Robbin, A. (1992). Social scientists at work on the electronic network. *Electronic Networking: Research, Applications and Policy*, *2*(2), 6-30.

electronic

networking:

research, applications and policy

Volume 2, Number 2 • Summer 1992

Meckler

Electronic Networking (ISSN 1051-4805) is published quarterly by Meckler Corporation, 11 Ferry Lane West, Westport, CT 06880 (203) 226-6967; Internet: Meckler@jvnc.net; CompuServe (70373, 616) Applelink: Meckler.

Subscriptions: $95.00 (£60.00) per year. Please add $15.00 for subscriptions to Canada and Central and South America. Orders from North, Central, and South America should be sent to *Electronic Networking*, Meckler Corporation, 11 Ferry Lane West, Westport, CT 06880; from elsewhere to Meckler Ltd., 247–249 Vauxhall Bridge Road, London SW1V 1HQ, UK, 01-931-9985. All back issues are available. *Electronic Networking: Research, Applications, and Policy* is indexed by ERIC for the ERIC database, and *Current Journals in Education* and by *Information Science Abstracts*.

POSTMASTER: Send all address changes to *Electronic Networking*, Meckler Corporation, 11 Ferry Lane West, Westport, CT 06880.

electronic
networking:
research, applications and policy

Contents

Vol. 2 ■ No. 2 ■ Summer 1992

Social Scientists at Work on Electronic Research Networks

Alice Robbin

The purpose of this article is to contribute to our stock of knowledge about who uses networks, how they are used, and what contribution the networks make to advancing the scientific enterprise. Between 1985 and 1990, the Survey of Income and Program Participation (SIPP) ACCESS data facility at the University of Wisconsin-Madison provided social scientists in the United States and elsewhere with access through the electronic networks to complex and dynamic statistical data; the 1984 SIPP is a longitudinal panel survey designed to examine economic well-being in the United States. This article describes the conceptual framework and design of SIPP ACCESS; examines how network users communicated with the SIPP ACCESS project staff about the SIPP data; and evaluates one outcome derived from the communications, the improvement of the quality of the SIPP data. The direct and indirect benefits to social scientists of electronic networks are discussed. The author concludes with a series of policy recommendations that link the assessment of our inadequate knowledge base for evaluating how electronic networks advance the scientific enterprise and the SIPP ACCESS research network experience to the policy initiatives of the High Performance Computing Act of 1991 (P.L. 102–194) and the related extensive recommendations embodied in Grand Challenges 1993 High Performance Computing and Communications (The FY 1993 U.S. Research and Development Program).

A significant body of research has been conducted on computer-mediated communication (CMC) since the early 1970s.[1] Laboratory experiments and field studies have contrasted use of CMC and other media in small group settings (Archer, 1990; Lea, 1991a and 1991b; Rice & Case, 1983; Smilowitz et al., 1988). Message transmission and content have been studied (Danowski, 1982, 1988; Stohl & Redding, 1987). Analysts have examined the relationship between the communication channel and system and organizational characteristics (Grote & Baitsch, 1991; Hiltz & Johnson, 1990; Rice, 1982; Rice & Shook, 1988; Rice & Torobin, 1986; Sproull, 1986). Concomitantly, researchers have investigated the process of individual and organizational innovation adoption (Rice, 1982; Rice, 1987; Rice & Associates, 1984; Rice & Torobin, 1986; Rice et al., 1990; Turoff, 1989; Williams et al., 1988). Much research effort has been devoted to the effects of CMC on social structure, social behavior, and the socio-emotional content of the messages exchanged between participants in a social network (Kiesler et al., 1984; Rice & Associates, 1984; Rice & Love, 1987; Sproull & Kiesler, 1986, 1991a, 1991b). Some investigations have been designed explicitly to examine task-related activities and/or productivity gains (Olson & Bly, 1991; Pappa, 1990; Pappa & Tracy, 1988; Rice & Case, 1983; Rice & Shook, 1988; Steinfield, 1986; Weedman, 1991).

During a period of exponential growth in scientific networks, few researchers however, have specifically addressed the role of CMC in advancing the scientific and technical enterprise. Nonetheless, we have accumulated a substantial body of evidence about the nature of communication in the scientific and technical community,[2] and a substantial amount of CMC research has been conducted on subjects

Alice Robbin <Internet: ROBBIN@SSCB.SSC.WISC.EDU> is senior scientist at the University of Wisconsin-Madison, associated with the Institute for Research on Poverty since 1984. She has written extensively on the data delivery system for social scientists; been principal investigator of various research grants; and was a finalist in the 1989 Computerworld/ Smithsonian Institution Award for Innovative Use of Information Technology. She and a colleague in the Department of Mathematics are currently developing a prototype of an information system for geometry preprints and will evaluate the diffusion of scientific discovery through the electronic network. Dr. Robbin can be reached at the Institute for Research on Poverty, 3412 Social Science Building, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706.

located in university settings. The rapid expansion of national and worldwide scientific electronic networks has been due principally, some might contend, not to empirical evidence that demonstrates their utility, but to successful advocacy by members of the scientific community. Anecdotes and subjective reports by computer and physical scientists have been relied on heavily to support assertions that CMC and the electronic networks have been responsible for improved intellectual "connectivity" and scientific advances (see Comer, 1983; Jennings et al., 1986; Markoff, 1990 and 1992; National Research Council, 1989). Illustrative examples of applications of high-performance computing and computer communications technologies have served as powerful tools to build executive and legislative constituencies to support expansion of the electronic networks (see *Grand Challenges*, 1992).

