a number of reasons, including mobility, affordability, and ease of use. In most cases, cell phones do not require any special training -- you simply talk. Illiterate people can use them. In poor villages, people share and rent them. In addition, the whole concept of interactional co-presence has been changed by cell phones, instant messaging, GPS, SMS, group communication and interaction. As advances in Cyberinfrastructure drive additional convergence between cell phones and video-based media, and barriers to access remain low, social interaction will continue to evolve as it is affected by these technologies. These developments provide opportunities for behavioral and social scientists to study Cyberinfrastructure-guided activities and to gauge their impacts, both positive and negative, on society and culture.

Cyberinfrastructure allows for new kinds of interactions which cut across many different activities and scales. Understanding how technologies of information and communication affect relationships and social networks is in many ways a lynchpin of future social science. *Social network analysis* has developed the theoretical and methodological apparatus to conceptualize and analyze cybercommunities as *multidimensional networks*. Consider, for example, scientific cybercommunities. The entities within these multidimensional net-

works include individuals, data sets, documents, analytic tools, and concepts. The network links between these entities include communication, collaboration, co-authorship, citation, and co-citations between these individuals. Additional links within the multidimensional network represent associations between individuals and specific data sets they generate or utilize, the documents they publish or access or credential (rate), the analytic tools they develop or deploy, the concepts they investigate, and the things they do. Social network analysis and the social and behavioral sciences more generally can help us understand how interactions are developing and changing in cyber-communities.

A wide range of concerns, from cultural change to civic engagement and international security, can be better understood by research on social networks. Use of technology can (but does not necessarily) build both weak ties and strong ties, expanding loose networks of affiliation as well as building social capital. We know also that technology can foster a range of organizational forms that affect relationships and networks. It builds possibilities for centralized, top-down communication and the social structures associated with it, and also possibilities for emergent, fluid, networking, communication and social structures.

Learning about society's implementation of Cyberinfrastructure can also help scientists implement better Cyberinfrastructure for themselves. There is a tremendous benefit to be had from taking a closer look at day-to-day practices of distributed communication. The blend of social interaction and entertainment inspires a tremendous amount of both learning and creative motivation in both children and adults. A core to understanding how to build better Cyberinfrastructure for science is developing a deeper understanding of the ways that Cyberinfrastructure is used to enhance and change human interactions in the "real world."

#### **Moonshot Challenge: Better Information for Better Policy**

Cyberinfrastructure has the potential to provide a wealth of tools and useful information for policymakers, from the ability to obtain customized reports based on up-to-the-minute data and in real time, to the ability to integrate and elicit useful information from disparate sources of information to better make qualified assessments.

Cyberinfrastructure which provides current information on changing technologies and their impact on workers and firms would be of tremendous value. For example, policymakers who are dealing with difficult trade negotiation issues in a particular industry could access current information on the changing nature of key industries in the U.S. and around the world. Report materials could include standard summary statistics but might also include videos, case studies, and expert contacts in the particular topics of relevance to the negotiations.

Moonshot Challenges for such capabilities include the development of a real-time and comprehensive information management and retrieval system to combine this information and present it in useful and relevant ways. The development and deployment of such Cyberinfrastructure can help provide a critical competitive edge to policymakers and leaders and provide a more comprehensive information foundation for key decisions, laws and policies.

### ***Workshop Recommendations for Changes in Social Interaction:***

- *Understanding Social Networks in Cyber-Communities* – Research is needed on how social networks develop and change on the Web. This research can not only help us design better Cyberinfrastructure, it can also provide insights on basic questions in social and behavioral science.
- *Understanding How Organizations Change to Leverage Cyberinfrastructure* – Research studies should identify and examine exemplars where Cyberinfrastructure has improved supply chain processes and communication and control processes. Contemporary Cyberinfrastructure evolves over time and cannot be wholly designed; yet developmental processes can be carefully and systematically observed and, where possible, measured. As a result, best practices can be emulated and problems avoided.

### **C. Changes in Jobs and Income**

Jobs and income lie at the core of every worker's interaction with the economy. The impact of Cyberinfrastructure (CI) on these areas has already been significant, and its significance will only grow in coming years. Any study of the impact of Cyberinfrastructure on jobs and income cannot consider the US economy in isolation. To a large extent, the effects on the US economy are best studied as corollaries of the effects on the global economy. Consequently, the scale and complexity of the problems to be studied are enormous. Detailed knowledge of national economies and firms and their interactions must be developed; data must be collected, integrated, and handled on a global scale; and computational tools for processing very large datasets and extracting meaning from them must be developed.

Although rapidly changing technology has driven demand for skilled workers for the last several decades, dramatic changes in the core structure of the relationships between firms and workers have been particularly evident in recent years. Boundaries of firms, the contractual relationships between firms and workers, and the locations of firms and their workers are all changing rapidly. Especially for IT-related products and services (which include much of the existing physical and virtual Cyberinfrastructure), global multinational companies have been restructuring the location and distribution of their activities in rich and complex ways. Moreover, the effects of globalization are being felt increasingly in other manufacturing and service industries, including such industries as health care, which were once thought to be immune to outsourcing. Restructuring and reallocation of activities within and between firms that increasingly cross national boundaries poses new challenges for national and international markets and policymakers.

