

# Issues for Science and Engineering Researchers in the Digital Age

Office of Special Projects  
Policy and Global Affairs

National Research Council

National Academy Press

**NATIONAL ACADEMY PRESS 2101 Constitution Avenue, N.W. Washington, DC 20418**

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Grant No. EIA-9812623 between the National Academy of Sciences and the National Science Foundation. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

International Standard Book Number 0-309-07417-7

Additional copies of this report are available from National Academy Press, 2101 Constitution Avenue, N.W., Lockbox 285, Washington, D.C. 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, <http://www.nap.edu>

Printed in the United States of America

Copyright 2001 by the National Academy of Sciences. All rights reserved.

# THE NATIONAL ACADEMIES

National Academy of Sciences  
National Academy of Engineering  
Institute of Medicine  
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. William A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. William A. Wulf are chairman and vice chairman, respectively, of the National Research Council.



## AD HOC COMMITTEE ON BEING A SCHOLAR IN THE DIGITAL AGE

**JAMES J. DUDERSTADT** (Chair), President Emeritus and  
University Professor of Science and Engineering,  
Millennium Project, University of Michigan

**WILLIAM ARMS**, Professor of Computer Science, Cornell  
University

**DANIEL E. ATKINS**, Founding Dean of the School of  
Information and Professor of Information and Computer  
Science/Engineering, University of Michigan

**MARK H. ELLISMAN**, Professor of Neuroscience and  
Bioengineering, and Director of the Center for Research in  
Biological Structure, University of California

**EDWARD A. FOX**, Director, Digital Research Laboratory, and  
Professor of Computer Science, Virginia Polytechnic  
Institute and State University

**PAUL MESSINA**, Assistant Vice-President for Scientific  
Computing, California Institute of Technology

**HELEN NISSENBAUM**, Research Associate and Lecturer at the  
University for Human Values, Princeton University

**BEN SHNEIDERMAN**, Professor of Computer Science, Head  
of Human-Computer Interaction Laboratory, and Member  
of the Institutes for Advanced Computer Studies and for  
Systems Research, University of Maryland

### *Project Staff*

**ANNE-MARIE MAZZA**, Project Director

**REBECCA BURKA**, Administrative Associate

**KEVIN ROWAN**, Project Assistant

**ERIC BROWN**, Research Associate

**DUNCAN BROWN**, Consultant Science Writer

**SUE BACHTEL**, Editor

## GUIDANCE GROUP

**VINTON G. CERF**, Senior Vice President, WorldCom,  
Ashburn, Virginia

**DAVID T. KINGSBURY**, Chiron Corporation, Emeryville,  
California

**PETER SHAMES**, Space Telescope Science Institute, The Johns  
Hopkins University, Baltimore, Maryland

**DONALD SIMBORG**, KnowMed Systems, Berkeley, California

**LEE S. SPROULL**, Stern School of Management, New York  
University, New York, New York

**DOUG VAN HOUWELING**, University Corporation for  
Advanced Internet Development, Ann Arbor, Michigan

**ROBERT A. WELLER**, Woods Hole Oceanographic Institution,  
Woods Hole, Massachusetts

**WILLIAM A. WULF**, President, National Academy of  
Engineering, Washington, D.C.

### *Project Staff*

**DEBBIE STINE**, Project Director

**ERIC BROWN**, Research Associate

## Preface

The advance of information technology presents enormous opportunities in the conduct of research. In many ways, today's electronic tools of communication and computing make possible heightened productivity and creativity. At the same time, use of these tools challenges many of the traditions of academic research, including its ethical standards, the role of collaboration among colleagues, and the relationship between student and teacher.

These challenges are sharpened by the sheer speed with which these tools have penetrated society and the academic professions. Until the mid-1990s, electronic mail and remotely accessed databases primarily were used by a few specialists whose professional interests outweighed the clumsiness and slowness of the interfaces. Today these tools are so powerful and easy to use that they are all but *de rigueur* for most academic disciplines. Tomorrow they will be universal, viewed as necessities. In the meantime, applications of information technology grow rapidly more sophisticated and automated, adding media and capabilities that open up new ways of learning, communicating, and creating knowledge. Scientists and engineers may be concerned about the broader implications of these changes, including how they may affect and possibly threaten existing norms and conventions defining good scientific research prac-

tices as well as how they may define the ethical and professional responsibilities of researchers.

Recognizing the extent of these changes, the National Research Council (NRC) formed the *Ad Hoc Committee on Being a Scholar in the Digital Age*. The Committee was asked to develop a guide for students and faculty that would “provide advice on issues, such as the following:

- Opportunities and implications of conducting research in a digital environment
- Scientific, technical, and structural barriers to making the best use of these opportunities
- Electronic age communication within the research community: from the casual exchange of gossip and information to obtaining data to the preparation of articles and the dissemination of research results
- New behavioral norms and issues associated with those new norms.

Key points that could be addressed in a student/faculty guide include:

- Development and use of communal assets such as digital libraries and databases
- Use of computation as a third research paradigm with theory and experimentation
- Changing research apparatus
- Collaborative research
- Communication of research
- “Netiquette” (from a researcher point of view)
- Ethical issues
- Career patterns

The Committee may make recommendations, but they are not the main focus of the project—rather it is to develop guidance for scholars.”

The Committee originally intended to produce a guide for scholars. After months of deliberation, however, it decided that it could develop best a document that discussed the evolving nature of scientific and engineering research in the digital age, as well as one that identified issues for students, faculty, and



administrators in this changing environment. Because of rapid changes in the then current environment, the Committee did not believe it could provide a “how-to” manual for the young scholar that would apply to all disciplines and that would remain current. Thus, this report reflects the Committee’s discussion of the issues facing those involved in scientific and engineering research as they respond to the opportunities presented by new digital technology. It outlines some of the broad developments that are in progress and sketches their impacts on the conduct of research and on research institutions.

The Committee met as a whole twice in Washington, D.C., and members also communicated frequently by e-mail and telephone as they drafted and refined the report.

The report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the NRC’s Report Review Committee. The purpose of this independent review is to provide candid and critical comments to assist the institution in making its published reports as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process. We wish to thank the following individuals for their review of this report: Jon Amastae, Center for Inter-American and Border Studies; Edward Ayers, University of Virginia; Peter Denning, George Mason University; Robert Gallager, Massachusetts Institute of Technology; Richard Kouzes, Pacific Northwest National Lab; and Bruce Schatz, University of Illinois.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations nor did they see the final draft of the report before its release. The review of this report was overseen by William Howard (appointed by the NRC’s Report Review Committee) who was responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered. Responsibility for the final content of this report rests entirely with the authoring committee and the institution.

The report is organized in four chapters:

- Chapter 1, “A Revolution in the Tools of Scientific and Engineering Research,” presents a broad view of the issues involved.
- Chapter 2, “The Research Process in a Digital World,” reviews the changes to be expected in practical terms, at the level of the individual researcher or team.
- Chapter 3, “A Digital Infrastructure to Support Tomorrow’s Research Communities,” describes the impacts of new information technology at the level of the research community and its constituent institutions.
- Chapter 4, “Changing Expectations of the Researcher,” reviews the ethical, legal, and social challenges of the new tools and their applications. Throughout its report, the committee has included URLs to additional sources of information, guidance, and debate.

The staff of the National Research Council’s Policy and Global Affairs Division assisted the Committee. Debbie Stine served as initial staff director, before being succeeded by Anne-Marie Mazza. Duncan Brown prepared much of the initial draft material. Rebecca Burka and Kevin Rowan managed the administrative side of the Committee’s work. Graduate student and intern Eric Brown gave the Committee and staff the benefit of his perspective as a young researcher. Sue Bachtel provided editorial assistance.

# Contents

1	A Revolution in the Tools of Scientific and Engineering Research	1
	Introduction, 1	
	Digital Tools, 2	
	Challenges to the Institutions of Research, 5	
	Virtual Communities of Researchers?, 7	
2	The Research Process in a Digital World	9
	Introduction, 9	
	Observation and Inquiry, 11	
	Creativity and Analysis, 12	
	Collaboration, 13	
	Experimentation and Simulation, 15	
	Data Archiving and Access, 17	
	Publishing and Dissemination: The Changing Meaning of Research Communications, 19	
	A Future of Continued Innovation, 22	
3	A Digital Infrastructure to Support Tomorrow's Research Communities	24
	Introduction, 24	
	Local (Institutional) Infrastructure, 26	
	National Infrastructure, 29	

	Support for the Global Research Community, 34	
	Software and Communications Standards, 36	
	Into Unexplored Territory, 37	
4	Changing Expectations of the Researcher	38
	Introduction, 38	
	Teamwork, 39	
	Professional Development, 40	
	Rewarding Information Technology Achievements, 41	
	Publications and Dissemination, 42	
	Patents and Copyrights, 43	
	Ensuring the Integrity of Research in the Changing Digital Environment, 45	
	The Enduring Foundations of Research, 47	
	References	48
	Appendix A	
	Committee Member Biographies	55

# 1

## A Revolution in the Tools of Scientific and Engineering Research

### INTRODUCTION

*“The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important item is the same as was used in the days of square-rigged ships. But there are signs of a change as new and powerful instrumentalities come into use.” (Vannevar Bush, 1945)*

Digital computers and networks are revolutionizing the way we conduct research, just as they are recasting the way we do business and spend our leisure time. It grows easier every day to collect, manipulate, and disseminate information. Researchers everywhere are developing new ways to use these digital systems in every phase of their work. These new tools—both hardware and software—make possible more creative and productive research, including new forms of collaboration on a global scale.

Advanced information technologies have become more than simply faster, more precise, and more automated research tools. They have opened up new ways of learning and communicating for anyone in the business of creating knowledge. No discipline in science or engineering is exempt.

These new ways of conducting research are also challenging many traditions of academic research. All research institutions, including universities, libraries, corporate and government laboratories, publishers, professional associations, and research funding agencies, will need to adapt. At the same time, utmost care must be taken to preserve the core values of research. Applications of technology must be monitored to ensure that they do not threaten researchers' expression of these values.

An equally profound challenge, however, is improving the rising flood of information in digital form and the capacity of the researchers to receive, assess, and act on this information. As the tools get better, the task of research may get harder, since the data is growing more complex and diffuse. In some fields, more information is helpful. In others—especially those involving complex questions at the frontiers of knowledge—it may be overwhelming and distracting. For many fields, what will be needed is better information, and better ways to handle it.

## DIGITAL TOOLS

Computers connected by high-speed networks to other devices throughout the world are powerful information systems. Public and private networks permit transmission of nearly instant voice, images, and other information to wide audiences around the world—and all at low cost.

And digital technology continues to advance at a nearly inconceivable rate. For the past four decades, the speed and storage capacity of computers has doubled every 18 to 24 months. Cost, size, and power consumption have shrunk at about the same rate. The data capacity of networks has increased one thousand-fold in just the past decade. Traffic on the Internet is doubling every three months.

The uses that will be made of advanced digital tools are difficult to anticipate. Some of the new applications are already having a fundamental impact on academic life. It is impossible to predict long-term effects, both positive and negative.

In education, teachers use computers to enhance conventional classes, for example, by creating Web sites for course materials, showing demonstrations, and using programmed instructional packages. The new technology has also created a new form of education. In some instances, course materials have been put online, and the number of face-to-face classes has sometimes declined. Most radically, technology permits so-

called distance education, in which students and instructors may never meet face-to-face. These situations raise one of the greatest drawbacks of the digital age—increasing depersonalization. Some can compensate for this, but others can become increasingly isolated. Institutions of learning must take great care in moving into this arena.