The conclusion reached from a review of the extensive literature on CMC is that we have an inadequate knowledge base about whether and in what ways electronic networks have furthered developments in science and technology and contributed to scientific productivity. There are, for example, few published accounts of actual implementation of scientific projects in the electronic environment to support the contention that electronic networks play a critical role in advancing the scientific enterprise (for an exception, see Cinkosky et al., 1991). Gould (1990, p. 12) notes that, "There is little empirical evidence on the nature of a network's ability to enhance the research process and increase productivity." McClure and colleagues (1991, p. 88) concur that, "There is little empirical evidence, for example, about how many researchers are regular users of the existing national network structure, what the vast majority of researchers use networks for, and about how networks affect R&D work."

Relatedly, Olson and Bly (1991, p. 220) comment that, although a great deal of research has been conducted on the nature of work groups, "very little research in any field has specifically focused on how work is performed in and managed by groups," and, in particular, work groups supported by information technology. (For an exception, see Galagher et al., 1990.) Consequently, we know little about, for example, the content of electronic scientific communications, substitution of CMC for other communications media, the development of intellectual networks, the extent of increased "connectivity" between geographically dispersed or formerly functionally fragmented members of a scientific community, or research outputs or other productivity gains that have resulted from CMC.

A potent explanation for this inadequate knowledge base is that the research agenda for advancing science and technology through computation and networks has been set by engineers and computer scientists. This research agenda has been validated by policy language that unquestionably regards improvements in national competitiveness as almost wholly deriving from technical and engineering efforts. Neither the High-Performance Computing Act of 1991 (U.S. Congress, 1991) nor *Grand Challenges 1993: High Performance Computing and Communications (The FY 1993 U.S. Research and Development Program)* (*Grand Challenges*, 1992) acknowledges that building infrastructure requires knowledge of how scientific activity takes place in human organizations — through cognitive, psychological, and social processes and structures through which information is transmitted and new knowledge is created. But as social scientists Sproull and Kiesler (1991b, p. x) note, "because the technology is used for communication, it has an impact on the most critical process in an organization: whether and how people communicate." As such, improvements in the physical components of the infrastructure must be coupled with improvements in our knowledge base about communication in scientific organizations so that we can demonstrate empirically whether and in what ways networks advance the scientific enterprise.

Toward this end, this article presents preliminary data on the communications of social scientists as they carried out their work on a large data set using electronic networks. The data derive from a recently completed assessment of the SIPP ACCESS research project, which developed an Information System for Complex Data (ISCD) for social scientists and policy analysts in the public and private sectors.[3] The R&D goal was to develop a prototype data facility that would increase the efficiency and effectiveness of the research process by improving access to, providing efficient and low-cost retrieval of, and fostering scientific communication about large-scale, complex and dynamic statistical microdata through electronic research networks.

Part One of this paper introduces the reader to scientific communication and feedback problems faced by social scientists who analyze complex data sets and relates the design of the SIPP ACCESS project to these problems. Part Two summarizes the literature on computer-mediated scientific communication and discusses its relevance to the project's goal of improving communication to enhance scientific productivity. We also briefly explain the measurement problem associated with establishing a causal link between scientific activity and productivity, which led to a decision to study the process between inputs and outputs —the intermediate steps in the research activity—and to develop non-economic performance-effectiveness indicators of the output derived from the SIPP ACCESS project.

Part Three profiles the communication process of social scientists in the SIPP ACCESS research network and examines one non-economic performance-effectiveness indicator of this process. We first describe the source of our data, electronic mail (e-mail) archived by the project between Spring 1985 and August 1990. We then discuss the CMC process that unfolded during this period, including growth in network traffic, relationship between communication, and type and duration of use of the facility, participants in the communication process, and subject matter of the messages. Lastly, we examine improvements in data quality to show how computer-mediated scientific communication contributed to advancing the social scientific enterprise.