We face significant challenges both in understanding the impact of Cyberinfrastructure on jobs and income and in contributing to the development of new Cyberinfrastructure tools that affect jobs and the workplace in positive ways. Contributions in the latter class would include tools that improve the efficiency of the global supply network; tools that improve the experiences of workers (including traditionally less employable workers) by delivering education and building more useful (not just *potentially* useful) IT tools; and tools that match supply and demand for specialized labor in ways that benefit both worker and employer.

Everything is becoming global, but not at a uniform pace. Globalization of goods-producing industries has been ongoing for the last several decades; Cyberinfrastructure tools developed in recent years have the potential to revolutionize the efficiency of the global supply chain for many such industries. Globalization of service industries is a more recent phenomenon, accelerating dramatically during the last ten years. Cyberinfrastructure lies at the heart of this development; the geographical location of service providers has become much less important in a world with universal, instantaneous and increasingly sophisticated connectivity.

Data on workers and firms underlying our understanding of the determinants of jobs and income are typically collected and analyzed within national boundaries. While integration of national datasets across countries provides some help in understanding the impact of globalization, harmonization of national statistics is not sufficient. For outsourcing and offshoring issues, we need to understand how firms in specific industries are changing the structure of their operations in the U.S. and abroad and in turn how these changes are impacting the mix of workers at locations in the U.S. and abroad.

One step that would significantly aid the data infrastructure needs in this area is a set of universal detailed geocodes of activities of workers and firms. The administrative and survey data sources available on workers and establishments should be extended to include detailed geocodes on activity. When a company outsources or moves a division or activity to some other location within the U.S., these new geocode standards will give us a way to capture the consequent changes in both activities and locations, both inside the U.S. and abroad.

Organizational changes in production activity induced by rapid technological changes from the IT and Cyberinfrastructure revolutions have had a profound impact on the demand for skills. In turn, these changes have had differential impacts on the demand and employment of workers by age, gender, race, and disability. Studies of the impact of Cyberinfrastructure on firms should be accompanied by parallel studies of the impact of Cyberinfrastructure on workers.

Convergence of private and societal interests should be encouraged. For example, given the scarcity of skilled workers, the private sector has the incentive to develop technologies that can be used readily by less skilled and otherwise disadvantaged workers, while funding priorities in the public sector could accelerate such developments that increase the employability of the less skilled and disadvantaged.

***Workshop Recommendations for Changes in Jobs and Income:***

- *Cyberinfrastructure and the Changing Workplace* – Changes in the organizational structure of workplaces (private, public, and nonprofit sectors) due to Cyberinfrastructure should be studied closely in order to identify successes and failures, both economically and in terms of the quality of the work experience. The focus should be on virtual firms, businesses facing complex data integration challenges (e.g., the medical sector), firms at the technological frontier of Cyberinfrastructure (e.g., customized products, complicated supply chains), and those at the forefront of globalization.

- *Cyberinfrastructure and the Global Distribution of Jobs and Income* – For outsourcing and offshoring issues, we need to understand how firms in specific industries are changing the structure of their operations in the U.S. and abroad and in turn how these changes are affecting the mix of workers at locations in the U.S. and abroad. One step that would significantly aid research in this area would be a set of universal, detailed geocodes of activities of workers and firms. Some key research questions are: What are the key technological developments and barriers preventing the globalization of specific types of production activities? Are there limitations in Cyberinfrastructure itself that prevent globalization? What are the limits of globalization?
- *Supporting Customization to Enable “Sell, Make, Deliver”* – Cyberinfrastructure is changing the production process of goods and services from standardization to customization. Whereas goods were once typically made before they were sold and then delivered, today it is common for goods and services to be first sold, then made or customized, and finally delivered. Customization of the production process is pushing the technological frontier of Cyberinfrastructure. Research questions include: What are the key Cyberinfrastructure tools and algorithms for supply network management that facilitate customization? How is customization affecting workers, firms, globalization, and outsourcing?
- *Cyberinfrastructure and the Workforce* – Organizational changes in production activity induced by rapid technological changes from Cyberinfrastructure have had a profound impact on the demand for skills. In turn, these changes have had differential impacts on the demand and employment of workers by age, gender, race, and disability. Studies of the impact of Cyberinfrastructure on firms should be accompanied by parallel studies of the impact of Cyberinfrastructure on workers. Education after entry into the workforce and/or job changes is a particularly important area of inquiry.

#### D. Privacy and Trust in the Networked World

The convergence of Cyberinfrastructure with nanotechnology suggests a massively networked world where even the most mundane devices will be “smart.” However, there is considerable potential for intrusion and surveillance in such a data-rich, networked world and there is legitimate, growing fear about the creation of a surveillance society. There is also the problem of “trusting” communications that are not authenticated in time-honored ways such as through personal testimony, authoritative sources, or refereed procedures. These concerns go far beyond those already discussed with respect to malevolent behavior, and they involve questions about the rights and responsibilities of people and organizations in society and in cyberspace and the possibilities for trustworthy communication over the Internet.

For example, are Internet Service Providers (ISP’s) responsible for malevolent traffic sent by their customers to other parts of the Internet? If so, does an ISP have the right to monitor all the activity that takes place on a customer’s machine? Alternatively, do individuals have the right to use their machines in private and unconstrained ways, and presumably to be held accountable/responsible for any damage that they do to others? Multi-disciplinary

research is needed to inform the design of systems of rights and responsibilities in cyberspace by appropriate social, legal, ethical, and economic principles and behavior. These rights and responsibilities need to be implemented in Cyberinfrastructure components that are affordable, usable, and maintainable. And the components must fit together in a consistent system design.