Libraries are also changing, as almost all new information is created in digital format and much of it is available online. The need for a student or a researcher physically to be near a library is diminishing. Specialized information previously unavailable to the general public is now available. For example, the Library of Congress is mounting its collections on the World Wide Web; the National Institutes of Health offers medical information to everybody, not only doctors; and the Cornell University Legal Information Institute Web site emphasizes service to professionals who are not lawyers. Stanford University, among others, has established a “digital library.” Many universities have licenses to such digital libraries and offer access to all their faculty and staff. The Association for Computing Machinery has placed all of its publications in a digital library.

Publishing is also changing. Posting research results on the Internet offers researchers a fast and low-cost method of disseminating their results around the world. For example, physicists post preprints of papers in the online archives at Los Alamos National Laboratory. The Association for Computing Machinery and the Institute of Electrical and Electronic Engineers, among others, offer similar services. Publishers are converting their journals to electronic format and are eliminating paper versions of some journals. Since electronic journals are not constrained by the cost of printing, researchers can augment research papers with raw data and examples.

Almost every field of research is making use of computing, and in some disciplines the methods of research are changing fundamentally:

- The Internet permits social scientists to conduct surveys more quickly, more cheaply, and in greater detail. For example, data mining is being used to detect patterns in structured and unstructured sets of data.
- Remote access to instrumentation, including the remote operation and shared control of robot-like data-harvesting or data-refinement tools, is increasingly the rule. Early examples of this form of “telescience” include network access to an

automated telescope on a mountain in Chile, to an ultrahigh-voltage electron microscope in Japan, to a synchrotron in Palo Alto, and to a supercomputer in San Diego. All these instruments are now online, and accessible from digital portals by authorized researchers throughout the world.

- Online databases let researchers exchange, analyze, and manipulate protein structures, infrared spectra, recordings of artistic performances, and a vast array of other information. (The Genome Database is at the heart of the Human Genome Initiative.)

- Computer simulation has become a standard tool of research in many fields of science and engineering. Some say it has become a “third modality,” as important as theory and experiment. This kind of simulation, never possible before, has made it routine for researchers to design new drug molecules and test their properties “in silico,” understand the complex behavior of networks of brain cells and synaptic clusters in large-scale computer simulations, watch model galaxies collide, and “feel” the forces exerted as a drug docks in a protein.

- New ways to represent information have appeared using combinations of written text, dynamic images, and sound. Instructions on how to create new sensory environments are packaged in new, dynamic modes of communication (*IEEE Spectrum*, 1996). Multimodal methods for representing data are growing rapidly, challenging and expanding users’ skills and making use of seldom used abilities of the human nervous system to accommodate broader and more varied visual, auditory, and tactile input. Likely, the most advanced video games and flight training environments, with their heightened interactivity and their density of information, foreshadow the Web-based data exploration environments of the near future.

- Desktop publishing, digital video, and Web site development are emerging as new arenas of creativity.

All of these things are happening now. Foreseeing their full effects on research is impossible. As researchers use these technologies, and as they organize their lives and careers around them, personal and professional incentives and social interactions may change.

The rapidly expanding availability of primary sources of data in digital form may be shifting the balance of research away from working with secondary sources such as scholarly publications. Researchers today struggle to extract meaning



from these masses of data, because our techniques of searching, analyzing, interpreting, and certifying information remain primitive. New automated systems, and perhaps new intermediary institutions for searching and authenticating information, will develop to provide these services, much as libraries and scholarly publications served these roles in the past.

It is clear that profound changes are taking place in the way research is conducted and disseminated. The value researchers place on open communication about methods and results and the diligent search for truth, however, appears to remain constant.

## **CHALLENGES TO THE INSTITUTIONS OF RESEARCH**

The institutions of research include organizations such as universities and laboratories, libraries, museums, publishers, and professional associations, as well as the formal and informal norms of professional behavior. Together they define the boundaries of what is possible and permissible. Those boundaries are shifting strongly, and cannot be discerned with confidence, but some generalizations are possible.

### **Universities**

The evolving applications of information technology offer opportunities for cooperative relationships among widely dispersed people who are interested in sharing new experiences and intellectual pursuits. The resulting opportunities for investigation and communication will multiply far beyond the bounds of the physical campus. Research careers in the future may be built independent of today's institutions, unless those institutions can offer attractive infrastructure and resources not available elsewhere.

### **Libraries**

Libraries will expand their functions, offering users remote access to the world's information through text, images, sound, and other forms of information not yet envisioned. Libraries are moving beyond their traditional role as selective stockpilers of paper documents, exploiting and developing new tools for identifying, retrieving, and disseminating information in all of its forms. They are certain to be integrated with the Internet

through the successful and flexible World Wide Web interface (see, for example, Fox, 1999).

One function of libraries that will remain, and perhaps increase in importance, is helping users navigate, analyze, and evaluate the integrity and reliability of information. The very wealth of information available already makes it hard to find a particular piece of data. The techniques of information retrieval will need to keep pace with the large and complex distributed databases that are increasingly housing the world's information. Libraries, as organizations, may change dramatically, but the need for information professionals to support research is unlikely to diminish. They will require increasingly technical expertise and training to play the pivotal role that researchers and institutions demand of them.

### **Electronic Publication**

Scientific and engineering journals are being supplemented, and in some cases supplanted, by digital forms of publication, both formal and informal. Most journals today have electronic counterparts on the Internet, and some have abandoned paper circulation altogether. In the past decade or so, some researchers have come to use press releases and Web sites to announce results before they are published in peer-reviewed journals. A vast literature of uncertain authority has arisen.

Electronic publication has great advantages: low cost, speed, and the potential for great depth of reference and cross-reference (i.e., through hyperlinks to source material and other documentation). But it challenges many of the customs of research. The academic professions and institutions will need to develop responses that preserve traditions of open inquiry and communication but do not stifle these important digital tools.

Electronic publication allows articles to be constantly revised to reflect new findings. But those who cite results from such a source must be able to identify just which version they are referring to. There are techniques for tracking changes over time, but they are not used uniformly today (Brown and Duiguid, 1995).

Peer-reviewed publication has always played an important role in linking research results with particular researchers or teams. Through conventions of peer review and publication, results and discoveries are credited to researchers. First publi-

cation typically leads to attribution of priority for results and discoveries. Electronic publication offers an enormous variety of publications, including peer-reviewed online journals, internal Web-based research notes, and individual Web pages, many of which are not peer reviewed. It also makes the discovery of previously published work more difficult. Thus, electronic publication may weaken existing conventions for attributing credit and priority (Nissenbaum, 1998). As with any form of "gray literature," quality control is the responsibility of the author who chooses to cite such information.

Electronic publication, because it can be evanescent, also requires new forms of archiving. If a journal goes out of business in the paper world, all of its back issues normally remain available and securely indexed in libraries. If the same thing happens in cyberspace, its contents and their cataloging information vanish forever unless systematically archived. Standards for preserving electronic data and publications are under development (See <http://www.arl.org/preserv>; Waters and Garrett, 1996, also discuss these issues.)

Profound changes are taking place in academic publishing as publishers and researchers struggle with the opportunities and challenges of electronic information, but the underlying need has not changed. Researchers will always need ways to publish their research that are timely, accurate, and available to the entire academic community.

### **VIRTUAL COMMUNITIES OF RESEARCHERS?**

Research is a process by which knowledge is created from information, disseminated to others, and preserved for future generations. These activities include hypothesis, observation and inquiry, data collection, analysis, discussion, dissemination, and reanalysis. Historically, most research has been conducted in universities, institutes, laboratories and similar organizations that provide access to books, facilities and instruments, and professional colleagues. Now, digital networks have greatly enhanced the ability for researchers to work "virtually" with colleagues around the world, regardless of their physical location. Distance education for graduate student researchers promises the same freedom from geographic constraints. While electronic communication can provide students with access to expertise beyond their immediate advisors and mentors, it cannot replace the close mentoring relationship that many

consider to be at the heart of today's graduate education. While no one expects the imminent demise of university campuses, the technology is advancing so rapidly that future researchers may organize themselves around global digital networks. Individuals will differ in the extent to which they may rely on these evolving networks. The world's universities and their laboratories and other facilities over the next few decades will be joined by new types of institutions, able to exploit the evolving information technology and sometimes competing with them. (Noam, 1995; Wulf 1995; Lenzer, 1997; O'Donnell, 1998; Duderstadt, 1999).

## 2

# The Research Process in a Digital World

### INTRODUCTION

Digital tools can make the routine aspects of nearly every research task easier, from observation to publication. Computer systems have become more powerful and are used to create models of highly complex phenomena—such as cloud cover patterns. The proliferation of data on the world’s networks is making available at a mouse-click information that has traditionally required hard, time-consuming work to gather. Digital technologies not only facilitate traditional research tasks, but also enable previously impractical analysis (i.e., being able to simultaneously retrieve information from diverse archives and to correlate it). Furthermore, it is increasingly feasible to provide network access to remote instruments such as telescopes, microscopes, and specialized manufacturing services (Roberts, 2000). Equally important, research results can be made publicly available online much sooner than by traditional means.

Through access to e-mail and desktop videoconferencing, researchers may now work more informally and quickly with colleagues. Traditional institutional boundaries (and loyalties) are blurring, as researchers more frequently engage in long-term projects with colleagues from other institutions and disciplines.

But these new digital tools—because they are so powerful and easy to use—can be misused by the unsophisticated or abused by the dishonest. They may offer new temptations for plagiarism and fraud. Researchers may find themselves overwhelmed by the massive volume of data on the networks as they seek ways to winnow sound information from nonsense. Protecting the integrity of research will require vigilance and ingenuity and probably the development of new technologies to enhance the security of data and prevent forgeries, use of false identities, and unauthorized changes to publications or data.

Just as important, researchers need better ways to select from among the proliferating sources of digital information. After all, searching for and reading an article takes time, a precious commodity. In the age of hard print, researchers had cues about what to read, such as tacit hierarchies among journals, but these tools have yet to be developed for electronic publishing.

Well-intentioned researchers may be seduced by the power of their computers, and substitute off-the-shelf software packages for their own careful analysis. (Powerful statistical software packages, for example, may be misused by people who do not understand the proper application of statistics.) Blind use of the powerful tools of digital technology, without regard to underlying assumptions, can lead to errors that are difficult to avoid and detect.

The tools of information technology, like other products of our industries, can often fall short of our needs and expectations. It is easy to be carried away with their great promise, but we need to temper our enthusiasm. In real life, we need to remember that human factors may dominate the impact and acceptance of electronic tools. Technology changes rapidly; keeping up to date can be expensive. In particular, the more complex systems for global communications and collaboration involve major investments in engineering and logistics (see, for example, National Research Council, 2000).

For purposes of discussion it is useful to divide the research process into several interactive activities: articulating a hypothesis, creativity and analysis, observation and inquiry, experimentation and simulation, data archiving and access, and publication and dissemination. For most researchers, digital computing and communications technologies are changing each of these activities. This chapter attempts to trace some of the

changes and to forecast the impacts on the lives and careers of researchers.

## OBSERVATION AND INQUIRY

“Credit must be given rather to observation than to theories, and to theories only if what they affirm agrees with the observed facts.” (Aristotle, *On the Generation of Animals*, 350 B.C.)

Discovering facts remains the foundation of research. The new tools of information technology give every researcher increasingly powerful means of acquiring data, verifying it, and manipulating it in search of patterns. Researchers can now observe from the subatomic level to the outer reaches of the universe. Huge masses of data, collected for many purposes over many decades, are accessible online to researchers everywhere.

The proliferation and use of remote-sensing devices is adding to the flood of data that is reliable, sophisticated, and less costly. For example, digital sensors on satellites, aircraft, and buoys, collect weather and other environmental data that are processed and made available for downloading, accessible to anyone with a modem.