Part Four summarizes the benefits that resulted from the electronic networks and offers a series of policy recommendations. These recommendations link the assessment of our inadequate knowledge base for evaluating how the electronic networks advance the scientific enterprise and the SIPP ACCESS research network experience to the policy initiatives of the *High-Performance Computing Act of 1991* and the related extensive recommendations embodied in *Grand Challenges 1993.*

PART ONE: THE SIPP ACCESS PROJECT

The SIPP ACCESS project was established in response to national concerns that a large investment had been made in the production of longitudinal panel surveys between the 1960s and early 1980s but that subsequent use of these data had been much smaller than anticipated. We believed that at least four conditions had precluded use of complex data. The intellectual and capital investment required for exploiting these data had been lacking: widespread, advanced training in using complex data and adequate computational equipment were unavailable. The size, scope, and complexity of these data were significant impediments to timely and efficient access and retrieval. At the time we wrote our proposal in 1984, appropriate technologies for efficient and low-cost data reorganization and retrieval, communication of scientific information, and exchange of data were largely unavailable to the social research community. The existing social science infrastructure for conducting research and policy analysis was not designed to respond optimally to the dynamic environment of data production, distribution, and utilization.

Nature of the Problem

Information science provides a perspective on why the use of complex data had been limited.[4] One emphasis has been on information flows or communication interactions between entities in an information system. Clark (1986, p. 83) and Dolby, Clark, and Rodgers (1986, p. 96) have noted two problematic characteristics of data: the communication chain is long (from data collection through analysis and presentation), and information is not documented. A great deal of information does not get communicated; indeed, it is permanently lost. The data producer allocates resources primarily to the data collection and production processes, with the result that there is inadequate support for subsequent data use. Lack of institutionalization of the data producer's "memory" after a project is completed has meant that many public data sets cannot be shared by scholars unassociated with the original data collection. Despite efforts to record the decision-making process that leads to the production of a public use data set, documentation is rarely deemed adequate for the task at hand, and it has always been very difficult to locate relevant information. This creates a problem of *access* to data (David & Robbin, 1981).

Communication failure also characterizes data *use* environments, where there is either no communication or only intermittent communication between a data producer, user, expert, and the knowledge resource (data set and accompanying description). The primary source of expertise, the data producer, has little or no communication with the secondary analyst to help solve the methodological, data quality, or analytical issues arising from a complex scientific design and processing decisions. Thus, researchers spend a considerable amount of time through trial and error during the analysis stage, making discoveries about the quality of the data set. Secondary sources of expertise, that is, those analysts who spend a considerable amount of time investigating the data, do not have a vehicle to collate and communicate their expertise and experiences with the data. This results in similar, time-consuming rediscovery by a new cohort of secondary analysts.[5] Research findings are widely circulated in advance of publication to elicit comment and criticism, but many novice users fail to identify or retrieve these papers. In addition, the secondary analyst's discoveries about the utility of the data are not communicated to the data producer before changes are made in the scientific design of a data collection. Furthermore, expertise about the data is not institutionalized in an expert; data repositories have not served as this vehicle or as a source of expert advice for methodological management, or analysis problems.

The widely accepted implementation of a knowledge resource described precludes the iterative accumulation of essential knowledge about data, knowledge that is generated dynamically and interactively. Contributions are made by many entities in a communication system that include libraries, archives, and human or machine experts. Instead, the knowledge resource pertaining to a data set remains static, and updates to it are intermittent and rare. Applying this communications perspective, we concluded that an information system designed primarily to disseminate data, without feedback, would fail to meet research needs.

Finally, our assessment of the data delivery system also indicated that the organization supporting researchers' activities was equally important to how the researcher approached complex data. Researchers were geographically dispersed, employed in a variety of occupations, and subject to differing constraints. Variation in these dimensions affected the scarcity and location of computational resources, the availability of communication networks, the ability of the researcher to delegate technical tasks to programmers and assistants, and the access that the researcher had to expert consulting.

Design of SIPP ACCESS

Our proposal to the National Science Foundation suggested a different framework for a data delivery system from the one that had been implemented between the 1960s and early 1980s in the United States and elsewhere. Complementing local and national data archives which had provided the structural foundations for improvements in data delivery, we proposed a prototype of an Information System for Complex Data (ISCD) to stimulate the production and sharing of knowledge about a complex data set by the data producer, project staff *cum* expert, and social scientists in a research network.

At this time, the U.S. Bureau of the Census was fielding a longitudinal panel survey called the *1984 Survey of Income and Program Participation* (SIPP). It was designed to provide a comprehensive and integrated profile of economic well-being of Americans (Ryscavage, 1987). The SIPP was a response to known deficiencies in our poverty statistics, as well as a recognition of the interactions among social program participation, work, income and assets, health, education, and family structure. The scientific design of the SIPP also responded to a deficiency in survey methodology — namely, that cross-sectional surveys fail to identify the dynamic nature of behaviors and events; a longitudinal data collection was needed. SIPP was produced by the U.S. Bureau of the Census at a cost of about $50 million over a two-and-a-half year period.

A great deal of information does not get communicated; indeed, it is permantly lost.