Responses to Cyberinfrastructure involve trust as well as rights and responsibilities. Cyberinfrastructure is increasingly pervasive, reaching into all aspects of our professional and personal lives. But can people and organizations trust that infrastructure, both in the technical sense and the organizational sense? Specifically,

- Can you believe what you see and hear? Are these data, text, images, and sounds, legitimate? Have they been produced by the individual or organization that claims to have produced them? These questions relate both to integrity and identity.
- Should you believe that a particular part of the infrastructure will provide the service it advertises – and only that service?
- More importantly, how can you make these trust decisions? What tools can help people determine the origin, identity, or integrity of the data or service? What interfaces are intuitive and usable by the broadest range of unsophisticated users?
- How can organizations support their users in making trust decisions? How do organizational policies enhance—or hinder—privacy, security, and trust?

By example, consider the following scenarios:

- You receive a message from an organization offering a service; is the message really from that organization? If you offer payment will they provide that service, and if not, what organizational or technological recourse do you have? If you provide personal data as part of the service, can you trust them to use the data only as agreed for that service?
- You see a photograph in a newspaper, television, or Web site; is this photograph an image of something that actually happened, or has it been modified?
- Your movements and activity are tracked throughout your workplace, your city streets, and perhaps your home; what technical and organizational systems may allow this data to be used by systems to provide benefits (to productivity, community security, or personal convenience) and yet allow you to understand and control the limits on your privacy?
- Your organization insists on the use of a specific Web service, but to make it work you must reduce the security settings in your browser; or, the service requires excessive personal information that may be later exposed or used inappropriately.



Surveillance might reduce malevolent behavior, but it might also invade privacy, reduce feelings of trust, and reduce risk-taking and innovation. Developing the proper balance in a cyber-world requires rethinking many fundamental issues. A key question is whether political and social systems will evolve at a pace to stay ahead of what technology might soon allow. SBE-funded researchers are in an excellent position to address questions about technology-driven social and cultural change and the evolution of institutions to cope with these changes.

***Workshop Recommendations for Privacy and Trust in a Networked World:***

- *Rights and Responsibilities in Cyberspace* – What are the rights and responsibilities of people and organizations in cyberspace? Multi-disciplinary research is needed to make sure that designs for systems of rights and responsibilities in cyberspace are informed by appropriate social, legal, ethical, and economic principles.
- *Trust and Cyberinfrastructure* – New research on trust should build on the existing NSF Cyber Trust program and require deep collaborative efforts by social science and technical researchers. Collaborative research teams can address key problems such as:
  - How do end users think about privacy and trust, as related to their actions in and through the Cyberinfrastructure?
  - How do these notions vary across cultures around the world?
  - How do these notions change over time?
  - How can the next-generation Cyberinfrastructure support users' real desires for privacy, and provide them tools to make reasonable trust decisions?

**E. Challenges and Opportunities**

The questions raised by Cyberinfrastructure are profound. What kinds of social networks will be created by Cyberinfrastructure? How can they be used for good? How might they be used malevolently? How does the information overload of emerging technologies relate to limitations of attention? How does this affect education, particularly in those areas that require real concentration? Are new social divides being manifested – whether economic or generational—between those who have considerable facility and feel at ease with the new technologies and those for whom Cyberinfrastructure remains largely mysterious and alien? Are certain kinds of jobs going to be eliminated, thus increasing the divide between rich and poor? Will privacy be possible in the networked world? How can people's legitimate desire for privacy be reconciled with the desire for constant connectivity and interaction?

The recent history of electronic voting machines illustrates the promises and pitfalls of Cyberinfrastructure and the need for collaboration between computer scientists and social scientists. After the highly contested presidential election of 2000, many people thought that electronic voting systems would solve the problems with existing voting



systems because of their ability to provide ballots in different languages, to accommodate those with disabilities, to prevent some forms of mistakes, to provide voters with a way to check their votes, and to count votes quickly and reliably. Computer scientists, however, soon pointed out that existing electronic systems were vulnerable to hacking and to other forms of manipulation. Social scientists pointed out in turn that the ultimate risk of using electronic systems depended upon both the systems' vulnerabilities and the potential threats that might be mounted by human actors given the social structures in place for conducting elections.

Widespread media coverage of these concerns has led to public doubts about electronic systems and to calls, in some quarters, for a halt to their use, but we still do not have an authoritative cost-benefit analysis of the use of electronic versus other voting systems that takes into account the vulnerabilities of all types of systems and the potentials for threats to them. And we do not have a clear-cut analysis of the costs and benefits of one promising solution, the provision of a paper trail, to the vulnerabilities of electronic voting systems. Surely this is an area where computer scientists and social and behavioral scientists need to collaborate. Moreover, it demonstrates the importance of these types of collaborations.<sup>6</sup>

#### **Moonshot Challenge: Cyberinfrastructure to Guard against Natural and Societal Threats**

Early warning systems have become fundamental to minimizing the human, physical, and economic damage of natural and societal disasters. One only has to go through airport security to experience how the airline industry uses early warning to promote the safety of its flights. Man-made disasters such as terrorism, disease epidemics such as the spread of SARS, and natural disasters such as earthquakes are all the target of sophisticated systems for early warning and prevention when possible.

As society becomes more complex and inter-connected, such systems are barely keeping up with both the number and severity of potential disasters and threats. Critical Cyberinfrastructure in the form of data gathering and management, tools and technologies for integrating diverse data sources and eliciting critical information from raw data, and organizational and social structures for disseminating information and addressing threats are critical to ensure the safety and stability of society, and constitute a Moonshot for today's scientific communities and technologies.