Sifting all this data in a timely way is an enormous challenge to researchers, but it is critical in order to maximize the value of the data collections. Computer systems that are powerful enough to browse files with hundreds of billions of bytes and display visually millions of pixels may also provide powerful tools for gleaning insights from massive data collections. One such system at the Jet Propulsion Laboratory enabled environmental researchers to examine detailed radar data of the entire Amazon basin and zoom in and out of areas of particular interest; previously they could examine data only on small areas of the region, one at a time.

These developments pose a challenge for researchers, since results must be replicable. It must be possible for other researchers to repeat an experiment and confirm the results. Software programs and hardware will need to be validated and documented, as will records of provenance and other items of “metadata” (data about data, especially about documents themselves and their histories), to ensure consistency and accuracy of data that is produced in one laboratory and used by another.

## CREATIVITY AND ANALYSIS

Creativity for the researcher is the ability to approach the solution to a problem by using existing tools and information in new or unconventional ways and/or by creating new tools. Easier access to information and more powerful ways to manipulate it may speed up the creative process, but also may bias or redirect the process in ways that we cannot predict.

New tools also will make possible “brute force” exploration of possible problem solutions, and powerful techniques for composing, manipulating, and assessing alternatives. These techniques and the accompanying improvements in access to peers and mentors for consultation could make research a more sociable activity, some believe (for example, Shneiderman, 1998). At the same time, the ability to reach a colleague by electronic mail does not guarantee an instant collaboration. Many researchers find that they are swamped with requests for advice, help, and mentoring, and are simply unable to respond to requests from those they do not already know.

Information tools also may increase pressure to draw research topics more directly from observations, rather than predominantly from the curiosity of researchers. The sheer quantity of digital information about the world is sure to shape many researchers’ choices in defining their future activities. This can be a daunting task, but the potential rewards are great. Information abundance can contribute to much progress in many fields if the information is well organized, tagged with descriptive metadata, and presented with appropriate textual or graphic displays. For example, there will be an increasing need to create compact textual displays that are controllable by users so that the data can be organized chronologically, clustered by source, ordered by relevance, and identified for trustworthiness. Even richer possibilities exist for visual displays that show trends in data over time, variations by source, gaps in normal patterns, or outliers that are significant. Exploratory visual “data mining” has proven to have powerful benefits in pharmaceutical drug discovery, digital libraries, and gene data presentation. Novel strategies include parallel coordinates or starfield visualizations for multidimensional data, zoomable timelines for temporal data presentations, and tree maps or hyperbolic trees for hierarchies. Rapid progress is expected as research prototypes are made commercial and embedded in scientific research systems (Card, Mackinlay, and Shneiderman, 1999).



## COLLABORATION

For the researcher, the continuing rapid increases in computing power offers not only access to information but whole new ways of communicating with colleagues. While some researchers can remember the 100 bit-per-second modems of the 1960s, others are already using the gigabit-per-second networks provided by Internet2 and other network providers. High speed networks in the gigabit-to-terabit-per-second range and high resolution digital displays (tens of megapixels) will allow an increasingly high degree of fidelity in reproducing human interaction at a distance, from routine conversation to multisite conferencing. Telepresence and virtual environments enabled by high speed networks are likely to become important tools in supporting scientific collaboration.

Digital communications with other humans are advancing in step with new digital technology. Students and their mentors are in touch with other researchers, through e-mail, Web sites, and discussion lists. Through these means, online communities are flourishing.

### Enlarging Research Boundaries

Research communities are enlarging their boundaries as the global research community is knit together by digital networks. Synchronously, researchers working in the morning in Washington, D.C., can interact with counterparts working late in the afternoon in Europe. Writing late at night, they can exchange ideas with colleagues in Indonesia who have begun the morning. Almost any request theoretically can be addressed in less than 24 hours by members of a research team no matter where they happen to be. Geographically distributed groups can now better collaborate, enabled by ready exchange of data, programs, videoconferencing, and software version control systems that function across networks.

The widespread adoption of the English language for international research communication, evolving technologies of machine translation, and multilingual information retrieval along with new digital tools for translating and retrieving information in many languages have also reduced barriers to collaboration. Most conferences have become international, with most of the submission and reviewing and an increasing portion of the discussion and selection aided by electronic tools.

These human communications are increasingly including voice and video, along with text. As sensors and tele-operated actuators improve, researchers may use interfaces that offer touch and action-at-a-distance. With the techniques of "virtual reality," some users may soon communicate with one another through simulated environments, using "telepresence," perhaps guiding their own software representations, or "avatars," to interact in a virtual world with those of their colleagues (*IEEE Spectrum*, 1996). Collaboration is supported and mediated by information technology in many ways, and often coordinated by digital libraries.

Digital tools to support research collaboration can begin with basic tools such as e-mail, audio/video conferencing, and screen-sharing programs that are freely available. The next generation of synchronous and asynchronous tools will be tailored to the needs of specific communities such as environmental scientists using NASA remote-sensing data sets or physicists sharing complex and expensive real-time experimental equipment. Scientists could benefit from software tools to facilitate and speed scheduling, transform data to accommodate each other's needs, and even negotiate delicate questions of authorship.

Researchers will always want to meet in person. Still, as the capacity of electronic networks grows, the physical distance among collaborators may become less of a consideration. With the trend toward remote use of instruments, databases, and other resources, researchers may find that they identify less with their home institutions and academic departments (which have traditionally been responsible for supplying and maintaining the facilities and infrastructure of research), and more with their far-flung collaborators.

### **Blurring Disciplinary Boundaries**

Some people argue that the traditional boundaries among disciplines also are blurring. The global scope of researchers' information networks works against narrow specialization. Many researchers in the traditional academic environment build careers within their disciplines or subdisciplines. Will the disciplinary specialist be relevant in the future, when many new, interesting and significant problems will require a multidisciplinary approach? Some say the need for specialized knowledge will fade as researchers are able to roam through

networks that contain the knowledge of the world, effortlessly extracting whatever they want to know.

On the other hand, digital technologies can facilitate contributions to multidisciplinary projects from a number of diverse specialists. Multidisciplinary modeling and simulation is becoming more tractable though emerging techniques for integrating discipline-specific modules into an overall simulation program, for example, those that link chemistry and fluid flow modules to model the behavior of combusting materials. And the Internet—with its capacity to help members of small groups find one another and sustain themselves as communities—may increase the academic tendency toward specialization.

## EXPERIMENTATION AND SIMULATION

Digital technologies make it possible to examine the physical world at an increasingly fine scale, both in space and time. They also make it possible to simulate the world in more lifelike ways, for example, in modeling the complexity of natural systems such as climate. Researchers have direct access to increasingly sophisticated instruments, and to the archived data from those and other instruments.

### **Simulation: A “Third Modality”?**

In fact, many researchers speak of computation as the “third modality” of scientific investigation, on a par with theory and experimentation. For many years computers have been used to simulate natural phenomena, through statistical methods—such as the Monte Carlo codes used to analyze the transport of neutrons in fission chain reactions—or deterministic solutions of the equations of motion for complex systems—such as the molecules comprising a physical system or the stars contained in a galaxy. Today the extraordinary power of advanced scientific computers enables predictive simulation of complex phenomena directly from fundamental microscopic principles. Numerical simulations may tend to draw more closely together modelers, theoreticians, and experimentalists, because their tools (digital sensors, simulations, networks, and databases) are increasingly related.

Digital tools enhance our understanding of natural phenomena. They do not reduce the importance of reliable experimental and theoretical analyses.

### Remote Instrumentation

One important advantage of new digital research tools is remote access to facilities that must be shared because they are costly or are located in extreme environments such as Antarctica or outer space. These facilities include, for example, both instruments for data gathering and machines for fabricating semiconductor devices. Remote access to these devices eliminates the need for expensive and time-consuming travel. More important, they allow a theoretically infinite number of observers to gain access simultaneously to instruments and technical support and to obtain multiple views of the same phenomena. Researchers in remote facilities may also meet responsibilities at a home institution, such as teaching and supervising research assistants.

Remotely accessible facilities include the following:

- The National Science Foundation's (NSF) National Nanofabrication Facility (<http://www.nnf.cornell.edu>);
- The intermediate voltage electron microscope at the University of California at San Diego (<http://www-ncmir.ucsd.edu/CMDA/> and <http://www-ncmir.ucsd.edu/MIBC/>);
- MOSIS, a remotely accessible facility for prototyping integrated circuits;
- The synchrotron X-ray data collection facility at the Stanford Linear Accelerator (<http://ssrl.slac.stanford.edu/>);
- The NSF-funded chain of incoherent scatter radars for measuring electromagnetic phenomena in the magnetosphere;
- The planned NSF Polar Cap Observatory;
- The Telescopes in Education (TIE) program, which provides educators and students with convenient access to a growing network of professional-quality telescopes, through pioneering the remote operation of telescopes by computer;
- The virtual nuclear magnetic resonance facility at Battelle Pacific Northwest Laboratory (<http://www.emsl.pnl.gov:2080/docs/collab/virtual/EMSLVNMRF.html>).

### High-Performance Computing

"Supercomputers" were originally designed for highly efficient parallel processing (for computation-intensive calculations such as chemical interactions and fluid dynamics). In the

future, they are just as likely to be used as “superstorage” systems for handling huge amounts of digital data. Some of the largest parallel-processing supercomputers today, for example, are in “video-on-demand” servers for communities of users, such as in museums, to handle the enormous volumes of graphical data that must be transferred. Managing terabytes or even petabytes of geographic or other spatial information, in addition to datasets, text, and multimedia information, is the purpose of specially designed superstorage systems with hundreds or thousands of disks and processors. One demonstration of such a system, used for mapping the surface of the earth, is Microsoft’s Terraserver project (<http://terraserver.microsoft.com>). Another example is Knowledge System’s (<http://www.ks.com>) PetaPlex series, originally funded by the National Security Agency.

Supercomputers and superstorage systems also enable the use of advanced, discipline-specific “search engines” for data-intensive applications that also involve large computations. For example, by comparing, pixel-by-pixel, co-registered Landsat satellite images of the same area that are taken at different times, it is possible to detect ground motion due to earthquakes. This type of computation requires large memories and takes months on a workstation; on a supercomputer it can be carried out in a few hours. Automated classification of large volumes of remote-sensing data (such as information on soil types) is also possible on powerful enough systems.

### **Distributed Computing**

One area of research that the burgeoning networks have made possible is so-called distributed computing: using varied collections of computers, data collection and visualization engines, linked by communications networks, to solve large-scale computing problems. Projects such as “Distributed.Net” or the “Search for Extraterrestrial Intelligence,” which rely on loosely knit groups of volunteer computer users from all around the world, are changing the paradigm for solving complex problems by distributing processing activity over networks of thousands or even millions of computers.

### **DATA ARCHIVING AND ACCESS**

Preserving knowledge is one of the most vital and yet rapidly changing functions of the university. For centuries the

intellectual focus of every university has been its library—a collection of written works, maps, and special collections preserving the knowledge of civilization. Today, such knowledge exists in many forms—as text, graphics, sound, algorithms, and virtual reality simulations—and it exists almost literally in the ether, distributed in digital representations over worldwide networks, accessible to anyone, and certainly not the prerogative of the privileged few in academe. The library is becoming less a collector of information and more a navigator of information—a facilitator of retrieval and dissemination.

In a sense, some have observed, the library and the book are merging (see, for example, O'Donnell, 1998; Wulf, 1995). Hypertext links, embedded in tomorrow's books, will lead the reader through the maze of information seamlessly. The Web provides a model for this new form of information navigation. Some think the advanced search engines of today will pale by comparison with the software now being developed, which will enable us to use artificial intelligence techniques to collect, organize, relate, visualize and summarize information.

### **Ensuring the Quality of Information**

The most challenging function of such a digital library will be to ensure the quality of information. As anyone who has searched the Web knows, the information there is highly variable in usefulness. The development of tools for capturing the useful and rejecting the useless is an important area of research in itself. So too will be work on tools to help humans rapidly review summaries, visualizations, and other organizations of information as we move from a black-and-white view of "relevance" to a fuzzier structuring that allows flexible human control.

