Revolutionizing Science and Engineering Through Cyberinfrastructure:

Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure

January 2003

Daniel E. Atkins, Chair
University of Michigan

Kelvin K. Droegemeier
University of Oklahoma

Stuart I. Feldman
IBM

Hector Garcia-Molina
Stanford University

Michael L. Klein
University of Pennsylvania

David G. Messerschmitt
University of California at Berkeley

Paul Messina
California Institute of Technology

Jeremiah P. Ostriker
Princeton University

Margaret H. Wright
New York University

Disclaimer

This report was prepared by an officially appointed advisory panel to the National Science Foundation, however, any opinions, findings, and conclusions or recommendations expressed in this material are those of the panel and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

1.0 Introduction

Scientific and engineering research has been crucial in both the *creation* and the advanced *application* of the amazing products of the digital revolution begun some sixty years ago – a revolution that increasingly undergirds our modern world. Advances in computational technology continue to transform scientific and engineering research, practice, and allied education. Recently, multiple accelerating trends are converging and crossing thresholds in ways that show extraordinary promise for an even more profound and rapid transformation – indeed a further revolution – in how we create, disseminate, and preserve scientific and engineering knowledge. We now have the opportunity and responsibility to integrate and extend the products of the digital revolution to serve the next generation of science and engineering research and education.

Digital computation, data, information, and networks are now being used to replace and extend traditional efforts in science and engineering research, indeed to create new disciplines. The classic two approaches to scientific research, theoretical/analytical and experimental/observational, have been extended to *in silico* simulation to explore a larger number of possibilities at new levels of temporal and spatial fidelity. Advanced networking enables people, tools, and information to be linked in ways that reduce barriers of location, time, institution, and discipline. In numerous fields new distributed-knowledge environments are becoming essential, not optional, for moving to the next frontier of research. Science and engineering researchers are again at the forefront in both creating and exploiting what many are now seeing as a nascent revolution and a forerunner of new capabilities for broad adoption in our knowledge-driven society.

A vast opportunity exists for creating new research environments based upon cyberinfrastructure, but there are also real dangers of disappointing results and wasted investment for a variety of reasons including underfunding in amount and duration, lack of understanding of technological futures, excessively redundant activities between science fields or between science fields and industry, lack of appreciation of social/cultural barriers, lack of appropriate organizational structures, inadequate related educational activities, and increased technological ("not invented here") balkanizations rather than interoperability among multiple disciplines. The opportunity is enormous, but also enormously complex, and must be approached in a long-term, comprehensive way. It is imperative to begin a well-conceived and funded program to seize these opportunities and to avoid potentially increasing opportunity costs.

This report is from a Blue Ribbon Panel convened by the Assistant Director for Computer and Information Science and Engineering (CISE)[1*] of the National Science Foundation (NSF) to inventory and explore these trends and to make strategic recommendations on the nature and form of programs that NSF should take in response to them. The charge to the Panel is premised on the concept of an advanced infrastructure layer on which innovative science and engineering research and education environments can be built. The term *infrastructure* has been used since the 1920s to refer collectively to the roads, power grids, telephone systems, bridges, rail lines, and similar public works that are required for an industrial economy to function. Although good infrastructure is often taken for granted and noticed only when it stops functioning, it is among the most complex and expensive thing that society creates. The newer term *cyberinfrastructure* refers to infrastructure based upon distributed computer, information and communication technology. If *infrastructure* is required for an *industrial* economy, then we could say that *cyberinfrastructure* is required for a *knowledge* economy.

The charge to the Panel is to 1) evaluate current major investments in cyberinfrastructure, most especially the Partnerships for Advanced Computational Infrastructure (PACI)[2]; 2) recommend new areas of emphasis relevant to cyberinfrastructure; and 3) propose an implementation plan for pursuing these new areas of emphasis. The full text of the charge is included as Appendix E.

The base technologies underlying cyberinfrastructure are the integrated electro-optical components of computation, storage, and communication that continue to advance in raw capacity at exponential rates. Above the cyberinfrastructure layer are software programs, services, instruments, data, information, knowledge, and social practices applicable to specific projects, disciplines, and communities of practice. Between these two layers is the *cyberinfrastructure* layer of enabling hardware, algorithms, software, communications, institutions, and personnel. This layer should provide an effective and efficient platform for the empowerment of specific communities of researchers to innovate and eventually revolutionize what they do, how they do it, and who participates.

Although the term cyberinfrastructure is new, NSF investment in envisioning, creating, deploying, and using computational-based infrastructure is not. Previous NSF programs have created key capabilities and experience that have already done much to enable a next big step up in the power, ubiquity, and application of advanced cyberinfrastructure. They have been instrumental in creating the vision and demand for more. By *advanced* we mean both the highest-performing technology and its use in the most leading-edge research.

**Pointers to references are noted with superscripts and the citations are listed in Section 7.*

In the 1960s NSF funded some of the very first academic computing centers and in the 1970s funded early activities in computational science. Beginning in the mid 1980s the Advanced Scientific Computing (ASC) initiatives together with NSFNET provided the research community access to machines at the top of the computation pyramid. The NSFNET transitioned into the commercial Internet, and a decade later the ASC program evolved into a more comprehensive source of high-end computing and related services. Two Partnerships for Advanced Computing Infrastructure (PACI)[2] were formed: one centered at the National Center for Supercomputing Applications (NCSA)[3] at the University of Illinois, Urbana-Champaign, and the other at the San Diego Supercomputer Center (SDSC)[4] at the University of California, San Diego. Recently the NSF made awards for terascale capability facilities to the Pittsburgh Supercomputing Center (PSC)[5], and then awards for a Distributed Terascale Facility (tera*grid* capability)[6] to a project consortium including NCSA, SDSC, Argonne National Laboratory, the Center for Advanced Computing Research (CACR) at the California Institute of Technology, and the PSC.