We selected the SIPP panel survey as a test data set. These panel data are considered perhaps the most complicated set of survey data ever created, containing about seven million observations and more than 20,000 variables that represent 36 months of data for sample members. In size, this data set is far smaller than the terabytes of data collected by projects in the physical sciences, but it is structurally more complex. Longitudinality itself implies a complex data structure because variation over time occurs in measurements, analysis units, and aggregation.

In the SIPP panel data, reference time and the periodicity of measurement produce three modes of time series. Matching people across interviews to create a longitudinal sample requires expensive data processing on a mainframe. Two central problems are to identify who is in the sample at one time and to determine how long they remain in the sample over the length of the panel. The standard sequential (flat) public use file throws away much of the information that is built into the survey's complex file structure about the relationship of measures on different units of analysis. It also obscures the various sample universes defined by the questionnaire's skip patterns and relational structure. The Bureau of the Census released nine cross-sectional files (one for each interview) and a 32-month time series subset of the sample population and variables over a three-and-a-half year period between Spring 1985 and Fall 1988.

We created a central facility organized around a relational database management system (Ingres RDBMS) to allow sharing of the SIPP data by a number of users and to reduce the very large overhead associated with retrieving data from longitudinal panel surveys and creating analysis files. To improve access to statistical data, the ISCD's infrastructure integrated information, statistical data, computers, software, and communications. Information and discoveries about and experiences with the data would be shared by all analysts associated with the research network and permanently available through the libraries and archives of the central node. We believed that learning time by future analysts would be lessened, and that scientific output would be increased because the cost of access and the potential for error would be reduced, the scientific design of the data would be clarified, and research results would be more rapidly generated. Discoveries about the data communicated to the data producer would lead to future improvements in data quality. Finally, by diffusing knowledge about and fostering adoption of the SIPP ACCESS innovation, we hoped that a new generation of social researchers would be trained to employ new technologies for managing large-scale, complex data.

A public facility was created at the University of Wisconsin-Madison Physical Sciences Laboratory (PSL). The PSL facility was available to analysts located in the United States and elsewhere by telephone dial-up and an evolving infrastructure of electronic networks, including BITNET, ARPANET, and the Internet. Database design and development work were carried out on computers at the Institute for Research on Poverty (IRP) and Center for Demography and Ecology (CDE), located on the main campus about 20 miles from the PSL facility. The CDE and IRP computers were clustered, and the cluster was linked to PSL through the university's local area network. After database development was completed, files were transferred over the local network to the PSL facility relational database management system and Optical Archive Store (OAS).

PART TWO: EMPIRICAL RESEARCH ON CMC USE BY THE SCIENTIFIC COMMUNITY

As noted in the introduction, the corpus of empirical research on use of CMC to further the scientific enterprise is small. Here we review findings from selected studies that examined whether the availability of the computer altered the medium, content, or frequency of scientific information exchange or stimulated scientific productivity that contributed to the production of new knowledge. We exclude from this review studies of subjective satisfaction with computer-based conferences for scientific group discussion (e.g., Hiltz & Johnson, 1990; Tombaugh, 1984). We also exclude studies that examined the task-related content of electronic conference messages because the research was not designed explicitly to study the scientific process or outcomes (e.g., Rice & Love, 1987; Weedman, 1991).

Early Research, 1970s – Early 1980s

Much of what we know about the use of CMC by the scientific and technical community derives from the pioneering work of Hiltz, Turoff and associates on the Electronic Information Exchange (EIES) computer-conferencing system sponsored by the National Science Foundation (Hiltz, 1984; Hiltz et al., 1978; Hiltz & Turoff, 1978; Kerr & Hiltz, 1981; Turoff & Hiltz, 1978; Turoff & Hiltz, 1982). Their field investigations, conducted between the early 1970s and early 1980s, produced no conclusive findings about the substitution of computer communication for mail or telephone. Hiltz (1984) found, for example, that the computer could lead to an increase or decrease in telephone communication, but such change was a function of the group context (i.e., nature of the problem). Because there was great variability in group structures, there was no way to predict accurately whether substitution would occur. However, electronic communication appeared to have stimulated additional communication that would not have taken place in the absence of rapid and reliable message transfers (p. 173).

Research that sought to examine productivity gains achieved through CMC was inconclusive. For example, Johansen and colleagues' (1978) longitudinal study of the potential of computer conferencing did not yield results that indicated that scientific productivity had been increased by computer networks. Electronic networks did not appear to increase scientific collaboration at one research site (Gerola & Gomory, 1984). On the other hand, the EIES project surveyed users on productivity-related factors and found small effects due to CMC. Kerr and Hiltz (1981, pp. 118–119) reported that EIES members of the Joint Electron Device Engineering Council believed that computer conferencing had improved the quality of work, face-to-face interactions, and decision making. Hiltz (1984), while noting that "objective measures of scientific productivity were difficult to obtain" (p. 177), found that, "Overall, 52 percent of the respondents to [Hiltz's] surveys report[ed] general impacts on working patterns," including "broadened professional perspectives, increases in communication or connectivity, change in perspective of self and cosmos caused by the communication medium, and specific work habits" (p. 176). Multivariate analysis revealed that the perceived productivity increase was related to time spent on the EIES system, number of people actively communicating with, and number of new people met through the system, but these three variables explained only 29 percent of the variance (pp. 183–184).