### **Archiving Research Data**

Another growing function of the university library may be the archiving of research data. Electronically stored information is highly perishable, unlike paper documents, which can last for centuries. In particular, research notes or supporting data for a report can be erased in the blink of an eye if they are not systematically preserved. The individual researcher is generally responsible for maintaining these records, but in the

long run a more comprehensive, consistent, and reliable method is needed.

This role may be assumed by digital libraries, which would thereby become true information utilities, ranging from the personal to the global (*Communications of the Association for Computing Machinery*, 1998). In the broad sense, digital libraries are human and technological systems and institutions, with collections and processes that have at least some digital aspects, usually at least an electronic catalog. At the upper end of the development scale are information systems that support rich hypermedia collections, helping people communicate and collaborate over space and time as they use data, information, and knowledge (Lesk, 1997).

Digital libraries must have effective preservation plans, since they are assuming the roles traditionally assigned to archives and museums, as well as libraries. By integrating the collection, organization, discovery, retrieval, reuse, publication, and dissemination roles now spread across many institutions, digital libraries will make it easier to relate data, information, and knowledge and their many representations. But disciplines need to develop data format and metadata standards to enable fusion, cutting through data in different ways.

Digital methods will allow these representations to be viewed according to their origins (for example, a state or national digital library), topic or discipline (a digital library of computing), genre (a digital library of reports), media (a digital library of music), or combinations of these and other perspectives. In any case, an effective preservation program requires planning (e.g., specification of allowable standard representations) as well as ongoing activity to accommodate changes in media and format.

### **PUBLISHING AND DISSEMINATION: THE CHANGING MEANING OF RESEARCH COMMUNICATIONS**

Publishing is a fundamental part of research. It is the means by which researchers share their work with others, expose it to critical examination, give credit to previous work, and establish their own claims. Traditional methods of publication are intertwined with the technology of printing, and the economics of distributing and storing printed volumes. Information technology opens a variety of new communication channels to researchers. These channels—many under the

direct control of the researcher—are powerful means of communication. They allow researchers to publish incrementally, to include interactive materials, and to include supporting information for which there would not be room in a traditional publication. Using the new technology, researchers can communicate more broadly and directly with colleagues and the public.

Researchers today have on their desktops the technology for sophisticated multimedia production, incorporating text, graphics, audio, and video. Many of them have established their own Web sites and other channels for communicating with colleagues and the public throughout the world. Information that is openly available on the Internet is accessible to people around the world. This is a great advance over print journals or specialist monographs, which are available only to members of organizations with expensive research libraries.

These techniques place a heavy responsibility on the researcher. He or she must be accountable not only for the quality of the work, but for the completeness and candor of the presentation. The rise of cheap and capable information technology in the past decade or so has been paralleled by a growing tendency by researchers to publish their own results, before serious peer review. In physics the practice has emerged of depositing an open access version in a preprint archive at the same time as the paper is submitted to a journal. Some researchers, however, who seek to establish their priority in achieving a certain result in this way may jump the gun, propagating statements that turn out later to be misleading or incorrect when subjected to review. Questions of research priority have grown sharper for many researchers in the past decade or so, since, in some fields, research results are potentially valuable intellectual property. (For a discussion of the changing status of research priority, see Nissenbaum, 1998.)

For the consumer of information, the early publication of research is valuable, since these communications contain the latest information. On the other hand, the information may be of uncertain validity. Professionals need to develop their own techniques for assessing such information. Members of the public may lack those techniques.

An additional concern is archiving. If results are published on the Internet, the duty of preserving all relevant information (including background data and research notes) falls on both the researcher and the webmaster.



## Electronic Journals

The peer-reviewed journal has been the touchstone of authoritative research results for more than 350 years. Today thousands of journals, printed on paper, record progress in specific disciplines and subdisciplines. Publication in an established journal implies that the report has been subjected to peer review and careful editing. Readers may therefore rely on such reports. Publication in a leading journal also can confer status on a researcher. However, the number of journals poses a serious financial challenge for many libraries. Some journals can cost libraries \$10,000 or more for annual subscriptions. Some journals have dabbled in electronic publishing and others have switched completely to that mode of distribution. Most have hesitated to abandon their traditional way of operating.

Meanwhile, researchers themselves have taken advantage of electronic information technology to introduce new kinds of informal "publication," such as press releases and personal Web sites. These channels have the virtue of openness and speed. They also make it possible to include background information such as laboratory data, additional graphs, and even sound and video, which would not be available in a journal publication. The information conveyed, once it is presented, can be tested by the standard techniques of research. But many researchers have pointed out that these informal channels, as well as electronic versions of report series, lack quality control, and often muddy the waters of research more than they clarify them. (Ginsparg <http://www.lanl.gov/blurp/pg96unesco.html>), Walker (<http://www.amsci.org/amsci/articles/98articles/Walker.html>), and Harnad (1991, 2000) (<http://www.princeton.edu/~harnad/nature.html>), among others, discuss the possibilities of electronic publication.)

The National Institutes of Health (the federal agency that sponsors most of the nation's fundamental biomedical research) has proposed a new "electronic publishing site" to be known as "PubMed Central" (Pear, 1999). This site would serve many of the purposes of a journal. It is envisioned as a "two-tiered" site, with one tier containing peer-reviewed reports much like those to be found in a printed journal, and the other (larger) tier containing information on virtually all publicly funded biomedical research. The proposal can be found at <http://www.nih.gov/welcome/director/ebiomed/ebiomed.htm>. Such a site would give researchers everywhere, regardless of their

professional or geographic positions, equal access to the latest news of progress in biomedical science. (For example, researchers at institutions whose libraries cannot afford subscriptions to all-important journals would not be handicapped.) It would thereby speed progress.

Critics of the proposal, including editors of many conventional journals, worry that it would propagate poor quality science, which, rather than speeding progress, might impede it as large numbers of researchers attempt to replicate incorrect results that would have been exposed by vigorous peer review.

Other models also are evolving. The Association for Computing Machinery (ACM) is typical of academic publishers who have converted traditional journals to digital production. The processes of selection, editing, and review are unchanged. Currently, individuals and libraries can subscribe to either online or print versions, but demand for the print versions is falling rapidly and the ACM expects to phase them out over the next few years. ACM is engaged in a "portal" project, to bring the entire computing literature into *its digital library*. In Japan, NACSIS, a government agency, is stimulating change in the journal literature by providing electronic publishing and digital library services to scores of professional societies. In the United States, the National Science Foundation (NSF) is developing the National Science, Mathematics, Engineering, and Technology Education Digital Library (NSDL) to support undergraduate learning. It will be built as a "federated enterprise" (with resources submitted by educators and researchers). The Computer Science Teaching Center (<http://www.cstc.org>) has taken a similar approach (carefully testing and reviewing submitted materials for quality and usability). In other collections, resources may be submitted, made available after brief editor review, and then receive additional tags identifying how widespread their adoption is, as well as annotations describing use cases; such collections could operate in a fashion similar to the NIH's PubMed Central.

### A FUTURE OF CONTINUED INNOVATION

No one is wise enough to predict the future of research in the digital environment. That future will be shaped by the individual efforts of researchers, students, administrators, entrepreneurs, and others working to make the most of their tools. Our research system's greatest strength is its capacity for

innovation at the individual level, and innovation by its nature is unpredictable.

One possible outcome is that, to the extent that distance and time lose importance, it will become possible to do good research anywhere, without losing access to the necessary information tools. However, the significance of this effect remains to be seen; researchers at poorer colleges and universities are likely to lag those in richer institutions in gaining access to the new digital tools.

Another possible result is a shift in emphasis of scientific research, farther from more-or-less direct observation of nature toward observation that is mediated by the available instruments, networks, and databases on the world's information networks. We may also confidently predict a growing reliance on computer simulation as an adjunct to experimentation.

The extent to which the university campus will lose its relevance to scientific research as virtual communities of scholars evolve is uncertain. Researchers and administrators have been discussing the impacts for a decade or more (see, for example, Lenzer, 1977; National Research Council, 1993, 1994, 1996; Sproull and Kiesler, 1991; Casper, 1995; Noam, 1995; Wulf, 1995; Casper et al., 1998; O'Donnell, 1998; and the Vision 2010 project supported by the Carnegie Foundation <http://www.si.umich.edu/V2010/home.html#indexmap>]). It is clear that many of the functions of the local campus (such as the traditional library and some of the delivery of "mass market" undergraduate and technical education) are being threatened by information technology.

## 3

# A Digital Infrastructure to Support Tomorrow's Research Communities

### INTRODUCTION

Research in the digital age requires a new kind of infrastructure—digital libraries and databases, access to networks, adequate communications bandwidth, supercomputers, and various support services. These resources are the basis of new kinds of worldwide research communities that are taking advantage of the richer communications possibilities offered by the global network. The elements of this infrastructure are growing cheaper and more widely accessible by the day. But they will not provide themselves. Traditional academic relationships, built around these resources, are likely to be altered by the spread of a digital research environment. New relationships, based on community of interest, are becoming the common currency of research.

The responsibility for providing, maintaining, and controlling the many elements of the research infrastructure remains to be determined. Will the authority of academic departments be eroded if researchers learn to value networks (national and international resources) and extended communities more than they value the crucial social organization of departments, schools, and universities? Will the increasing remote use of instruments reduce the power of academic institutions, as researchers organize themselves in virtual institutions centered

on major instruments and facilities? To the extent that these questions can be answered in the affirmative, researchers will experience important shifts in their fundamental professional relationships. Every researcher, in shaping a career, should be aware of this potential.

No one can answer these questions with certainty today. Change will come from the bottom up, through the efforts of individual researchers, seeking their own paths to knowledge and developing the tools and relationships to get them there. The young—because they have grown up immersed in digital media environments—will probably be the most prolific in devising unexpected new tools and techniques. This effort in total will be enormous and hard to monitor, much less predict. Much of it will take place invisibly, in dormitory rooms and computer labs. The applications of these tools will be constantly surprising. Some of them will change the world.

For institutions of higher education, overseeing this unpredictable surge of innovation will be constantly challenging. Unlike other forms of academic infrastructure, such as electric power and water, the campus information infrastructure of buildings and laboratories cannot be planned in detail, or in isolation from the needs of the end-users. The best institutions can do in this case is provide enough bandwidth and suitable points of connectivity, while leaving it to researchers, their departments, and their colleagues to configure the details of the local area network. For every university the investment required will be very large.

This large investment means that researchers at every level will have to be informed about choices so that they can guide their institutions in making technology decisions. The growing disparity in Internet access between rich and poor has been called the “digital divide” (<http://www.ntia.doc.gov/ntiahome/digitaldivide/>). Vigorous efforts are being applied to cope with the economic, policy, training, and design issues (<http://www.digitaldivide.gov/>), but persistent attention to these issues will be needed at every level (Shneiderman, 2000). The large disparity between advanced and developing nations also raises concerns about the capability of researchers to participate in international projects. Bridging this divide, nationally and internationally, will bring benefits by ensuring diverse participation in research activities and enabling previously marginalized researchers to contribute.

For researchers themselves, of course, these institutional

questions are not the dominant ones (although every researcher would be well advised to be aware of them). The important questions are those that bear directly on research and education. In pursuing their careers, young researchers will experience growing freedom to build relationships on the basis of community of interest—locally, nationally, and globally. It may even become possible to build rewarding research careers outside of the traditional academic structure. The most successful academic institutions will be those that most effectively facilitate the expanding communities of researchers, while providing the local conditions to attract the best faculty and students. Those conditions include stimulating colleagues, rich information resources, and departmental reward structures that are open enough to recognize innovation and creativity in nontraditional areas.