<sup>6</sup> American Association for the Advancement of Science prepared by Mark S. Frankel, "Making Each Vote Count: A Research Agenda for Electronic Voting," 2004.

# Potential Roles of SBE and CISE V

The history of electronic voting systems in the last five years shows that the success of Cyberinfrastructure will depend to a significant degree upon information technology's ability to grapple with the fundamental complexity and ambiguity of communication, social interaction, and culture. No single academic discipline or point of view is sufficient to comprehend all the implications of Cyberinfrastructure. On the technology side, the possibilities are exciting and daunting. On the human side, new challenges will arise from the unpredictable uses to which tools are put by actual communities of users. Both SBE and CISE can play key roles as research collaborators and expert consultants.

Moreover, true collaborative research is needed between SBE and CISE researchers. In order to achieve this, both intellectual and material interfaces must be shared. For example, it is not sufficient for SBE researchers to be told about Cyberinfrastructure possibilities if they do not possess the technical expertise to understand their ramifications. Many SBE researchers lack the technical know-how to participate without significant support from Cyberinfrastructure experts. Similarly, CISE researchers often lack sufficient domain-specific knowledge to appreciate the complexity of the technical problems that truly need to be solved by SBE researchers. The level of knowledge required by both sides will require true collaboration between the two research communities to make a joint research initiative successful. SBE researchers must become familiar with emerging Cyberinfrastructure and CISE researchers must learn about the social sciences.

The preceding pages have outlined the three tasks that must be addressed:

1. CISE developing Cyberinfrastructure to support and enable the SBE sciences
2. SBE helping CISE design scientific infrastructure
3. CISE and SBE working together to assess the societal impact of Cyberinfrastructure.

Table 1 shows what each group, SBE researchers and CISE researchers, can bring to a collaborative partnership for tackling these tasks.

## **A. Cyberinfrastructure for the SBE Sciences**

The first task listed on Table 1 is to develop Cyberinfrastructure to support and enable the SBE sciences. The first step in doing this is defining SBE needs for data collection, computing power, tools for data comparison and measurement, methods of data storage, and communication and collaboration. SBE researchers should take the lead in doing this, but they must also collaborate closely with CISE researchers. In turn, CISE researchers and their community can provide either "off-the-shelf" solutions for SBE problems or entirely new solutions.

Broadly, the expertise of CISE researchers is needed in such areas as data handling (instrumented collection, management, searching and mining, authenticated access and secure transmission, complex query handling, non-rectangular databases); operations research (optimization and modeling, decision support, supply network management); and human-computer interfaces. CISE researchers could help develop new research methods to federate, validate and analyze massive social datasets constructed from multiple sources; design feasible algorithms for computationally overwhelming social science methods; develop toolkits and techniques for creating controlled experimental environments in distributed virtual labs connecting large numbers of human subjects for simultaneous real time experimental interactions; and develop ways to protect people's privacy while allowing researchers to use large datasets with individual information.

**Table 1 – Roles of SBE and CISE for Each Task**

TASKS	ROLE FOR EACH GROUP	
	Social, Behavioral and Economic Sciences (SBE)	Computer and Information Science and Engineering (CISE)
<p>CISE developing Cyberinfrastructure to support and enable the SBE sciences</p>	<p>Identify SBE needs and requirements for:</p> <ul style="list-style-type: none"> <li>• Data collection</li> <li>• Computational analysis, simulation and modeling</li> <li>• Tools for data comparison and measurement</li> <li>• Data storage, management, and preservation</li> <li>• Communication and collaboration, etc.</li> </ul>	<p>Work with SBE researchers to identify needs and requirements and frame them as technology problems</p> <p>Design, develop and deploy technical solutions for SBE problems</p> <p>Work with SBE community to target Cyberinfrastructure tools to community needs, and to assist the community to use them effectively</p>
<p>SBE helping CISE design Cyberinfrastructure</p>	<p>Work with CISE community to structure Cyberinfrastructure organizations and infrastructure to promote coordination, functional social dynamics, effective decision-making, conflict resolution, etc.</p> <p>Work with CISE community to define effective incentive and allocation structures that promote stability, efficiency, and usability of Cyberinfrastructure</p> <p>Work with CISE community to better define and develop mechanisms for discouraging malevolent behavior</p>	<p>Incorporate models, frameworks, incentives, policies, and other mechanisms from the social science community into tools and technologies comprising Cyberinfrastructure</p> <p>Target enabling Cyberinfrastructure tools and technologies to social, behavioral and economic science applications</p> <p>Develop Cyberinfrastructure solutions to enable SBE:</p> <ul style="list-style-type: none"> <li>• Data collection and integration</li> <li>• Data analysis and modeling</li> <li>• Data comparison and measurement</li> <li>• Data archiving</li> <li>• Communication and collaboration</li> </ul>
<p>CISE and SBE assessing the societal impact of Cyberinfrastructure</p>	<p>Describe and assess social impacts of Cyberinfrastructure on:</p> <ul style="list-style-type: none"> <li>• Human interaction</li> <li>• Jobs and income</li> <li>• Privacy</li> <li>• Social and institutional frameworks, etc.</li> </ul>	<p>Develop coordinated instrumentation, tools and technologies for assessing the social impacts of Cyberinfrastructure</p> <p>Assess the technical vulnerabilities of Cyberinfrastructure</p>

As Cyberinfrastructure develops for the SBE sciences, social and behavioral scientists must find ways to use it fully. There must be outreach regarding the importance and role of Cyberinfrastructure in social and behavioral science research, and there must be efforts to overcome the resistance of some in the community to large-scale projects. Education is needed in the SBE community about how Cyberinfrastructure approaches are becoming a reality and about how they require resources and research teams at larger scale than previously seen.