The Terascale Initiative is providing network access to high-end computing through physically proximate clusters of commodity computation servers. The more recent Distributed Terascale Facility is continuing the exploration of new modes of computing by extending the concept of clusters to that of wide-area grids of supercomputers allocated dynamically to a common problem over both wide distance and multiple organizations.

Two other highly relevant initiatives are the NSF Middleware[7] and the Digital Library Initiatives[8]. The NSF Middleware Initiative and Integration Testbed is an ongoing effort to develop, disseminate, and evaluate software that allows scientists and educators easily to build and share new distributed applications, share instrumentation, and share access to common data repositories. The Digital Library Initiative has been a major catalyst in creating the vast information sources and new services of the Internet including Google. Likewise, basic research in computer and information science over many years has produced much of what we now know as the Internet and the Web.

The NSF CISE Directorate supported most of the initiatives cited above. But also emerging across all NSF directorates are a variety of multidisciplinary research communities, working in partnership with computer and information scientists and engineers, to explore how to revolutionize both what problems they explore, as well as how they go about exploring them. Generic names for such cyberinfrastructure-enabled environments include *collaboratory, co-laboratory, grid community/network, virtual science community,* and *e-science community.* Examples of specific science-driven pilot projects include the Network for Earthquake Engineering Simulation (NEES)[9], the National Virtual Observatory (NVO)[10], the National Ecological Observatory Network (NEON)[11], the National Science Digital Library

(NSDL)[12], the Grid Physics Network (GriPhyN)[13], and the Space Physics and Aeronomy Research Collaboratory (SPARC)[14]. Taken together with the CISE-based activities, these new projects are *building out* in terms of broader scientific application, and they are *building up* in terms of function and performance. They provide a glimpse into an exciting future.

Mission-oriented research agencies are also initiating similar projects, for example the NIH Biomedical Informatics Research Network (BIRN)[15], the Department of Energy (DOE) National Collaboratories Program[16], and the DoE project for Scientific Discovery Through Advanced Computing (SciDAC)[17]. Relevant international programs include the UK E-science program[18], parts of the EU 6th Framework Project[19], and the Japanese Earth Simulator Center[20].

As indicated by the title of this report, the scope of our exploration and recommendations goes well beyond the topic of cyberinfrastructure in isolation or as an end in itself. Building, operating, and using advanced cyberinfrastructure must be done in a systemic context that exploits mutual self-interest and synergy among computer and information, and social science research communities who see it as an *object of research*, and other ("domain science") research communities who see it as a platform in *service of research*. More specifically, we need highly coordinated, large, and long-term investment in

1. *fundamental research* to advance cyberinfrastructure;
2. *development activities* to create and evolve the building blocks of advanced operational cyberinfrastructure;
3. institutions with people and facilities to provide *operational support and services*; and
4. *high-impact applications* of advanced cyberinfrastructure in all areas of science and engineering research and allied education.

We envision the creation of thousands of overlapping field and project specific collaboratories or grid communities, customized at the application layer but extensively sharing common cyberinfrastructure. The cyberinfrastructure should include grids of computational centers, some with computing power second to none; comprehensive libraries of digital objects including programs and literature; multidisciplinary, well-curated federated collections of scientific data; thousands of online instruments and vast sensor arrays; convenient software toolkits for resource discovery, modeling, and interactive visualization; and the ability to collaborate with physically distributed teams of people using all of these capabilities. This vision requires enduring institutions with highly competent professionals to create and procure robust software, leading-edge hardware, specialized instruments, knowledge management facilities, and appropriate training.

Furthermore, cutting across all these coordinated endeavors we need specific activities to benefit education, general science awareness, and policymaking. We need coordinated participation by academia, private industry, non-NSF government agencies and laboratories, and state, regional, and national centers. A program in this area should be interagency and international. It must address very complex interaction between scientific, technological, and sociological challenges and opportunities.

The Panel's findings and recommendations have been informed by extensive interaction with broad areas of the scientific and engineering research communities through 62 presentations at invitational public testimony sessions (see Appendix D); 700 responses to a community-wide survey (see Appendix B); review of dozens of prior relevant reports; scores of unsolicited emails and phone calls; 250 pages of written critique from 60 reviewers of an early draft of this report; panel members attending conferences and workshops concerning visions and needs of specific research communities; and hundreds of hours of deliberation and discussion among Panel members. The members of the Panel have backgrounds in areas widely relevant to creating, managing, and using advanced cyberinfrastructure. They include high-performance computing, visualization, technology trends, digital libraries, databases, distributed systems, middleware, and collaboration technology. Members of the Panel also have considerable collective experience in industrial management and academic administration.

In the next section of this report we present our vision for an Advanced Cyberinfrastructure Program that we recommend be initiated immediately under the leadership of the NSF. We next summarize trends and issues that we believe are converging to motivate, justify, enable, and to some extent prescribe the Advanced Cyberinfrastructure Program we described in Section 2. In the remaining sections we discuss the principal requirements for achieving this program, primarily organizational and financial. We also discuss the role of the current major centers and projects now providing advanced cyberinfrastructure, particularly, as we were specifically asked, the PACI programs. Section 7. contains references. Supplementary material is included in five appendixes.