### LOCAL (INSTITUTIONAL) INFRASTRUCTURE

Research institutions (colleges and universities, departments, and research laboratories) are responsible for local infrastructure (such as campus networks and libraries, personal computing resources for students, faculty, and staff, and access to specialized resources such as supercomputers or virtual reality environments). Academic departments generally provide instruments and facilities for research, and control their use. Overall, information technology to support faculty, students, and staff often involves 10 percent of the annual operating budget of a university.

But beyond the financial challenge is the challenge of providing the sophisticated computing environments that many young researchers demand. Information technology has become a strategic investment for universities, critical to the vitality of their academic missions and administrative services. The great diversity of researchers' needs will require an equally diverse network infrastructure. Some humanists will need access to digital libraries and graphics processing. Some scientists and engineers will need massively parallel processing. Social scientists will need the capacity to manage massive databases, e.g., data warehouses and data mining technology. Artists, architects, and musicians will require multimedia technology. Business and financial operations will need fast data processing, robust communications, and high security. Emerg-

ing areas of research will place their own demands on the university infrastructure.

### **Expertise and Support**

Most young researchers are highly knowledgeable about computing and how to use it for their research. Universities provide basic networking and computing service, but little expertise in the use of academic applications. The extent of this support varies from institution to institution. Much of it is provided informally. Some of the formal support is automated, in the form of remotely accessible databases presenting lists of the answers to “frequently asked questions,” software setup scripts, and the like.

In most institutions, less central support is provided as time goes on, as computing expertise becomes a basic skill that every researcher has. The support that is available, to the extent that it is distributed, may be less accessible to users.

Faculty development programs will need to provide graduate students and faculty members with the research and teaching skills needed for success in a digital environment. While many young researchers are highly knowledgeable about computing and how to use it for their research, many senior faculty members lack the knowledge and skills to operate confidently in the digital environment, and find the marvelous new tools of information technology to be baffling sources of job stress (*Science*, 1999). Students are far ahead of the faculty in their understanding of digital tools. Few universities have devoted adequate attention or resources to the needs of faculty (in sharp contrast to the massive amounts spent on student needs). The Faculty Development Institute of the Virginia Polytechnic Institute and State University (<http://www.edtech.vt.edu/idi.html>) represents one serious attempt to provide those tools for faculty and staff. Others can be found at George Mason University, MIT, and Cornell University.

### **Personal Computing and Communication**

Universities and colleges are considering to what extent they should be responsible for providing faculty and students with computers and communications tools such as cellular phones and PDAs. The costs involved in these decisions are substantial. For example, some may provide their staffs with

personal computers in both home and office, but require undergraduate students to supply their own. Some may provide home computers and ISDN or cable-modem connections, wireless telephones, and other devices. Others will regard computers as commodities, much like telephones or calculators, to be provided by the individual. The dramatic declines in prices of digital systems make it increasingly feasible to lay these responsibilities on the individuals (many of whom will also prefer to exercise their own judgments in choosing the systems they need).

### **Software**

It was once the norm for researchers to write or customize their own special-purpose computer software. This practice had the advantage of giving researchers a hands-on appreciation of the logical structures and appropriate applications of the software. On the other hand, for outsiders, results obtained through such special purpose software could be hard to interpret and were generally impossible to audit.

Over the past two decades, home-coded software has been steadily replaced by off-the-shelf commercial software packages for calculation, data collection and analysis, and data mining. These general-purpose packages are comparatively well documented and reliable. But many researchers have little understanding of the limitations and constraints of the software tools they use. Many of these users do not need to understand the operation of their tools in detail. For some it is vital; fortunately powerful software tools and high-tech languages are available to make this task easier.

### **Local Area Networks**

To nurture the digitally mediated research communities of tomorrow, universities will need to provide students, faculty, and staff with robust, high-speed networks. Both Internet access to off-campus resources and “intranet” networks to link students, faculty, and staff are essential. While the processing power of computers is continuing to increase, it is often the growing bandwidth of communications that is more vital to researchers. Today, switched Ethernet at 100 million bits per second is common.



At some universities the information technology infrastructure is centrally coordinated. Other institutions have allowed it to evolve with few constraints, perhaps encouraging some units to serve as “islands of innovation,” able to look ahead, explore new technologies, and serve as pathfinders for the rest of the university.

## **NATIONAL INFRASTRUCTURE**

A national infrastructure—long-distance communications, nationally significant instruments (many accessible by digital networks), software libraries, and high-performance computer facilities—might be needed. It is appropriate for these resources to be provided at the national level, by government, consortia of universities, or industry (or combinations of these institutions). They must be further developed, maintained, and calibrated.

### **Digital Libraries and Databases**

The long-term impact of digital networks on academic libraries is impossible to predict. Libraries are in a state of sweeping change, in both technology and function. The most obvious sign is the proportion of library information that is now available in digital formats. Online catalogs and information services have almost totally replaced card catalogs and printed indexes. Over half of all scientific journals are online, with the proportion in other fields rising rapidly. Activities such as JSTOR and the American Memory project at the Library of Congress are digitizing large quantities of historic documents. Researchers in some disciplines can rely almost entirely on digital information. While only extreme enthusiasts envision a world where all library materials are online, the importance of digital information looks likely to expand indefinitely.

However, the changes are much greater than simply replacing traditional materials with digital equivalents. One major change is that a vast amount of information that is important to research is provided on the Internet, with access open to all (in open archives, research Web sites, open-access periodicals, and government repositories). In some disciplines, it is possible to teach advanced courses or carry out research using open access materials only. Open access has benefits to both the author and reader, in encouraging the flow of information. Many people

see it as the natural form for research information, but somebody has to pay for managing information and no general economic model has yet been established for providing open access to all materials. Authors have to decide whether to publish their research in an open access form.

Another major effect of networked information is that readers no longer need to visit a library building. Libraries are increasingly providing services to remote users. Often they are connecting users who never enter the library with online services provided by remote publishers.

Digital libraries are surrounded by considerable hyperbole. Advocates often imply that multimedia and virtual reality are replacing text as the medium of communication and that digital libraries in the near future will supply tools for navigating oceans of information, provide researchers immediate access to all relevant literature and allow them to search by concept rather than conventional text-based retrieval. All of these concepts have been demonstrated in research laboratories, but the real world is more prosaic. Deriving new knowledge from materials in libraries remains a painstaking, skilled task, whether the materials are digital or print.

New formats of information are emerging, albeit slowly. Networked databases are important in some fields, such as the Protein and Genome databases. Research papers are providing more of the raw data and intermediate results. Online collections, such as the Perseus Project's collections of classical and humanities materials, are rich in links that relate concepts at a much finer granularity than conventional footnotes and references. Surely, such new approaches to information will continue to grow.

Digital information provides new challenges to libraries and to researchers. Libraries are very conscious of the difficulties in long-term preservation of digital material, with its fragile media and ever changing formats. Digital information is so easy to change that materials go through countless versions and revisions; this leads to difficulties in citation. Although hard data is lacking, many observers sense an increase in academic plagiarism, made easier by digital information on the Internet.

Finally, digital libraries and electronic publishing have placed stress on the framework of copyright and patents, which facilitates the exchange of information while providing income for publishers and authors. Who should own the product of research: researchers, universities, publishers, or perhaps the

taxpayer who funds so much research? In the current state of uncertainty, every researcher needs to understand the basic concepts of copyright, both as an author and a reader. In most cases, authors can be guided in their decisions by asking themselves—and their colleagues and mentors—whether a proposed transaction, such as transfer of copyright, helps or hinders the underlying objectives of communication.

### **Wide Area Networks**

The Internet in its present form was the result of initiatives by the federal government, the technical development of the ARPANet, which began in the 1960s, and the national expansion led by the National Science Foundation in the late 1980s. Today, the federal government continues to support research in the technology of information processing and communications, but the Internet is almost entirely a commercial enterprise.

Funds for specific disciplines are provided through numerous agencies, including the National Science Foundation, the National Institutes of Health, the National Endowment for the Humanities, the Department of Defense, the Department of Energy, the Department of Education, and others. In recent years, Congress has voted considerable funds to support computing in education. Although there are continual national efforts to present these activities as a coherent national plan, in practice any coordination is informal. However, there is general agreement that the nation needs to pay attention to this need.

A multi-agency program (the National Coordination Office for Computing, Information, and Communications) has been established to coordinate federal information technology R&D. Among those activities is a proposed federal research and development initiative that would support (a) long-term research in software, interfaces, and high-end computing; (b) development and acquisition of newly powerful computer systems and associated software; and (c) research on economic, social, and work force implications of the Information Revolution (<http://www.cra.org/Policy/it2.html>).

### **Advanced Scientific Computation**

Many important scientific and engineering problems are so computation-intensive as to be beyond the capacities of today's systems. These problems can be found in many fields, includ-

ing atmospheric modeling; simulation of black holes and general relativity; chemical reaction studies; cryptography; drugs by design; ecological analysis of dynamic systems; electron transport in complex media; fusion energy studies; genomics; geophysical and astrophysical turbulence; global systems simulation and modeling; high energy and nuclear physics; microbial interactions in soil, water, and vegetation; modeling metabolic systems; prime number mathematics; protein folding; quantum chromodynamics; subsurface transport; supernova science; and theoretical biology.

The federal government traditionally has supported the advance of computing by funding national supercomputer centers to support both particular government missions and the needs of the broader scientific community. It will continue to play these roles.

### **Online Access to Important Instruments**

One significant use of the world's data networks and computational grids has been to enable remote access to a diverse array of sophisticated research instruments. Such "telescience" represents a new model for scientific collaboration and investigation and provides important challenges, advancing our thinking about new approaches to surmount limitations in several important areas. Telescience tools are now evolving to include the seamless linking of remote data acquisition with distributed computation-intensive analysis and intelligent comparisons of new data with data in repositories. A most demanding form of the telescience applications of the future will employ a type of computation-based steering, "closing the loop" of remote control of data acquisition, data refinement, and data analysis. This advance will enable the process of data acquisition to be improved, in more or less real time, by feedback from the analysis to enable intelligent or knowledge-based, remote steering of research or diagnostic instrumentation.

Automated sensors are available in a wide and growing variety (including high-resolution optical, chemical, and temperature sensors). Placed on oceanographic buoys, spacecraft, and many other platforms, these sensors output their data to computers on the global network, so that a scientist or engineer at a desktop workstation anywhere can obtain direct information about environmental or physical conditions anywhere, in nearly real time. The data thus acquired can be processed,

analyzed, and shared with colleagues. These instruments are doing the work once done by expensive oceanographic voyages, for example, but with better spatial and temporal resolution and vastly lower cost.

Another kind of remote access that the network offers is the ability to study human behavior and relationships in immensely greater detail, and in near real time. Psychologists are studying interpersonal relationships in electronic groups; sociologists and anthropologists are studying electronic group dynamics; economists are studying electronic auctions. This new tool raises ethical and methodological questions involving privacy, informed consent, protected populations, and other considerations.

Access to these new resources is controlled in the traditional manner, by the government agencies and other organizations that own them. Competition for their use may be intense. NSF's nanofabrication facility, for example, is available on a competitive basis to anyone in the country, in any discipline, for remotely specifying and building specific micro-electromechanical devices needed for research projects. It is always fully booked.

### **Preservation of Data**

A new kind of data facility will be needed to archive the unpublished data underlying research, such as laboratory data and notes. Unlike laboratory notebooks, electronic notes and data are extremely perishable; their accidental or deliberate erasure may undermine the support for important research results. Furthermore, electronic notes do not bear the personal stamp of handwritten notes; techniques for electronic "watermarking" or "fingerprinting" of notes and other data will be needed to validate their sources. Secure electronic date stamping will also be needed to verify the sequence of events (and also to settle questions of research priority). Extensive work has been done on electronic notebooks, their legality, fingerprinting, time stamping, and other issues (see, for example, <http://www.emsl.pnl.gov:2080/docs/collab/research/ENResearch.html>).

