Office of Special Projects

MEMORANDUM

Date: January 6, 1999

To: William Arms
Paul Messina
Mark H. Ellisman
Daniel E. Atkins
Edward A. Fox
Ben Shneiderman
Nancy G. Leveson
Joshua Lederberg

From: Jim Duderstadt, Chair

Subject: Conference Call #1: On Being a Researcher in the Digital Age

First, let me thank all of you for agreeing to participate in this project. I believe we have a wonderful opportunity to provide guidance to young scholars/researchers. Please find enclosed the agenda (Tab 1) and related materials.

After brief introductions (Tab 2) and a review of the project description (Tab 3), we will spend the majority of the time discussing the draft outline (Tab 4) developed by the virtual guidance group (Tab 5). We will want to hear from you regarding both your general and specific comments. Are there gaps in the outline? Does the outline accurately capture the key questions and concerns facing the young scholar/researcher? Should any sections or items be added or deleted?

Next, we will address next steps (Tab 6). This project envisioned one meeting of this panel to gather input from other experts. We are interested in holding a workshop in February/March with the expectation that we could have several sessions on specific issues of concern. For example, we might want to hear about efforts of the National Archives and Records Administration as they deal with the issue of preservation of data and electronic records whose formats are obsolete or are becoming obsolete. Similarly, the Association of Research Libraries is struggling with implications of digital information, and legal scholars are dealing with intellectual property rights in the digital age. Please give some thought to the issues we might want to address at the workshop. In addition to the workshop, the project provides for several campus focus groups to provide comments on the outline and report.

Finally, during this first conference call of our panel we will need to review bias/conflict so please take a moment to review and complete the attached forms (Tab 7). (Please fax your form to Brett Willette--202-334-1667.) As part of this process we will want to be certain that we are bringing multiple perspectives to the table and that we have the right mix of individuals involved in this activity.

CONFERENCE CALL INFORMATION

DATE:	Monday, January 11, 1999
START TIME:	3:00 p.m (EST)
END TIME:	4:00 p.m. (EST)
PHONE NUMBER:	1-800-226-7400
PASS CODE:	3670 #

On January 11th, at 3:00 p.m. please dial the "1-800" number, when prompted please dial the access code number and the (#) button. You will then be connected to the call. Please announce your presence so we know you have been connected. You can call in to the conference call from ANY phone in the United States if you are traveling.

If you experience any problems, please contact Brett Willette at 202-334-2455 immediately.

c: Anne-Marie Mazza (project director)
Brett Willette (research associate)
Eric Brown (research assistant)
Duncan Brown (writer)

On Being a Researcher in the Digital Age

Conference Call #1
Jim Duderstadt, Chair

AGENDA

3:00	Introductions	Tab 2
3:15	Panel Charge and Background on Activities to Date	Tab 3
3:25	Comments on the Outline	Tab 4
4:05	Next Steps	Tab 6
4:20	Bias Forms	Tab 7
4:30	Adjourn	