Recent Research, Mid-1980s – 1990

More recent evidence about the substitution of CMC for other communication media comes from research-in-progress conducted by the Center for the History of Physics of the American Institute of Physics. The Center is documenting the history of institutional collaboration in high energy physics since the 1970s (American Institute of Physics, 1990). Project staff have found that beginning in the middle 1980s, fax and electronic mail began to be substituted for mail and telephone; this occurred because scientists found the traditional mail too slow and the international telephone lines clogged and undependable.[6]

Hesse and associates (in press) investigated the relationship between access to electronic networks and scientific outcomes of oceanographic research. Sources of data included interviews with oceanographers, program officers in funding agencies, and network administrators and a survey of 338 subscribers of SCIENCEnet who were sent a ninety-three-item survey on their professional background, research activities and network use between September 1988 and early 1989 (pp. 6–7, 10, 12, 25–26). They examined three types of scientific outcomes: publication, professional recognition, and social integration. Hesse and his colleagues found that the more active network users were also more active and productive scientists, associated with more prestigious research institutions, received more professional recognition, published more, and knew more oceanographers (p. 13). Their results suggested that access to networks could benefit oceanographers who were geographically on the "periphery" (i.e., remote from resources).

McClure and colleagues (McClure et al., 1990, 1991) conducted an exploratory research project on scientific use of the electronic national networks. They investigated how many researchers were regular users, how they used these networks, and how these networks affected their work. Their goal was to describe the effect of existing networks on the research process and scientific communication. McClure and his colleagues relied on questionnaires and focus groups (136 researchers) in different organizational settings to gather personal background, work history, and network-related knowledge and use. These researchers found that the major effect of the networks on users was to increase the efficiency of scientific communication. For some participants in the study, networks had made an important contribution to data access, analysis, and interpretation (1991, p. 102). McClure and colleagues suggested that the majority of the researchers had not yet been affected by electronic networks. Networks complemented the traditional patterns of scientific communication, but the research process had not been altered (1991, pp. 103, 111–112). Overall, however, these investigators concluded that networks "encouraged researchers to communicate more often with a broader range of individuals," although most of the communication was of a personal nature and not linked to research activities (1990, p. 27).

Relevance to Scientific Communication in the SIPP ACCESS Research Network

Research conducted by Hiltz, Turoff and their associates during the 1970s and early 1980s, provided the SIPP ACCESS project with an initial framework for investigating scientific communication. Although the focus of their earlier investigations was the electronic conference, they offered a host of insights about the communication process, social structure, user satisfaction, and computer technology. The McClure and Hesse projects were conducted concurrently with ours and, as a result, their findings serve as a source of comparison and important additions to our knowledge base on scientific communication in the research network. In particular, as relevant to the discussion in this article, the McClure study noted that electronic networks had facilitated resource sharing and provided access to data that would have otherwise been unavailable. The Hesse project demonstrated a statistically significant relationship between use of networks and scientific activity.

These few studies on CMC use by the scientific community, as well as studies about the conduct of science, illuminated two problems for us. First, when we began our project in the mid-1980s, most of our knowledge about the conduct of science derived from studies of the "hard" sciences. There was no research literature written specifically about the conduct of quantitative social science, although there were a few studies that examined social scientists (cf. Hagstrom, 1965, 1970; Crane, 1972; Hargens, 1975). Without institutionalized recordkeeping and reporting by the social science community, we had no basis for comparing electronic and other media communications about statistical data. The Hiltz and associates work did not yield evidence of the conduct of scientific investigations that relied on access to complex and dynamic statistical data.

Second, investigators who attempted to assess the relationship between the output of science and information technology faced significant methodological problems. One problem they encountered was associated with the measurement of scientific activities or outputs and their relationship to productivity.[7]

"Research cannot prove that the network use causes productivity to increase."

Hesse and his associates (in press, p. 6) comment: "Despite such claims [that networks have enhanced scientific productivity], there has been no systematic empirical research on how computer network support for science is related to important scientific outcomes, including productivity." According to these researchers, "Research cannot prove that the network use causes productivity to increase" (p. 6), because it is not feasible to conduct randomized experiments or collect longitudinal data owing to a highly dynamic network system, population, and resources.

Although we may debate their assertions about experiments and measuring a "moving target," these are different methodological problems from the measurement of scientific productivity. (For an experimental design using electronic networks, see Bikson & Eveland, 1990.) The problem is that objective measures of gains are elusive. In a 1986 report, the Office of Technology Assessment (OTA) noted the difficulties of relating R&D expenditures to scientific productivity or other economic benefits (U. S. Congress, Office of Technology Assessment, 1986). In OTA's view, the principal benefits of research, new and unexpected knowledge, could not be assigned a direct economic benefit. Furthermore, scientific activity had indirect effects, such as "spillovers" and "spinoffs," which were difficult to quantify.