## SUPPORT FOR THE GLOBAL RESEARCH COMMUNITY

Advancing information technology will enable a globalization of research. Collaborators increasingly often will be in separate countries, often continents apart. They will increasingly use remotely sited automated instruments and facilities. Standards for communications protocols, data acquisition and data processing software, and data preservation must be applied globally, to ensure that data and results are comparable, reliable, and verifiable. These tasks extend beyond the limits of national sovereignty. They will be carried out by the members of the research community (often acting individually), and sometimes through national governments as signatories to international technical conventions.

### A New Infrastructure for Collaboration: The Collaboratory

Information technology's prime virtue as a builder of relationships is its ability to enhance teamwork. Researchers in many fields are experimenting with this aspect of the technology by building "collaboratories" (a portmanteau word, coined in the 1980s, combining the words "collaboration" and "laboratory"). In a collaboratory, the available digital tools are integrated to provide researchers access to scientific resources (such as instruments and shared databases), wherever they are located. A collaboratory, according to one of the concept's earliest promoters, is "a center without walls in which the nation's researchers can perform their research without regard to geographical location—interacting with colleagues, accessing instrumentation, sharing data and computational resources, [and] accessing information in digital libraries" (Cerf, 1993; National Research Council, 1993; *New Scientist*, 1997; Wulf, 1989). A collaboratory provides the technological basis for interaction among researchers, instruments, and data independent of distance.

The collaboratory approach, built around networks of distributed computers, is expected to change not only scholarly work, but many other activities involving human teamwork, including the arts, business, and education (Rosenberg, 1991; McDaniel et al., 1994; Edelson et al., 1996; Kouzes et al., 1996; Kilman and Forslund, 1997; Casper et al., 1998; Olson et al., 1998). Some form of collaboratory may be the appropriate infrastructure for the great universities of the future. Experi-

mental research on both the social and technical aspects of collaboratory has been sponsored by the National Science Foundation since the mid-1980's; more recently the Department of Energy (<http://www.emsl.pnl.gov:2080/docs/collab/>) and the National Institutes of Health have begun collaboratory initiatives.

The most developed of the NSF-sponsored collaboratory projects is the Space Physics and Aeronomy Research Collaboratory (SPARC) ([www.si.umich.edu/SPARC](http://www.si.umich.edu/SPARC)). The SPARC environments allows far-flung teams to conduct real-time observational campaigns using hundreds of coordinated instruments (both ground- and satellite-based). Both the instruments and technical support experts are available on line. Computational models of the object of study (the magnetosphere, in this case) run in parallel with data gathering and help predict where to direct the instruments to the most interesting phenomena. The collaborative data gathering sessions through the collaboratory can be annotated, stored and played back for others to experience at a later date. Students participate in authentic projects in ways previously impossible due to constraints of travel and physical size of instrument rooms. Archival data and the relevant journal literature are also available. "Electronic workshops" occur between the data gathering campaigns to share insight about and interpretations of the experimental data and to collaborate on publications.

Other collaboratories include the following:

- The NSF-sponsored Graphics and Visualization Center, founded in 1991, includes some of the most advanced work in virtual reality applications and shared virtual environments. In particular, the center's Telecollaboration Project aims to "develop a distributed collaborative design and prototyping environment in which researchers at geographically distributed sites can work together in real time on common projects." Research is underway to provide both the hardware and software required.
- The Cooperative Online Resource Catalog (CORC) project of the Online Computer Library Center (OCLC) is exploring the cooperative creation and sharing of metadata (literally "data about data"—data that helps identify, describe, and locate networked electronic resources) by libraries, through the involvement of libraries and other institutions throughout the world. Its aim is to help libraries cope with the huge amounts of material becoming available on the Web (<http://www.oclc.org/oclc/research/>).

## SOFTWARE AND COMMUNICATIONS STANDARDS

Since investments in academic activities are small compared to those in commercial or consumer requirements, it often is necessary for there to be agreement across the global research world if a service is to develop fully. Standards thus are essential, but often lag far beyond the innovations that continuously appear in the world of information technology. Researchers therefore must unify around standards that are necessary in their communities, and must develop educational programs to allow them to benefit quickly from emerging standards.

There are two kinds of standards: those that involve human users and those that operate between digital systems. The former should be changed very slowly, and only for very good reasons, to avoid frustration and wasted time (since the effort to learn a language is considerable). Those that smooth communications among machines are more straightforward, (but as we all know, it frequently is not done correctly).

One example of the former type of standard concerns documents. In 1985, SGML, the Standard Generalized Markup Language, became an international standard. Even after more than fifteen years, few researchers really understand it, and tools supporting it are too expensive to facilitate widespread adoption. However, two descendants of SGML, namely HTML and XML, have been widely exploited by the scholarly community as well as the broader commercial and consumer sectors, and are generally thought to be important mechanisms for interchange. Backed by the World Wide Web Consortium (W3C, see <http://www.w3.org>), these two languages should be as well understood as are convenient word processors, since they allow document creation that enables long-term preservation as well as dual rendering (e.g., in electronic as well as paper publication forms).

Other standards efforts in machine-to-machine communication have shown considerable progress. Agreements on communication protocols have been at the heart of the success of the Internet. TCP/IP, along with other core services like FTP, Telnet, and SMTP (for mail), have made it possible for the Internet to flourish. The national standard Z39.50, and the corresponding international standard ISO 23950, enabled the WAIS service to flourish briefly just prior to the emergence of the World Wide Web, and have supported broad access to library catalogs. Simpler schemes have allowed federated



search across distributed information collections, as in the NCSTRL service (<http://www.ncstrl.org>) for computer science technical reports. This and related work on metadata, through the Dublin Core initiative, has enabled further sharing of scholarly resources through efforts like the Open Archives Initiative [Van de Sompel and Lagoze] (<http://www.openarchives.org>), which promises to move us toward universal access to broad segments of open literature. Researchers should partake of the new services stimulated by such efforts, both as readers and contributors, lest they miss out on the latest findings in their fields.

Other standards will no doubt emerge in key areas to support research activities. In the area of software development, “platform-independent” languages such as Java and Perl have made it easier to reuse and share information and tools. With so much research depending upon software, researchers will save enormous amounts of time if they make investments in purchasing reliable and relevant tools, or adopting free and helpful software (e.g., Linux, or that from the GNU project), as they pursue particular solutions.

### **INTO UNEXPLORED TERRITORY**

The exploding technology of computers and networks promises profound changes in the fabric of our world. That goes for everyone, of course, not just researchers. As seekers of knowledge, researchers will be among those whose lives change the most. What these changes will mean, for academic work and the larger society, remains to be seen. Researchers themselves will build this New World largely from the bottom up, by following their curiosity down the various paths of investigation that the new tools have opened. It is unexplored territory. At the same time, the hoped-for benefits of these systems will depend on their being made available widely and equitably. That is a challenge that the community of researchers, working with public and private funders and regulators of research and technology development, will need to take on over the long term.

## 4

# Changing Expectations of the Researcher

### INTRODUCTION

As earlier chapters have noted, the revolutionary new tools of research and communication clash with some of the traditional values of research. The institutions of research—universities, academic departments, libraries, publishers, and research funding organizations—however, are adapting. The problem for young researchers is that these adaptations are proceeding at varying rates, depending on the specific institutions.

For many young researchers, this environment may seem disordered and chaotic. Digital aids to research such as having wide access to electronic publications places greater stress on the researcher to keep up. Researchers need guidance about what is important, what is highly specialized, and where to find it. Cooperation among individual disciplines of research and the creators of software who serve them will help smooth the process.

Some researchers will find this change healthy. Others will be discouraged. This chapter is intended as an aid for young researchers in navigating their scholarly activities in a digital environment. At the same time, certain characteristics of successful research have not changed. The values of cooperation, open communication, intellectual rigor, and persistence, are traditionally passed from generation to generation by the ap-

prenticeship of students to their mentors. These relationships will be prized and preserved, even as the research environment changes in all the ways this Committee has described. The value of personal relationships is likely never to be replaced by digital technologies.

## TEAMWORK

The conduct of research often involves collaboration and teamwork at least in the sense of building on the work of predecessors. In some cases the scope and/or scale of work demands a team approach, often requiring participants with complementary expertise. The word “teamwork” is used here in both a strong and weaker sense. In the strong sense it means tightly coupled intellectual collaboration with others who are on the “same team.” In the sponsored research world, a team may have received joint funding and have mutual self-interest in a successful outcome of the collaboration.

The word is also used in the weaker sense of participating in—giving to and taking from—a specialized research community of common interests. One may be a participant in and a good citizen of one’s research community, but at the same time compete for recognition and support. An example of this kind of teamwork is the increasing expectation that experimenters contribute data to a common electronic repository or archive. These practices require more explicit articulation and agreement about rules for data sharing and rights of usage. In some fields, new covenants and relationships of trust are being negotiated.

In the digital age, researchers find it increasingly a possibility—and thus an obligation—to maintain awareness of other research in one’s field and to consider one’s work within the context of others’ work, wherever these researchers are geographically located. Researchers can now follow the work in their specialization on a day-by-day basis, for example through Web sites. Through digital technology, researchers are increasing their evaluation not only relative to their home and national institutional peers, but also with respect to their peers around the world.

Technology may also be exacerbating the age-old tension between cooperation and competition in research. On the one hand, there is infrastructure and mutual gain through contributing data, processes, and findings to the common good, and this can be done at nearly instantaneous rates. On the other hand,

this same technology makes research results or even facilities readily available to one's competitors on a global scale.

There are some who hope that digital computer and communication technology will allow larger, more efficient interdisciplinary research teams to form that can better tackle complex societal problems. If this is the case, perhaps researchers will increasingly see it as an obligation to participate in team research. In addition, digital technology will enable some researchers to successfully assume different roles in different teams, as the leader of one, participant in others, occasional consultant in some, and observer/learner in others.

Technology also can enable young newcomers to a field to interact with the "veterans" of the field in new ways that transcend geographic and institutional boundaries. Students can seek mentoring and access to facilities beyond their local mentors and advisers. On the other hand, veterans may have little incentive to interact with junior colleagues. Technology can also provide easier interaction among the hierarchies of scholarly activities. But expectations for both newcomers and veterans need to be better defined. For example, research universities and four-year colleges need to offer meaningful research experiences to undergraduates in order to attract the best students into their graduate programs.

### PROFESSIONAL DEVELOPMENT

Information technology provides important tools for the conduct of research and teaching, communication, administration, and personal productivity. Increasingly powerful computers allow one to solve complex problems and simulate natural phenomena, while networks provide access to scientific literature and data and help one communicate and collaborate with colleagues. The impact of information technology on teaching can be equally profound, linking students and faculty through networks, allowing access to original materials, and supporting new forms of pedagogy such as distributed virtual environments. The array of tools for productivity—electronic mail, word processing, spreadsheets, contact managers, symbolic mathematical tools, presentation software, Web browsers and search engines, data warehouses, and multimedia—continues to expand.

Failing to keep pace with the rapid evolution of these tools can put one at a considerable disadvantage. While it is not

necessary to be an early adopter of each wave of new technology, some facility with mainstream technology such as the Internet and personal productivity tools has become essential for keeping pace with the research community. Recent surveys suggest that many researchers worry that they are falling behind their colleagues in their use of information technology. They fear the resulting isolation from their research colleagues and an inability to take advantage of the emerging tools for teaching and research necessary for the classroom and laboratory.

Institutions must provide sufficient resources to enable researchers to keep pace with developments in information technology through robust technology environments, staff support, and training opportunities. Today's researchers must make the personal commitment to take advantage of these opportunities—to keep pace with technology just as they try to keep pace with research advances in their fields.