Panelists

On Being a Researcher in the Digital Age

Chair
James J. Duderstadt

Digital Libraries and Databases
William Arms

Computation as a Third Paradigm
Paul Messina

Research Applications -- Remote Instrumentation
Mark H. Ellisman

Collaborative Research
Daniel E. Atkins

Communication of Research
Edward A. Fox

Ethical Issues
Ben Shneiderman

Software Engineering
Nancy G. Leveson

Researchers
Joshua Lederberg

JOSHUA LEDERBERG

Joshua, a research geneticist, is Sackler Foundation Scholar and President-emeritus at The Rockefeller University in New York. Dr. Lederberg was educated at Columbia and Yale University, where he pioneered in the field of bacterial genetics with the discovery of genetic recombination in bacteria. Because of their simple structure and rapid growth, bacteria have afforded geneticist a fruitful field for research, which has also ripened recently into many medical and industrial applications. In 1958, at the age of 33, Dr. Lederberg received the Nobel Prize in Physiology or Medicine for this work and subsequent research on bacterial genetics. Dr. Lederberg has been a professor of genetics at the University of Wisconsin and then at Stanford University School of Medicine, until he came to The Rockefeller University in 1978. A member of the National Academy of Sciences since 1957, and a charter of its Institute of Medicine. Dr. Lederberg has been active on many government advisory committees and boards, including the National Advisory Mental Health Council, and the Chair of the President's Cancer Panel. He has long had a keen interest in international health, and is serving term on WHO's Advisory Health Research Council. He recently co-chaired the study of Emergency Infections which has appeared under the imprint of the Institute of Medicine. Long interested in the communication of science to the public, and for public policy, Dr. Lederberg also authored a weekly syndicated newspaper column for six years; and he co-chaired the Carnegie Commission on Science, Technology and Government.

In addition to his lifelong research in the molecular genetics of bacteria, viruses, and the immune system, he has also been engaged in the applications of computer science for innovative discovery in biomedical science. For his work at Stanford on the DENDRAL program for interpretation of mass spectra, he was recently awarded the Allen Newell prize of the Association for Computing Machinery. In the period 1957-1977 he was also deeply engaged with NASA in mission design for the search for evidence of life on Mars, culminating in the Mars Viking lander experiment.

Dr. Lederberg has been awarded numerous honorary Doctor of Science and M.D. degrees, and the LL.D. from the University of Pennsylvania, as well as a Foreign Membership of the Royal Society, London, and the rank of Commandeur in the Ordre des Arts et des Lettres of the French Republic. In October 1989 he received the U.S. National Medal of Science, and in February 1997 the Maxwell Finland Award of the National Foundation for Infectious Diseases. He is an honorary life member of the N.Y. Academy of Medicine and its John Stearns awardee for 1996, and is a past chairman and now honorary life governor of the New York Academy of Sciences.

He retired as president of the Rockefeller University in July 1990 and continues his research activities there in the field of DNA secondary structure and mutagenesis in bacteria.

ON BEING A SCHOLAR IN THE DIGITAL AGE
OFFICE OF SPECIAL PROJECTS

PROJECT DESCRIPTION

The digital age offers both wonderful possibilities for improving the caliber of research as well as the potential for reducing the quality of research through lack of knowledge or improper conduct. At the *Future of Research Universities in the Information Age* planning meeting held at the Academy on December 3-4, 1996, participants urged that a guidance document similar to *On Being a Scientist*: *Responsible Conduct in Research* be developed in this area.

Participants stressed that the scientific process encompasses a wide range of technical, social, and procedural activities, each of which involves information--information is collected, combined, analyzed, derived, discussed, and distributed. Some, if not all, of these activities may and often do benefit from the application of computer and networking technology. All are subject to the knowledge, attitudes, and preferences of individual scientists, the prevailing norms or cultures in individual fields, and the constraints and incentives imposed by relevant institutions, from those that employ or educate researchers to those that fund or otherwise support them.

Sometimes lack of knowledge or improper conduct lead to actions that can reduce the quality of research. Concerns arise such as data property rights, confidentiality, improper or uncertain attribution, insufficient attention to copyrighted material, various degrees of review of online publications, impact on tenure decisions, etc. Or, scholars are not reaching the full potential of their research because they are unaware of the possibilities that the digital mechanism could lead to whole new areas of research and new methods of investigation and thus to science that was not and could not be done before.

A guide is needed that will both outline the possibilities and provide guidelines for beginning scholars so that proper conduct is more clearly defined

PLAN OF ACTION

Statement of Task

This guide for students and faculty will 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.

Expertise Required

A committee of approximately 10 individuals with background in electronic collaborative research, ethics and the Internet, digital libraries, remote research, computational models, and an understanding of the likely future potential and problems will develop the guide.

This committee will be overseen by a guidance group of knowledgeable experts [*see attached list*] who will nominate committee members and evaluate the results of their deliberations.