OTA recommended that analysts attend to the *process between inputs and outputs (outcomes)*, that is, the intermediate steps in the research activity, and OTA devoted considerable discussion to gaining insights on scientific productivity through noneconomic indicators. This, too, was the strategy that Hesse and his colleagues and Andrews (1979) and his colleagues employed in their respective studies of the scientific enterprise. Both groups of researchers responded to the problem of productivity measurement by examining the process of scientific activity and establishing a broad set of quantitative or qualitative "performance-effectiveness" indicators. We, too, followed this strategy in assessing the SIPP ACCESS project. As such, our discussion on electronic communication in the research network examines the communication *process* and relies on one noneconomic indicator of the "output of science" derived from scientific communication, improvements in data quality.[8]

PART THREE: COMPUTER-MEDIATED COMMUNICATION ABOUT COMPLEX STATISTICAL DATA

Electronic mail (e-mail) was intended to be the principal medium for communicating with SIPP ACCESS users for four reasons. First, e-mail was viewed as an efficient way to meet user information needs in a way that did not impede the development of the facility; the project staff was small, and everyone performed multiple roles (e.g., database designer, programmer, analyst, and consultant). Second, e-mail supported a cooperative work environment, ensuring that the entire staff would be informed about the needs of the user community and that the most qualified staff person would respond to a user.[9] Third, e-mail would encourage users to explain clearly and precisely what problem they encountered. Fourth, in contrast to telephone calls and face-to-face consulting, minimal work was required by the project staff to maintain a permanent record of communications.

Our discussion of electronic communication in the SIPP ACCESS network relies on a sample of e-mail messages for the period Spring 1985 to August 1990. The first section below discusses the sample and coding decisions. The second section profiles the electronic mail communications archived by the SIPP ACCESS facility.

Source of Data

Several thousand e-mail communications circulated between SIPP ACCESS users and project staff, among the project staff, between the project staff and the Bureau of the Census liaison, and between project staff and computer and data center staffs of the Physical Sciences Laboratory (PSL), where public access to the SIPP database was available, and the clustered computers of the Center for Demography (CDE) and Institute for Research on Poverty (IRP), where database design and development took place.[10] At the end of 1990, fifty-seven months of e-mail messages were reviewed to remove duplicates and messages pertaining to project administration, preliminary analysis by the data archive staff before the data were loaded into the SIPP databases, and internal communication among project members.[11] The sample consists of 1,646 e-mail messages that were deemed relevant to studying scientific communication between the period April 1985 through August 1990.[12] Thirteen hundred and forty-nine of these messages were transmitted between analysts and the project staff; fifty-four messages between project staff and data or computer center staff at PSL or CDE/IRP; and 243 messages between the U.S. Bureau of the Census liaison and the SIPP ACCESS project staff. Messages from data or computer center staffs were included only if they pertained to communications about an analyst's work in the databases.

Restricting the e-mail sample to communications between the analyst and expert and between the expert and data producer means that we do not report on the full array of relationships that occur naturally in an information system for complex data.[13] For example, while we archived e-mail between the project staff and the data archive and computer center staffs that was directly related to analysts, we lack records of analyst–computer center staff traffic unless a communication was forwarded or copied to us by either the analyst or the computer center staff. Similarly, we have no record of communications between the data producer and the analyst, nor could we ever obtain one.[14] The latter gap is more serious, and we will have more to say about it in a later part of our discussion. Furthermore, although we hoped that SIPP ACCESS would serve as a vehicle for analyst-to-analyst communication, we have few records to indicate that this occurred; the only e-mail for this dyadic relationship occurred if a member of the SIPP ACCESS staff conducted research and maintained a record of these communications.

E-mail messages were edited to record the following information: local (LAN) or remote (wide area) network communication, source of the message (person's name), target of the message (person's name), date (day, month, year, time of day), and subject(s) of message. Each person was then coded according to the role played in the information system (e.g., analyst, expert, data producer, data/computer center staff). "Analysts" conducted research on the SIPP data. The "data producer" was the U.S. Bureau of the Census. The "expert" was a member of the SIPP ACCESS project staff. The "data/computer center staff" were support staff of the IRP, CDE, or PSL computer facilities. Throughout our discussion we use the terms "analyst" and "researcher" interchangeably.

Contents were analyzed for their meaning and the subject(s) of the message coded and classified; coding required a significant level of knowledge about the SIPP ACCESS system and the SIPP data.[15] A keyword technique could not be applied because the descriptive language used by every analyst to explain problems differed significantly. Classifying the subject(s) of a communication was not always an easy task. Knowledge of the context of the communication was critical, such as the history of the series of questions that had been posed by an analyst and the responses provided by the SIPP ACCESS staff.