### REWARDING INFORMATION TECHNOLOGY ACHIEVEMENTS

Researchers who use information technology can improve the quality and productivity of their research, teaching, and administrative activities. To that degree they should be rewarded, through compensation, promotion, and tenure. Yet in many organizations (particularly in universities), personnel evaluations are based primarily on traditional criteria, such as peer assessment, scholarly publications, grantsmanship, graduate student supervision, and teaching, without additional consideration for the innovative use of technology. Junior faculty and research staff should keep pace with information technology to the extent that it improves their professional performance, but they should be realistic in expecting rewards for doing so.

Spending too much time developing information technology for the classroom or laboratory has hurt the careers of many junior researchers. Promotion reviews often fail to weigh properly achievements such as software and hardware development. Publishing in online journals or through sophisticated Web sites is usually discounted in favor of publishing traditional peer-reviewed articles in well-established journals. An elegantly coded subroutine developed to solve a research problem or novel application of multimedia in the classroom may go

unnoticed or unappreciated by senior colleagues accustomed to evaluating scholarly publications in peer-reviewed journals.

While this Committee can encourage promotion and tenure reviews to be more inclusive in evaluating young researchers, junior faculty and staff are wise to understand clearly the criteria used in making these important career decisions, and plan and conduct their efforts accordingly.

## PUBLICATIONS AND DISSEMINATION

The research process is incomplete without responsible publication. Although the technology and forms of publication are changing, the expectations for researchers are not.

In the “publish or perish” world of academia, the enhancement of reputations leading to promotion, tenure, and research funding are often motives for publishing, but they should never dominate the underlying purpose of communicating results to others.

The first expectation of all researchers is accuracy. Whether or not a publication is subject to peer review, the author is responsible for checking all facts and references, for identifying material that is doubtful and not making assertions beyond those supported by the evidence. A key element of establishing accuracy is proper attribution. Almost every work builds on other research. Correct and full attribution is essential acknowledgement of collaborators.

The next expectation is timeliness. Particularly when work has been supported by public funds, good researchers recognize an obligation to present their results while they are still fresh so that others can build on them. They also seek to publish research where it will be read. In science research, this is often a journal that is indexed by a service, such as Medline for biomedical research or INSPEC for the physical sciences, but sometimes an open-access repository or a Web publication may reach a wider audience. Some research has characteristics that inhibit timely publication: the peculiarities of patent law discourage early publication; industrial sponsors wish to keep results away from competitors; and some government organizations classify certain categories of results. Universities have fairly specific regulations about such restrictions. Individuals have an obligation to be aware of constraints on publication before starting the research.

Individuals have yet another obligation. The system of

publication is based on collegiality. Editing and reviewing are important tasks that everyone shares. The care with which these tasks are carried out is of fundamental importance: they are time-consuming and frequently tedious, but the quality of academic publication depends upon them.

Publication can be one of the most rewarding aspects of being a researcher. A steadily growing list of peer-reviewed publications that others read and use as a basis for their work is one path to excellence.

## PATENTS AND COPYRIGHTS

“The right to search for truth implies also a duty: one must not conceal any part of what one has recognized to be true.”  
(Albert Einstein)

Without the sharing of information, science would grind to a halt. But universities and academic researchers in many fields increasingly have come to think of their results as “intellectual property,” to be exploited for personal gain rather than shared. This shift has resulted in part from changes in patent and copyright laws.<sup>1</sup> These changes have been accompanied by industry’s growing reliance on academic research, and rapidly expanding federal research investments in areas that have strong interest for the private sector (such as biomedical research and computer science). University technology transfer offices work actively to encourage faculty members to patent significant results, so that the institutions and researchers may share the stream of royalties that result from licensing the patents to companies. Academic research is generally growing more competitive.

As the financial stakes rise, the tradition of scientific openness and sharing is endangered. To protect their positions in competition for funding or the race for scientific primacy, some researchers withhold results and techniques that are of general interest and utility. Others succumb to pressure from institu-

---

<sup>1</sup>Notably the Bayh-Dole University and Small Business Patent Act of 1980, as amended through the years (P.L. 96-517). The Act permitted federal grantees such as universities for the first time to retain title to federally funded inventions (if they licensed them speedily to industry). A discussion of the Act from the universities’ point of view can be found at <http://www.cogr.edu/bayh-dole.htm>

tional administrators to maximize patent royalties, or from industrial sponsors to protect trade secrets. For these reasons, research faculty may be reluctant to share information with colleagues in the traditional ways. They may even be tempted to withhold scholarly information from their graduate students—contradicting a fundamental feature of the mentor–student relationship.

Some protections of information are acceptable. The taxpayers who support academic research do so partly because they anticipate future commercial benefits. Commercial applications are often impossible without patent or copyright protection to the companies who wish to apply them. Society must balance these needs with the long-term health of the research enterprise, which depends on the flow of information. Generally, for example, companies that fund academic research are permitted brief periods (a few weeks) in which they may withhold information on results that are thought to be suitable for patenting, but agree to make everything else public immediately.

It is also necessary to protect copyrighted information. Journals and commercial databases, for example, are generally protected by copyright. The so-called fair-use doctrine permits such information to be used in education or research (thus encouraging the spread of knowledge through the community of researchers), but forbids wholesale reproduction and any commercial uses. The easy availability of information on the Internet has tempted some students and researchers to pirate information. Ken Abernethy and Tom Allen of Furman University have made a practical “Overview of Copyright Law in the Digital Domain,” aimed at the needs of researchers, available (at <http://s9000.furman.edu/DD/book/chap2/essay/prop1.html>). It includes references to more comprehensive and technical treatments. The Virginia Tech Web site for submitting dissertations for electronic publishing includes numerous copyright resources (at <http://etd.vt.edu/issues/> and <http://scholar.lib.vt.edu/theses/copyright/>).

Researchers are continually faced with decisions about patenting and copyright protection of their work. The choices are often difficult and subtle ones, despite efforts by research funding agencies, universities, and the companies to develop policies that draw guidelines between what is acceptable and what is not. Some universities offer online resources and training for graduate students and faculty in their rights and obliga-



tions. Each researcher must be aware of the laws and policies involved in using information obtained from others. The reader may obtain some guidance from *On Being a Scientist: Responsible Conduct in Research*, a related report of the National Research Council (1995).

### ENSURING THE INTEGRITY OF RESEARCH IN THE CHANGING DIGITAL ENVIRONMENT

Although most researchers routinely do not think explicitly about the fundamental values driving their activities, there is considerable agreement within the research communities of scientists and engineers that a core set of values at least implicitly guides the ideal practices in their fields. They include the pursuit of truth, knowledge, and understanding; the ability to change and control parts of the world; the ability to build things; the betterment of the human condition through health, preservation of the environment, general economic advancement, happiness, beauty, education, and stimulation of the young. The values served are universal. Ideally, scientific and engineering research serves universal ends as well. Intended for the benefit of all people, good research is not explicitly parochial, nationalistic, racist, class conscious, and so on.

Many norms and conventions governing the conduct of research serve these values. Strict methodological guidelines, for example, represent a research community's ideal standards for achieving sound, reliable results. The convention of peer review is intended to ensure that published work meets the standards of a particular research community and helps build a body of work that is trustworthy. Blind peer review protects against irrelevant biases. Open sharing of research findings stimulates progress by enabling researchers to build incrementally upon the achievements of others. To enable the union of diverse skills and expertise, collaboration and teamwork have evolved. Fair attribution of credit for discoveries and results promotes progress by encouraging excellence. The norms that reflect principles of ethical conduct in research serve not only general moral ends, but also the ends of scientific integrity, truth, and competence.

This report describes many of the new digital research tools for science and engineering. These tools have irreversibly revolutionized the way research is now and will be conducted. For the most part research communities are adapting by wel-

coming these changes as necessary improvements for fostering productivity, speed, efficiency, precision, collaboration, and creativity. This willingness to adapt is generally perceived as positive and progressive (Nissenbaum, 1998).

Amidst the general optimism, the Committee sounds cautionary notes about possible impacts of the new research practices. If core values are weakened or deflected by the use of digital technology, then the relevant practices should be carefully examined and redirected and, if necessary, rejected. A few examples illustrate the extent of these challenges:

- Electronic publication opens vast possibilities, which many have welcomed as ways to free researchers from the hold of commercial publishers. Publication practices have changed and will no doubt continue to change. We recommend, however, that as communities make these changes, they bear in mind how these changes affect the relevant values, such as truth and integrity. Because publication mediates the attribution of credit and priority we must consider how such norms will be affected by electronic publication.

- The traditional research value of data sharing is also being challenged. This value is not an absolute one, since researchers frequently don't reveal information about works in progress. Electronic sharing of information is different from traditional sharing in significant ways. For example, the researcher faces a greater loss of control if information sharing can so easily be disseminated by others. When devising digital norms, we must not necessarily adopt norms that are most similar to the old, but rather, norms that serve the valued ends to the same extent, or better, than the old.

- Researchers protecting intellectual property may interfere with teamwork and data sharing interests.

- Financial disparities among individual researchers and institutions with respect to access to costly elements of the digital infrastructure may produce unfair distributions of the advantages of having these important assets.

In all of these cases, researchers must apply the tests of their fundamental values.

## **THE ENDURING FOUNDATIONS OF RESEARCH**

Many excellent researchers in science and engineering will challenge disciplinary boundaries and confound expectations. Other more traditional and conservative researchers may be no less excellent or creative. Academic communities tend to retain traditional customs and governing structures. These researchers can do profound and important work, opening new horizons for their colleagues, without going beyond the limits of their disciplines.

The electronic tools of communications and computing enhance both innovative and traditional practices in research. But they may favor the innovators, by making it easier to work across boundaries and to ignore the rules. The community of researchers will be forced to examine those boundaries and rules increasingly as digital tools grow more powerful and pervasive. The community is certain to be strengthened by this radical review of its foundations.

## References

- Brown, John Seely, and Paul Duiguid. 1995. The social life of documents. Xerox PARC working paper, available at: <http://www.parc.xerox.com/ops/members/brown/papers/sociallife.html>
- Brown, John Seely, and Paul Duiguid. 2000. *The Social Life of Information*. Cambridge, Mass.: Harvard University Press.
- Bush, Vannevar. 1945. As we may think. *Atlantic Monthly*. July. See also <http://www.ps.uni-sb.de/~duchier/pub/vbush/vbush.shtml>
- Card, S., J. Mackinlay, and B. Shneiderman. 1999. *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann Publishers.
- Casper, Gerhard. 1995. Come the millennium, where the university? Address by the President of Stanford University to the annual meeting of the American Educational Research Association, San Francisco, April 18, 1995.
- Casper, T. A., W. M. Meyer, J. M. Moller, P. Henline, K. Keith, B. McHarg, S. Davis, and D. Greenwood. 1998. Collaboratory operations in magnetic fusion scientific research. *Interactions*, vol. 5, no. 3 (May), 56–65.
- Clark, David. 1999. The Internet of Tomorrow. *Science*, vol. 285, 353. July 16.
- Communications of the Association for Computing Machinery. 1998. [Special issue on digital libraries.] Vol. 41, no. 4, (April)
- Davis and Laogze Dienst. 1995. An architecture for distributed document libraries. *Communications of the ACM*, vol. 38, no. 4 (April) 47.
- Duderstadt, James J. 1999. Can colleges and universities survive in the information age? Pp. 1–25 in *Dancing with the Devil: Information Technology and the New Competition in Higher Education*. San Francisco: Jossey-Bass Publishers.
- Edelson, Daniel C., Roy D. Pea, and Louis M. Gomez. 1996. The collaboratory notebook. *Communications of the ACM*, vol. 39, no. 4 (April) 32–33.