Preliminary Work Plan

The first step in the project would be to develop and test a topic outline of the document with a focus group consisting of student and faculty members who are the likely audience for this document. An in-depth outline will then be developed and tested with a different focus group. Using the revised version of that outline, a science writer will then develop the text which will again be checked by a focus group.

The document will then be reviewed, completed, and disseminated.

The budget supports staff time for the development of the document including use of a science writer and several days of senior staff consultation. The panel developing the document will meet at least once. In addition, at least three focus group at universities outside of the Washington are included in the budget.

Product and Dissemination Plan

The product of this activity will be a guide--at this point estimated at approximately 60 pages--similar in format to *On Being a Scientist: Responsible Conduct in Research* [*enclosed*].

Some uses for this guide include:

- Library Internet training sessions
- University information technology training sessions
- Classes on Internet and ethical issues
- Briefing sessions for new students and faculty (either to university or laboratory) perhaps as part of a standard ethics package in conjunction with *On Being a Scientist*

A great deal of attention would be paid to the WWW version of the document which would be the primary way we would disseminate the document and let universities know of the existence of the document. At that point, universities could then order the document in bulk for a very low rate for use in training sessions, etc.

On Being a Researcher in the Digital Age
Outline

"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." (Vannever Bush, July 1945)

I. Introduction: How Digital Information & Communication are Changing the Nature of Research

A. The Changing Nature of Research

1. Electronic journals have made a wider range of research more accessible to researchers.

2. Researchers can now search library catalogs and view items in museum collections worldwide.

3. Online data allows biochemists download protein structures; chemists grab IR spectra.

4. Electronic correspondence allows physicists post preprints and social scientists conduct surveys.

5. Remote Instrumentation: A telescope on a mountain in Chile, a synchrotron in Illinois, a supercomputer in San Diego--all are online, to be programmed or guided remotely by researchers throughout the world.

B. Doing Science in a Digital Environment

The scientific process encompasses a wide range of technical, social, and procedural activities (see Figure 1 for one view from the CSTB National Collaboratories report), each of which involves information--information is collected, combined, analyzed, derived, discussed, and distributed. Some, if not all, of these activities may, and often do, benefit from the application of computer and networking technology. All are subject to the knowledge, attitudes, and preferences of individual scientists, the prevailing norms or cultures in individual fields, and the constraints and incentives imposed by relevant institutions, from those that employ or educate researchers to those that fund or otherwise support them.

C. Advice for Faculty and Students

1. The opportunities and implications of conducting research in a digital environment.

2. The scientific, technical, and structural barriers to making the best use of these opportunities.

3. The "how" of digital 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.

4. Dealing with new behavioral norms and issues associated with those new norms.

5. Information quality: issues for both producer and user

II. Development and Use of Communal Digital Assets: Digital Libraries and Databases

A. The Digital Library

1. What is it?

2. Its purpose

A digital library enables users to interact effectively with information distributed across a network. These network information systems support search and display of items from organized collections.

3. Its potential

From a researcher point of view, this has the potential to provide researchers to immediate access to all scientific literature and the ability to move to "concept search" as opposed to "document search" as in the past.

B. Databases

1. How networked databases work

By developing and using communal assets such as online databases, whole communities can take advantage of these resources. These communities can then post new and update old information in these databases for use by everyone. Of course, this results in some concerns such as who has "property rights" to this data. The perspectives on this issue continue to evolve as more of these databases are put into place.

D. Information Reliability

1. Electronic Lab Notebooks

a. Notorization

b. Watermarking

c. Fingerprinting

2. Archiving Research Publications

Because electronic publication allows for greater fluidity in the content of articles, methods of tracking how articles have changed over time are essential for research refernces.

3. Data Quality

E. Examples

A University of Colorado group funded by NASA built a prototype system for distributing satellite data over the network via an on-line access system. In this system, the user is responsible for locating, reviewing, and ordering the data, an approach that saves human time and cost in locating, copying, and shipping data tapes. During the recent oceanographic experiments in the western Pacific, satellite sea-surface-temperature maps were produced and made available over the Internet to researchers.