### B. SBE Helping CISE Design Scientific Infrastructure

CISE researchers and infrastructure developers are already working with the research community to define the technical possibilities for Cyberinfrastructure in the sciences and engineering. Among the many possibilities for Cyberinfrastructure are new kinds of sensing devices, more powerful computing systems, innovative ways to store and analyze data of all sorts, protocols for storing and documenting data, and methods of encryption and security for data. SBE researchers should become involved in these discussions to help assess and design better human-computer interfaces, organizations and institutions for Cyberinfrastructure, resource allocation and incentive systems, and ways to deal with malevolent behavior. For example, the design, implementation, and management of effective distributed infrastructure are as much problems for social science engineering as they are for computer engineering. Computer scientists have become increasingly aware of the need to integrate the motives and actions of human users in the design of Cyberinfrastructure systems. After all, humans are smart nodes in the network, and the performance of the system depends crucially on them.

#### **Moonshot Challenge: Cyberinfrastructure to Promote Effectiveness and Productivity**

To increase the benefits that come from advancing Cyberinfrastructure, it is critical for technology development to take into account user needs from the very beginning of design. Poor design can lead to problems ranging from small scale injuries of repetitive motion to large scale failures of critical systems, e.g. poor control interfaces of nuclear power plants or air traffic control stations which work against natural modes or capacities of human users. System design that promotes usability, reliability, amplifies human capacities for information processing, and compensates for constraints on cognitive and motor performance can pay off immensely. From a more efficient workforce to the reduction of risk in dangerous industries and military operations, user-centric design of Cyberinfrastructure can help promote effectiveness and productivity.

This Moonshot Challenge will require multi-disciplinary contributions from the science and engineering of ecological interface design, economics, and computer information science and engineering. With increased understanding between communities, the development of effective, productive, and user-oriented Cyberinfrastructure can achieve its potential.

Because Cyberinfrastructure is shared, its use requires cooperation between autonomous organizations and individuals, who are separately motivated and who typically have conflicting interests. The design of social systems that respect incentives to solve group coordination problems when agents have separate conflicting interests is at the core of social science research. Current leading edge research focuses on the design of various approaches – including market-like, bargaining, preference-revelation mechanisms, regulations, norms, and laws– for solving group problems.

Not only can SBE scientists help CISE researchers and infrastructure providers develop better ways of implementing Cyberinfrastructure for problems that have already been identified, they can also identify emerging problems that might

be solved by implementing or improving Cyberinfrastructure. Through their studies of scientific researchers, scientific organizations, and scientific institutions, SBE scientists will be able to find areas where technical advances are a key component of problem solutions.

### **C. CISE and SBE Assessing the Societal Impact of Cyberinfrastructure**

CISE and SBE researchers should work together to determine the social impacts of Cyberinfrastructure. CISE researchers are needed to describe technical possibilities. SBE researchers are needed to assess their social impacts. And both groups must work together to develop technical and institutional solutions for problems and opportunities created by Cyberinfrastructure. We have already mentioned many examples where collaboration could be helpful:

1. Deterring and controlling malevolent behavior using Cyberinfrastructure,
2. Producing better electronic voting systems,
3. Finding ways to introduce Cyberinfrastructure into society that mitigate adverse impacts,
4. Protecting privacy while allowing access to data.

The key tasks here are developing ways to measure the impacts of Cyberinfrastructure and encouraging computer scientists and social scientists to design Cyberinfrastructure that respects both human and technical characteristics.

### **D. Interdisciplinary, Not Just Multidisciplinary Research**

There is a growing research community comprising computer scientists and social and behavioral scientists who have started to address some of the central problems on the frontiers of Cyberinfrastructure design and management. One example of the emerging community is the ACM SIGecom and its annual E-Commerce conference.<sup>7</sup> This group is focused on intersections between computer science and economic theory applied to the design of computational markets and other mechanisms to support federated, distributed electronic infrastructure (unlike other groups which focus more on for-profit business applications).<sup>8</sup>

---

<sup>7</sup> See <http://www.acm.org/sigs/sigecom/>.

<sup>8</sup> For examples of the fruitful cross-disciplinary fertilization and collaboration that has been developing, see the various Proceedings of EC'xx, as well as, e.g., *Games and Economic Behavior*, vol. 35, issues 1-2 (2001) (a leading economics journal special issue devoted to economics and artificial intelligence), or *Agent-Mediated Electronic Commerce V: Designing Mechanisms and Systems*, LNCS 3048 (Springer-Verlag, 2004) (selected refereed papers from a computer science workshop).

Another example is the digital government research community and its annual dg.o research conference.<sup>9</sup> This group is supported by the NSF CISE Digital Government Program and has undertaken research on problems including digital archiving, multi-dimensional geospatial analysis, electronic rulemaking, and digital aerial analysis.<sup>10</sup> There are also interdisciplinary research communities studying the Internet, digital libraries, privacy, intelligent computer interfaces, and many other topics. The bilateral nature of the research problems in these areas where computer scientists and social and behavioral scientists have something to offer one another suggests the possibility for real inter-disciplinary research where each side learns from the other. Indeed, we believe that meeting the challenge of Cyberinfrastructure requires this kind of research.