Findings

In this section we provide descriptive statistics on these electronic mail communications. We examine the transmission of e-mail to and from the SIPP ACCESS staff, communications and their relationship to use of the SIPP ACCESS facility, the relationship between the sender (source) and receiver (target) of the message, and the contents (subject) of the message. The section concludes with a discussion about improvements in data quality.

Network Traffic

Electronic mail circulated locally at PSL and between nodes at PSL, CDE or IRP, and over the wide area network, via BITNET, ARPANET, and, later, the Internet. Table 1 shows the distribution of communications sent via the LAN and the wide area network. The number of messages transmitted between the SIPP ACCESS staff and members of the research network grew significantly over the years. Between 1985 and 1989, the number increased from year to year between about 200 and 300 percent, and there was a 100 percent increase in the rate of communication from January to August 1989 through the same period for 1990.

Table 1. Number of Local and Wide Area Network E-Mail Communications, April 1985–August 1990

Type of Network	1985	1986	1987	Calendar Year 1988	1989	1990[1]	Total
Local	14	43	104	178	420	394	1153
Wide-Area	5	1	36	146	223	82	493
Total	19	44	140	324	643	476	1646

1 *January–August only*

Communications sent via the wide area electronic network accounted for only 15 percent (N=493) of the total number of communications between 1985 and 1990; 1,153 messages were transmitted locally within the PSL-CDE-IRP systems.[16] The principal reason for this is that, in the early years of the project, the majority of PSL analysts used the telephone to dial into PSL. Therefore, only the LAN would have been relied on for electronic communications between project staff, analysts, and the Bureau of the Census liaison. Between 1987 and 1989, however, wide area network communication accounted for between 26 percent and 45 percent of the total number of communications that took place in those three calendar years.

This increase in use of wide area networking for communications was consistent, first, with an increase in the amount of work in the SIPP database being carried out by analysts and, second, with the apportioning of available and costly resources of the educational infrastructure. Analysts used the SIPP ACCESS facility to reduce the high costs of processing their analysis files or because they lacked adequate computational resources at their home institution. Although they had access to statistical software at PSL, nearly all researchers preferred to use their home site if it provided low-cost or free computing for statistical analysis. After they retrieved data from the database, they transmitted their analysis files to their home institution. They only communicated with the project staff when their analysis revealed problems with the data that they themselves could not solve.

Figure 1 below shows the geographic dispersion of e-mail communications for our sample, excluding Australia and Western Europe, and the different modes of network use. If e-mail exchanges on administrative matters, including communications with members of our advisory board, project planning, and conference planning and participation were included, Figure 1 would indicate a far more extensive communication network, extending throughout the United States and Western Europe. The large empty square indicates the University of Wisconsin-Madison LAN; the small square, wide area network, illustrating that e-mail was sent remotely from the home site; the circle, LAN use by dial-up; and the diamond, Internet

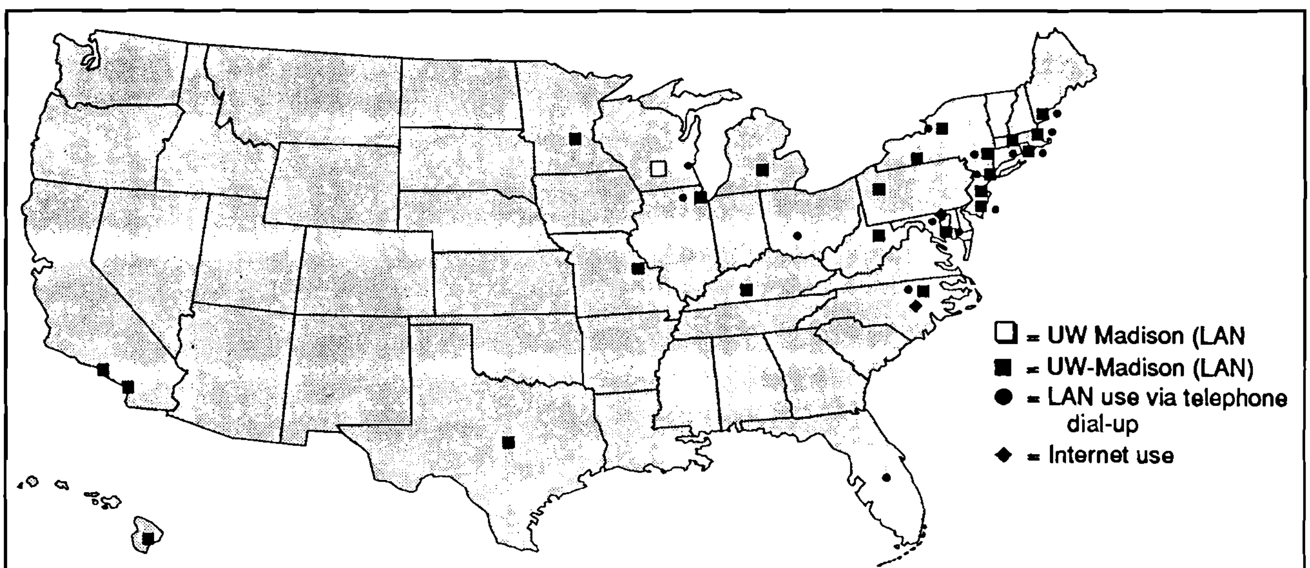

Figure 1. Geographic Dispersion of SIPP ACCESS Communications, April 1985– August 1990

connection leading to LAN use via computer-to-computer connection.