III. Use of Computation as a Third Research Paradigm with Theory and Experimentation

A. Concept of the "Third Modality"

Researchers now routinely talk of computation as the "third modality" of scientific investigation, on a par with theory and experimentation. It has blossomed in the space between the "field" and the "laboratory" becoming a new way to process information and derive knowledge.

B. Examples

1. The easy examples are those that simply automate what had been done manually, such as the reduction of data and the control of instruments.
2. Simulations of observed phenomena such as radio jets streaming out of galaxies and quasars and the patterns of plasma within a fusion reactor.

C. New Ways of Doing Research

The profound applications, however, are those that lead to whole new areas of research and new methods of investigation and thus to science that was not and could not be done

before. These applications have transformed the nature of scientific investigation; they have led to questions that would not even have been asked before.

1. Analyzing molecules that have not been synthesized.

2. Measuring the properties of a single neuron by "growing" it on a silicon chip.

3. Watching a model of galaxies colliding.

4. Systems that allow scientists to feel the forces as a drug docks in a protein.

5. Studying decision making processes in a variety of simulated organizational structures.

IV. Collaborative Research

A. The Collaboratory

1. Definition

A collaboratory, (a term derived by combining the words "collaboration" and "laboratory"), involves the integration of various digital tools, "to provide an environment that enable scientists to make more efficient use of scientific resources wherever they are located." It 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." The National Collaboratory - A White Paper in *Towards a National Collaboratory* (Wulf, 1989). It is an, "integrated, tool-oriented computing and communications system to support scientific collaboration." Cerf, V.G. et al., 1993 National Collaboratories: Applying Information Technologies for Scientific Research, NRC

2. Requirements for functional collaboratory

a. Shared secure data

b. Shared analysis

c. Shared instruments

d. Shared interaction space

3. The advantages of the collaborative approach

Research can now be conducted with greater ease than before across disciplinary, university, and national boundaries. "A collaboratory provides a technological base specifically created to support interaction among scientists, instruments, and data networked to facilitate research conducted independent of distance." Wulf, 1989.

4. How a "collaboratory" works

By forming a "collaboratory" researchers are able to create 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.

B. Examples

1. Genetics

Researchers studying the genome of the fruitfly were able to clone approximately 90% of the genome in the species onto 10,000 fragments by using a new Internet tool that facilitated information sharing and collaboration as opposed to approximately 3,00 loci mapped using conventional methods.

2. Existing collaboratories in process and sponsors, (appendix)

C. Efficacy of Collaborative Research

1. When is it a good idea to conduct electronic collaborative research? When is it not?

D. Preparing for and Participating in Collaborative Research

1. Things to consider before engaging in digital collaborative activities

E. Sociology of collaborative research

V. Changing Research Apparatus

A. Remote access to instrumentation

At the UCSD, an intermediate voltage electron microscope is available such that scientists all over the world can watch the results of explorations using this microscope. This avoids each campus having to fund such a microscope as well as the extraordinarily highly-trained technicians needed to support such equipment.

B. MEMS microinstrumentation

Via NSF's nanofabrication facility, anyone in the country in any discipline --even a university with minimal resources--can access this network to get a microdevice made that a researcher needs to conduct their research.

C. Sensor-equipped devices

Provide scientists with the ability to conduct research that is more portable, integrated, and shareable.

D. Using the Net to Study Human Behavior

Social scientists are using the net as a new way of studying relationships between people. Psychologists are studying interpersonal relationships in electronic groups; sociologists and anthropologists are studying electronic group dynamics; economists are studying electronic auctions. This new resource for social scientists raises several ethical and methodological questions.

1. How do researches navigate the research ethics issues surrounding privacy, informed consent, protected populations, etc.