***Workshop Recommendations for Building Interdisciplinary Communities:***

- *Cyberinfrastructure Savvy People* – We must design and implement programs to attract, educate, and retain both SBE and CISE researchers who can advance Cyberinfrastructure research. Support for predoctoral and postdoctoral fellows is one way to attract talent to a new area. Two other proposed approaches are programs such as the “Transformational Research Experiences for Scientists” (TRES) and the Immersive Summer Workshop integrating a broad spectrum of social scientists, domain experts, technologists, and other stakeholders in Cyberinfrastructure.
- *Mobilize the CISE-SBE Cyberinfrastructure Research Community* – We should provide encouragement and support for an interdisciplinary community of practice at the interface of the SBE and CISE communities. It will be particularly important to involve users, stakeholders, Cyberinfrastructure providers, students, postdoctoral researchers, and community practitioners from the start so that the scientific community can grow in a balanced and comprehensive fashion. This community can be developed and nurtured by:
  - Holding an annual NSF “PI meeting” or “All-Hands Meeting” for PIs with Cyberinfrastructure efforts at the interface of SBE and CISE.
  - Developing standard academic vehicles such as a new professional society or interest group, a new journal, community blogs and wikis, and a new annual “Cyberinfrastructure and Society” conference
  - Holding targeted conferences and workshops with the theme of “Cyberinfrastructure for x” where x is a key societal challenge such as community response, health, safety, etc.

<sup>9</sup> See <http://www.digitalgovernment.org/>.

<sup>10</sup> For examples of emerging inter-disciplinary research based on large-scale, complex problems in governments, see the dg.o Proceedings (<http://www.digitalgovernment.org/library/library/dgo2003/>); Communications of the ACM, special issue on digital government, vol. 46, no. 1 (2003) which includes refereed papers from SBE, information and computer scientists; and Arens et al., “Cyberinfrastructure and Digital Government” white paper, June 2003 ([http://www.digitalgovernment.org/library/library/pdf/dg\\_Cyberinfrastructure.pdf](http://www.digitalgovernment.org/library/library/pdf/dg_Cyberinfrastructure.pdf)).



# Summary Recommendations VI

In this section, we focus on the broad themes from the recommendations that are listed throughout this report and we provide summary recommendations for enabling and advancing Cyberinfrastructure for the social and behavioral sciences.

***Summary Recommendation 1: Develop and deploy enabling data-oriented Cyberinfrastructure targeted to the social and behavioral sciences.***

The social and behavioral sciences are data-driven, and many new advances, innovations, and discoveries are data-dependent. Perhaps the most dominant theme among the recommendations was for a reliable, functional, usable, and extensible data Cyberinfrastructure to enable social and behavioral science research, education, and practice. The broad categories of data Cyberinfrastructure tools and technologies called for include:

- *Systems for collecting and managing data.* Social scientists use a wide variety of sensor, text, audio, video, and other forms of data, and must organize, access, annotate, index, integrate, and manage data collections in ways that facilitate new discovery.
- *Toolkits for facilitating data integration, mining, analysis, and validation.* Such toolkits must target the wealth of administrative, transactional, and other kinds of data collections commonly used by social and behavioral scientists and must facilitate information extraction.
- *Facilities for preserving data over the long-term.* Social and behavioral scientists as well as other domains have an increasing need for community data collections to be preserved over the long-term. Participants suggested a “thousand year plan” for data preservation which can weather evolution of technology and can ensure that increasingly valuable input is available for social and behavioral science research, education, and practice.
- *Development of large-scale foundational data collections.* The development, mining and understanding of foundational data collections can create breakthroughs in understanding and accelerate new discovery. Cyberinfrastructure provides the potential of developing such collections for issues and in areas that were not possible before. For example, the development of both broad and detailed data on how U.S. firms in specific industries are changing the structure of their operation in the U.S. and abroad can provide critical information. Using such data, researchers can identify areas that may be ripe or not ready for globalization, explore the impact of outsourcing, and investigate other key issues critical for U.S. leadership and the economy.



**Summary Recommendation 2: Develop and deploy targeted toolkits, and virtual and computational environments for facilitating social and behavioral science research.**

Cyberinfrastructure provides the opportunity to use technology for enabling and facilitating social and behavioral science research in a multitude of ways. In addition to managing and deriving information from critical data, computational environments, robust software tools, and virtual environments provide a critical foundation for new advances and discoveries. Among the wide spectrum of tools and environments critical to enabling the social and behavioral science community, workshop participants focused particularly on the following:

- *Adequate computational environments.* Although much attention is focused on social and behavioral scientists' needs for data-oriented Cyberinfrastructure, important lines of inquiry for the SBE sciences, such as large-scale simulations of markets, social dynamics, and other environments, require large amounts of computational cycles, data storage, and network bandwidth between computational resources and data collection and archival facilities. In addition, human infrastructure is critical for enabling social and behavioral scientists to adapt computationally intensive methods to achieve good performance on large-scale computational resources, as well as to maintain community codes and packages of use for a broad spectrum of researchers.
- *Virtual world environments.* Virtually immersive environments can be used to allow the social science experimenter to create virtual worlds in which all stimuli but the experimentally manipulated stimuli are constant or evolve in known ways. This makes it possible to study human decision-making in complex or dangerous situations (e.g., driving, flying, reacting to surprising stimuli) and is a key use of enabling technology for research, training, and practice.
- *Research and tools for better algorithms and methods.* Computing power must be complemented by powerful tools. Scalable algorithms and tools for modeling and solving very large optimization and equilibrium problems, dealing with nonlinearities, analysis of non-rectangular datasets, customization of the production process, managing the discrete and combinatorial aspects of models, etc. are key to successful analysis, modeling and simulation for many researchers at the large scale.
- *Targeted toolkits and workbenches.* Considerable progress can be made when targeted and/or customized tools are developed to facilitate new research results. Cyberinfrastructure provides the potential to move beyond local projects and develop reusable tools, coordinate tools to better and more cost-effectively enable research, and to couple software and hardware, sensor, and other resources.

**Summary Recommendation 3: Instrument and design technologies to gather and provide key data for social scientists. Conversely, utilize human and computer interaction data to instrument and design Cyberinfrastructure technologies.**

The enormous amount of information potentially available from the Internet, personal digital devices, and the wealth of technology enablers makes it possible to develop detailed and sophisticated models of human behavior, interaction, communication, and collaboration. Many social and behavioral scientists want to explore designing and “instrumenting” technologies to access this valuable data. In addition, this data can be used to better design Cyberinfrastructure tools and technologies. Targeted recommendations focus on:

- *Portals and accessible user environments.* Portals and user environments are needed to provide broad access to data (both collected and in the field), computation, and other resources for social sciences. In addition, there was considerable interest in embedding data collection within Cyberinfrastructure access mechanisms so that measures of human satisfaction, usability, utility and productivity could be gathered continuously and used for evaluation.
- *Instrumenting technology for data collection.* Better usage of the Web for collecting survey information at a large scale whose results could be validated by smaller-scale comparisons with in-person and telephone surveys would provide the opportunity to do large scale surveys in a cost-effective manner. Also of interest was the instrumentation of small-scale technologies (cell phones, PDAs, etc.) in such a way that data on usage could be collected for the social and behavioral sciences.
- *Coding, managing, and extracting meaning from video and other forms of data.* Video, audio, textual, and related kinds of data present special opportunities for the SBE sciences. Participants thought that substantial progress can now be made in developing computerized methods for extracting meaning from those forms of data by coding text, speech, gestures, emotions, and facial expressions.

**Summary Recommendation 4: Ensure that confidentiality, privacy, and other social and policy considerations are included as part of the architecture of Cyberinfrastructure.**

Social scientists can benefit from the many kinds of data on human behavior and interaction that are now available in digital form, but these data raise issues about individual privacy and the confidentiality of information. In addition, everyday interactions on the Web raise issues of privacy and responsibility. Participants suggested a number of opportunities and challenges for social and computer scientists in these areas including the development of:

- *Privacy protection methodologies.* Participants commented that technological and institutional innovations must go hand-in-hand to protect the privacy of data providers. Encryption, artificial databases, automated disclosure, and other techniques must be explored and expanded to ensure that Cyberinfrastructure is trustworthy. This means that confidentiality and privacy methodologies must be integrated across the board in the software, hardware, data, computational, human, and institutional components of Cyberinfrastructure. At the same time, the goal should be to maximize access to

data for researchers by creating mechanisms that rely upon trust, responsibility, and ultimately (if necessary) sanctions for bad behavior instead of simply restricting access to data.

- *Policies designating rights and responsibilities in cyberspace.* Cyberinfrastructure will encompass a wide variety of coordinated resources and be shared by a broad community. Key to the successful development and deployment as well as use of Cyberinfrastructure will be an understanding of the social, legal, ethical, and economic principles which govern it.

**Summary Recommendation 5: Involve social and behavioral scientists in the design of organizational frameworks, incentive structures, collaborative environments, decision-making protocols, and other social aspects of Cyberinfrastructure.**

The component parts of Cyberinfrastructure are human, software, hardware, instrument, and other resources coordinated so as to interoperate “end-to-end” and to support multiple users simultaneously. At scale, this complex structure will need to involve appropriate user incentive structures, effective organizational frameworks, policy and privacy constraints, and a wealth of other social mechanisms to ensure stability, performance, and usefulness. The SBE perspective is fundamental to designing, developing, building, deploying, and managing successful Cyberinfrastructure. Participant recommendations include:

- *Research on governance structures for cyberinfrastructure.* Key topics include: What are useful approaches for governance, coordination and control in Cyberinfrastructure? What frameworks promote coordination and stability for distributed virtual organizations? Of critical importance is useful approaches for decision-making, conflict resolution, promoting coordination, and so forth in Cyberinfrastructure’s distributed environments.
- *Research on deterring and controlling malevolence.* The complex nature of Cyberinfrastructure is one of its key vulnerabilities. Both technological “fixes” which attempt to reduce vulnerabilities, and social, behavioral, and institutional “fixes” which attempt to reduce threats by increasing the costs/penalties (or reducing the benefits) of such behaviors are needed to reduce malevolent activity. Additional efforts must be initiated to assess threats, integrate technological and institutional solutions, and develop new approaches for dealing with a broad spectrum of potential threats.
- *Better resource allocation and incentive systems.* At scale, the need for Cyberinfrastructure resources will typically exceed the amount of available resources. Allocation mechanisms which encourage self-interested Cyberinfrastructure users to share resources in ways that benefit the broader group are critical for stability. Economic approaches and incentive-based schemes have the potential to be particularly effective for Cyberinfrastructure, especially in environments where there is often peak or urgent demand for key resources.