- Finholt, Thomas A., and Gary M. Olson. 1997. From laboratories to collaboratories. *Psychological Science* 8(1) 28-35.
- Fox, Edward A. 1999. The Digital Libraries Initiative: Update and Discussion, Introduction. *Bulletin of the American Society for Information Science*, vol. 25, no. 6 (Oct./Nov.) 7-11.
- Graubard, Stephen R., and Paul LeClerc, eds. 1998. *Books, Bricks, and Bytes: Libraries in the Twenty-First Century*. New Brunswick, N.J.: Transaction Publishers. 379.
- Harnad, S. 1991. Post-Gutenberg galaxy: The fourth revolution in the means of production of knowledge. *Public-Access Computer Systems Review*, vol. 2, 39-53. Available at <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad91.postgutenberg>
- Harnad, S. 2000. E-knowledge: Freeing the refereed journal corpus online. *Computer Law & Security Report*, vol. 16, 78-87, 2000.
- IEEE Spectrum*, 1996. "Distributed virtual environments."
- Jones, K. P., and P. Wilett, eds. 1997. *Readings in Information Retrieval*. The Morgan Kaufmann Series in Multimedia Information and Systems. San Francisco: Morgan Kaufmann Publishers.
- Kilman, David G., and David W. Forslund. 1997. An international collaboratory based on virtual patient records. *Communications of the ACM*, vol. 40, no. 8 (Aug.), 110-117.
- Kouzes, Richard, James D. Myers, and William A. Wulf. 1996. Collaboratories: Doing science on the Internet. *IEEE Computer*, vol. 29, Aug 40-46.
- Lenzer, Robert and Stephen S. Johnson, "Seeing Things as They Really Are," an interview with Peter Drucker, *Forbes*, 159 (1977) 122-128.
- Lesk, Michael. 1997. *Practical Digital Libraries: Books, Bytes and Bucks*. San Francisco: Morgan Kaufmann.
- Marvin, Carolyn. 1990. *When Old Technologies Were New*. London: Oxford University Press.
- McDaniel, S. E., G. Olson, T. Weymouth, C. E. Rasmussen, A. Prakash, C. R. Clauer, D. E. Atkins, R. Penmetsa, N. H. Manohar, and H. S. Shim. 1994. The Upper Atmospheric Research Collaboratory. *Proceedings of the CHI '94 Conference Companion on Human Factors in Computing Systems*, 21.
- National Research Council. 1993. *National Collaboratories: Applying Information Technology for Scientific Research*. Washington, D.C.: National Academy Press.
- National Research Council. 1994. *Realizing the Information Future: The Internet and Beyond*. Washington, D.C.: National Academy Press.
- National Research Council. 1995. *On Being a Scientist: Responsible Conduct in Research*. 2d ed. Washington, D.C.: National Academy Press, 40.
- National Research Council. 1996. *Meeting Summary: The Future of Research Universities in the Information Age*. Office of Scientific and Engineering Personnel, in collaboration with the Computer Science and Telecommunications Board, Commission on Physical Sciences, Mathematics and Applications; Government-University-Industry Research Roundtable; and National Academy of Engineering. December 3-4, 1996.
- National Research Council. 1997. *Meeting Summary: The Future of Research Universities in the Information Age*. Held on December 3-4, 1996.
- National Research Council. 2000. *Advanced Engineering Environments: Achieving the Vision, Phase 1*. Washington, D.C.: National Academy Press.

- New Scientist*. 1997. All the world's a lab. No. 2077 (April 12) 24–27.
- Nissenbaum, Helen. 1998. The puzzle of priority: Devising new norms and conventions in research for the context of electronic publication. Available from author (University Center for Human Values, 5 Ivy Land Princeton University, Princeton, N.J. 08544).
- Noam, Eli M. 1995. Electronics and the dim future of the university. Policy Forum. *Science*, vol. 270, 247–249. Oct. 13.
- Odlysko, Andrew. 1995. Tragic loss or good riddance? The impending demise of traditional scholarly journals. *International Journal of Human-Computer Studies*, vol. 42, 71–122. Available at <http://www.research.att.com/~amo>
- O'Donnell, James J. 1998. *Avatars of the Word: From Papyrus to Cyberspace*. Cambridge, Mass., and London: Harvard University Press. See also <http://ccat.sas.upenn.edu/jod/avatars/>.
- Olson, Gary M., Daniel E. Atkins, Robert Clauer, Thomas A. Finholt, Farnam Jahanian, Timothy L. Killeen, Atul Prakash, and Terry Weymouth. 1998. The Upper Atmospheric Research Collaboratory (UARC). *Interactions*, vol. 5, no. 3 (May) 48–55.
- Pear, Robert. 1999. NIH plan for journal on the Web draws fire. *The New York Times*, June 8, D1.
- Roberts, Lawrence G. 2000. Internet watch: (Beyond Moore's law: Internet growth trends. *IEEE Computer*, vol. 33, no. 1 (Jan.). Available at <http://computer.org/computer/co2000/r1toc.htm>.
- Rosenberg, Laurence C. 1991. Update on National Science Foundation funding of the "collaboratory." *Communications of the ACM*, vol. 34, no. 12 (Dec.), 83.
- Schatz, Bruce R. 1997. Information retrieval in digital libraries: Bringing search to the Net. *Science*, Jan. 17.
- Science Magazine*. 1999. Net news: High-tech anxiety. Vol. 285, no. 5434 (September 10), 1635.
- Shneiderman, B. 1998. Codex, memex, genex: The pursuit of transformational technologies. *International Journal of Human-Computer Interaction*, vol. 10, no. 2, 87–106.
- Shneiderman, B. 2000. Universal usability: Pushing human-computer interaction research to empower every citizen. *Communications of the ACM*, vol. 43, no. 5 (May), 84–91.
- Sodan, A. C. 1998. Yin and yang in computer science. *Communications of the ACM*, vol. 41, no. 4 (Apr.), 103–114.
- Sproull, Lee, and Sara Kiesler. 1991. *Connections: New Ways of Working in the Networked Organization*. Cambridge, Mass., and London: The MIT Press. 212.
- Van de Sompel and Lagoze. The Santa Fe Convention of the Open Archives Initiative, D-Lib Magazine, February 2000. <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>
- Waters, Donald and Garrett, John (chairs), *Preserving Digital Information: Final Report and Recommendations*, Task Force on Archiving of Digital Information, Research Libraries Group 1996, <http://www.rlg.org/ArchTF/index.html>.
- Weiser, Mark. 1994. The world is not a desktop. *Interactions of the ACM*, Jan. 7–8.
- Weiser, Mark, and John Seely Brown. 1996. The Coming Age of Calm Technology.

REFERENCES

51

- Xerox PARC, October 5 (available at <http://www.ubiq.com/hypertext/weiser/acmfuture2endnote.htm>).
- Wulf, William. 1989. The National Collaboratory - A White Paper in Towards a National Collaboratory. Unpublished report of a NSF workshop, Rockefeller University, NY. March 17-18.
- Wulf, William A. 1995. Warning: Information Technology Will Transform the University," *Issues in Science and Technology*, (summer, 1995) 46-52.





# Appendix



## APPENDIX A

### Committee Member Biographies

**James J. Duderstadt** (Chair) is President Emeritus and University Professor of Science and Engineering at the University of Michigan. He received his doctorate in engineering science and physics from the California Institute of Technology in 1967. He joined the faculty of the University of Michigan in 1968 and has served as Professor of Nuclear Engineering, Dean of the College of Engineering, and then as Provost and Vice-President for Academic Affairs. He was elected President of the University of Michigan in 1988 and served that role until July 1996.

**William Arms** is Professor of Computer Science at Cornell University, with research interests in digital libraries and electronic publishing. He has degrees from Oxford University, the London School of Economics, and the University of Sussex. During the 1970s, he was a faculty member at the British Open University, helping to develop some of the first computer science courses delivered by distance education. Subsequently, he was Vice Provost for Computing at Dartmouth College and Vice-President for Computing at Carnegie Mellon University. In these two positions he was responsible for early developments of campus networking, digital libraries, and the use of personal computing in education. Before joining Cornell in 1999, he spent four years at the Corporation for National Research Initiatives, as Vice-President.

**Daniel E. Atkins** earned a M.S.E.E. and Ph.D. in computer science from the University of Illinois, Champaign-Urbana, in 1967 and 1970, respectively. He joined the Department of Electrical Engineering and Computer Science as an assistant professor in 1972. In May of 1981, Atkins was promoted to the rank of full professor and in September of that year, in response to extraordinary opportunities in the College of Engineering, he joined the leadership team of the College as Associate Dean for Research and Graduate Studies. From January 1989 through July 1990 he served as interim Dean of the college of Engineering. Since the mid 1980s, Atkins has provided research leadership in the use of distributed computing/communication to support distributed forms for team-based knowledge work. He became founding Dean of the new School of Information in July 1992 and held that position until September 1998. He is now Director of the University of Michigan Alliance for Community Technology.

**Mark H. Ellisman** is Professor of Neuroscience and Bioengineering and the Director of the Center for Research in Biological Structure at the University of California, San Diego, where he has taught since 1977. He received his masters degree in neurophysiology in 1974 and Ph.D. in molecular, cellular and developmental biology in 1976 from the University of Colorado at Boulder. He now directs the National Center for Microscopy and Imaging Research, an internationally acclaimed technology development center and research resource established by the National Institutes of Health. He is a pioneer in the development of three dimensional light and electron microscopy and combined application of these image acquisition tools and computational technologies to achieve greater understanding of the structure and function of the nervous system.

**Edward A. Fox** holds a Ph.D. and M.S. in Computer Science from Cornell University. Since 1993 he has been at Virginia Polytechnic Institute and State University, where he serves as Associate Director for Research at the Computing Center, and Professor of Computer Science. For the Association for Computing Machinery he served during 1988-91 as a member of the Publications Board and as editor in chief of ACM Press Database Products. He is editor for the Morgan Kaufmann Publishers book series on multimedia information and systems and serves on many editorial boards.

**Paul Messina** is Assistant Vice President for Scientific Computing at the California Institute of Technology, Faculty Associate in Scientific Computing, and Director of Caltech's Center for Research on Parallel Computation. He is Chief Architect for the National Partnership for Advanced Computational Infrastructure, an UCSD-led partnership established by the National Science Foundation. He received his Ph.D. in mathematics in 1972 and his M.S. in applied mathematics in 1967, both from the University of Cincinnati. In 1997 the University of Lecce, Italy, awarded him an honorary degree in computer engineering, for his contributions to computational science.

**Helen Nissenbaum** is a Research Associate and Lecturer at the University Center for Human Values, Princeton University. She teaches and conducts research in the area of social and ethical dimensions of science and technology, focusing on information and communications technology. She is author of *Emotion in Focus* (University of Chicago Press), and co-editor (with D. J. Johnson) of *Computers, Ethics and Social Values* (Prentice Hall). In 1998, as co-principal investigator for an interdisciplinary study of human values in Web browser security, she was awarded a three-year research grant from the National Science Foundation. She serves on the Board of Directors of the Society for Philosophy and Technology, and as co-editor of the journal *Ethics and Information Technology*. She earned her M.A. in Education and Ph.D. in philosophy from Stanford University.

**Ben Shneiderman** is a Professor in the Department of Computer Science, Head of the Human-Computer Interaction Laboratory, and Member of the Institute for Advanced Computer Studies and for Systems Research, all at the University of Maryland at College Park. He is the author of *Software Psychology: Human Factors in Computer and Information Systems* (1980) and *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (1987, second edition, 1992, third edition, 1998). He is on the Board of Directors of Spotfire, Inc. and has been on the editorial advisory boards of nine journals. He received his Ph.D. from State University of New York at Stony Brook in 1973. He received an honorary doctorate of science from the University of Guelph, Ontario, Canada, in 1996 and was elected a fellow of the Association for Computing Machinery (ACM) in 1997 and of the AAAS in 2001.