2. What methodological considerations arise concerning the acquisition of data through data mining, data logging, etc.

VI. Communication of Research

A. Electronic Publishing

1. A more complex definition

2. How it works

23 Professional implications

Different types of electronic publishing relative to journal practices and peer review and the implications of institutional allegiances, questions of tenure and promotion, anonymity and privacy of a researcher choosing to publish in one type versus another and of e-pub versus paper-pub with no electronic version.

B. Who owns what on the Net

Networked publication of both raw and processed data and questions of intellectual property, confidentiality, and authorial credit.

C. Multimedia Presentations

The results of research can now be communicated with a richer, multimedia signal that integrates text with sound, graphics and video, and that can be interactive.

D. Electronic Copyright

E. Publication of Software and Databases

VII. Software Reliability

A major concern in the development of electronic experiments and models is the validity of the software used in conducting the experiment or forming the model. Researchers have a responsibility to be as certain as possible that the software they use is reliable.

A. Software Concerns Facing Digital Age Scholars

1. It is difficult to write correct programs and none of the complex programs today are bug free.

2. Coming to understand the problems in software and learning techniques to decrease the probability of serious era.

3. Determining the probability that what comes out of a digital experiment is truth.

4. Understanding that problems with software arise in computational not mathematical procedures.

B. Ensuring Best Quality Software

1. Write the software yourself

2. Buy software from reputable manufacturers

3. Hire competent people who use the best techniques available in the production of accurate programs.

4. Use total quality management when producing software

Testing software after it has been produced is much less accurate and efficient than preventing bugs during production.

5. Look at all data more critically when it comes from a program not written by yourself.

C. Box -- Dealing with software reliability problems.

VIII. Netiquette (from a researcher's point of view)

A. What is Netiquette

Rules that Internet users are inclined to accept and expected to follow. Newcomers who behave in a manner inconsistent with these rules are subject to reprimand from the group.

B. How Does Netiquette Impact the Digital Researcher

What do issues like the following imply in terms of proper behavior from a electronic age researcher point of view: Free Speech; Electronic Vandalism; Intellectual Property; Privacy.

C. Common Sense on the Net

Of course, in the end, researchers need to use their common sense and social skills to act properly in this ever-changing research environment.

IX. Ethical Issues (Obligations for Researchers on the Net)

A. New Rules for the Researcher

Issues such as properly allocating credit and documenting ever-changing research need to be addressed so that everyone is operating using the same rulebook.

1. Credit

2. Documentation

B. Special issues for electronic research on human behavior

1. Informed consent

2. Public -vs- private behavior

X. Career patterns

A. How Career Patterns are Changing

Career patterns are changing as the potential of the digital age is beginning to be realized.

B. How Digital Age Resources Effect the Research Mix

This effects issues such as where research is conducted, the mix of theoretical versus experimental research, big versus small projects, solo versus team efforts, etc.

XI. Conclusion

A. Recommendations
B. Linkages to new patterns in education

References

Bush, Vannevar. 1945. As We May Think. Atlantic Unbound. July.

NRC. 1997. Meeting Summary: The Future of Research Universities in the Information Age. Held on December 3-4, 1996.

NRC. 1993. National Collaboratories: Applying Information Technology for Scientific Research. Computer Science and Telecommunications Board.

NRC. 1994. Realizing the Information Future: The Internet and Beyond. Computer Science and Telecommunications Board.

NRC. 1994. Rights and Responsibilities of Participants in Networked Communities. Computer Science and Telecommunications Board.

NSF. 1996. Connecting and Collaborating: Issues for the Sciences. Report of a Workshop sponsored by NSF held at the UCSD on June 22-24, 1995. (Electronic Publication only at http://www.sdsc.edu/projects/CC/Paper)

Schatz, Bruce R. 1997. Information Retrieval in Digital Libraries: Bringing Search to the Net. Science. 17 January.

Wulf, William A. 1996. Warning: Information Technology Will Transform the University, (Online only publication at http://w3.scale.uiuc.edu/scale/library/Wulf.html)