**Summary Recommendation 6: Develop adequate funding models for Cyberinfrastructure which enable social and behavioral science research.**

Cyberinfrastructure is a complex undertaking that involves both new discoveries through research and the development and deployment of stable infrastructure that supports research. Different metrics of success apply to research and infrastructure. Successful research must demonstrate innovation and pushes forward the frontiers of knowledge. Funding models should support the exploration and validation of research ideas within a fixed timeframe. Successful infrastructure must demonstrate stability, usability, accessibility, and reliability. Funding models should support the deployment of usable and useful research and its smooth evolution over longer timeframes. Participants at the workshop frequently mentioned the importance of appropriate funding models for both research and infrastructure. Suggestions focused on:

- *Adequate funding for long-lived data collections and facilities.* Facilities which house long-lived data collections must be able to support and provide for their collections over the long-term. Adequate models for funding long-lived data and for transitioning data between technologies and/or facilities as appropriate over time must be part of the commitment.
- *Adequate funding for maintaining and evolving tools.* Many researchers depend on community tools to enable their research. When such tools are not adequately supported or maintained, the research may suffer. Participants suggested more responsible stewardship and maintenance of the most useful research tools and called for active agency support for non-commercial tools developed and used in the academic community.

**Summary Recommendation 7: Develop explicit venues for funding inter-disciplinary SBE and CISE research on the social impacts of Cyberinfrastructure.**

The development, deployment, and use of an increasingly useful and usable Cyberinfrastructure will have immense impact on society. Workshop participants felt strongly that the social impacts of Cyberinfrastructure must be understood and used to develop more functional Cyberinfrastructure and promote positive social networks and interactions. Suggestions for key research areas included:

- *Understanding social networks in cyber-communities.* Research is needed to understand how social networks develop and change on the Web. Such work will provide critical input into the design of Cyberinfrastructure in terms of demand for physical resources, governance structures, and malevolence. Similarly, research along these lines is also likely to provide insights on basic questions in social and behavioral science.
- *Cyberinfrastructure and the changing workplace.* Cyberinfrastructure is already bringing changes to the organizational and communication structure of the workplace. The ability to study these changes closely can help identify successes and failures, both economically and in terms of the quality of the work experience. Such studies are particularly important for virtual firms, businesses facing complex data integration

challenges (e.g., the medical sector), firms at the technological frontier of Cyberinfrastructure (e.g., customized products, complicated supply chains), and those at the forefront of globalization.

**Summary Recommendation 8: Develop the community for Cyberinfrastructure and social sciences through targeted funding programs, meetings, workshops, conferences, and other activities.**

Many of the participants in the workshop had never discussed common issues. Not only had many of the SBE and CISE participants never met each other, but it was also the case that within each community, many participants had never met. The workshop facilitated many informal projects and collaborations, and expanded the horizons of many who would like to focus on multi-disciplinary projects at the interface of the social and behavioral sciences and computer science.

There was considerable desire for a “next step” to the workshop – a venue in which the issues which began to be discussed at Airlie could continue and in which the community could come together. Suggestions included:

- *Development of new researchers at the SBE-CISE interface.* Participants suggested that support for predoctoral and postdoctoral fellows is one way to attract talent to a new area. Participants suggested new programs such as one for “Transformational Research Experiences for Scientists” (TRES) and an “Immersive Summer Workshop” integrating a broad spectrum of social scientists, domain experts, technologists, and other stakeholders in Cyberinfrastructure.
- *Organize the SBE-CISE cyberinfrastructure research community.* Participants sought ways to build the SBE-CISE community in a way that involves users, stakeholders, Cyberinfrastructure providers, students, postdoctoral researchers, and community practitioners from the start so that the community can grow in a balanced and comprehensive fashion. Suggestions for interaction included holding an NSF PI meeting for PIs with Cyberinfrastructure efforts at the interface of SBE and CISE, developing a “Cyberinfrastructure and Society” conference, or holding a series of targeted workshops.



The National Science Foundation promotes and advances scientific progress in the United States by competitively awarding grants and cooperative agreements for research and education in the sciences, mathematics, and engineering.

To get the latest information about program deadlines, to download copies of NSF publications, and to access abstracts of awards, visit the NSF Web Site at:

**<http://www.nsf.gov>**

- **Location:** 4201 Wilson Blvd.  
Arlington, VA 22230
- **For General Information (NSF Information Center):** (703) 292-5111
- **TDD (for the hearing-impaired):** 1-800-281-8749
- **To Order Publications or Forms:**  
Send an e-mail to: [pubs@nsf.gov](mailto:pubs@nsf.gov)  
or telephone: (703) 292-7827
- **To Locate NSF Employees:** (703) 292-5111

**NATIONAL SCIENCE FOUNDATION**  
ARLINGTON, VA 22230

OFFICIAL BUSINESS

RETURN THIS COVER SHEET TO ROOM P35 IF YOU  
DO NOT WISH TO RECEIVE THIS MATERIAL , OR  
IF CHANGE OF ADDRESS IS NEEDED , INDICATE  
CHANGE INCLUDING ZIP CODE ON THE LABEL (DO  
NOT REMOVE LABEL).

**PRESORTED STANDARD**  
**U.S. POSTAGE PAID**  
**National Science Foundation**