Whether communications took place via the LAN or wide area network depended on the network resources at the analyst's home institution and the type of use made of the SIPP ACCESS facility. Note that Figure 1 shows a concentration of communications with individuals located at institutions in the eastern part of the United States. This concentration indicates where a significant amount of poverty research was being conducted during that period.

Relationship of Communication to SIPP ACCESS Facility Use

Table 2 compares the communication patterns of five types of data facility analyst users. The column "Communicators" represents those analysts in a particular population from whom we received or to whom we sent an e-mail message (e.g., user of the PSL computer facility, workshop participant, or SIPPTEST, SIPPRUN, or PC-SIPPTEST database).[17] The column "Total" indicates the total number of accesses to the PSL computer facility for that particular population (totals from administrative records).

For example, 31 percent of those who logged into PSL (only to check mail, for example) actually communicated with the project staff via e-mail. Only 26 percent of the workshop participants ever communicated with the SIPP ACCESS staff. This statistic reinforces findings derived from administrative records that the majority of people who attended our training workshops never went on to use the SIPP ACCESS facility. On the other hand, 52 percent of the SIPPTEST database analysts communicated with the SIPP ACCESS staff. Excluding the four SIPP ACCESS staff who conducted research on the SIPPRUN database, 82 percent of the SIPPRUN database analysts communicated with the project staff. Lastly, 55 percent of the PC-SIPPTEST analysts had at least one electronic communication with the SIPP ACCESS project. Note that these five different populations of analysts are not independent from each other. For example, workshop participants used the SIPPTEST database, and some of them also used the SIPPRUN database. Some of them also bought the PC-SIPPTEST database.

We also examined the growth by calendar year in the number of persons with whom the staff communicated and the duration over which communications occurred. There was a significant reduction in the number of people who maintained a long-term relationship with SIPP ACCESS. Administrative records on accesses to the database support our inference that the majority of the analysts only used the data facility to complete a particular research project and that only relatively few analysts had multiple research projects that required using the facility over an extended period of time.

Dyadic Relationships

There are eleven different source and target communication relationships possible in an information system for complex data that includes the entities of analyst, expert consultant, computer/data center staff, and data producer. Our e-mail archive is principally a subset of the relationships between analysts and expert and expert and data producer, for reasons that we explained earlier.

Table 3 shows the growth in traffic flows by calendar year as a function of source and target. The project staff-analyst relationship accounted for 75 to 80 percent of the traffic during the entire period. Communications between the project staff and the Bureau of the Census accounted for, at most, 20 percent of the

Table 2. Users (Analysts) of the SIPP ACCESS Facility, April 1985–August 1990

Activity	Communicators	Users[1]
Login PSL	49	156
Attend Workshop	34	131
Use SIPPTEST	55	105
Use SIPPRUN	42	50[2]
Use PC-SIPPTEST	28	51
Total persons	121[3]	493[3]

1 *Includes users on any project connected to the SIPP ACCESS facility*

2 *Excluded are six project staff. Two SIPPRUN users who did not communicate with the SIPP ACCESS staff had other members of their team who did communicate.*

3 *Because persons engaged in several activities, the sum exceeds the total persons.*

dyadic traffic in a calendar year. Note that the number of communications sent (source) differs from the number of communications received (target). No one-to-one correspondence exists between source and target because some mail message may have required no response, particularly if the message served to communicate information or provide reference assistance. It was unusual to hear from a user again after the project staff had responded to the user's information request.

The number of communications by analysts as source or target increased substantially from April 1985 through August 1990. For example, analysts sent eleven messages in 1985 and 181 in 1990. We see a corresponding increase in the number of messages *received by analysts* (target of message): from six messages in 1985 to 268 messages in 1990. The SIPP ACCESS staff's communications *as source of message* also increased between 1985 and 1990, from eight messages in 1985 to 288 messages in 1990, with a significant increase from 1988 (N=83) to 1989 (N=361). A large increase in communications *targeted at* the SIPP ACCESS staff also occurred for this period of time (from thirteen messages in 1985 to 198 messages in 1990). Our administrative records show that analysts made the most extensive use of the SIPP ACCESS facility in 1988 and 1989. We did, however, register a very significant decline from 1989 to 1990 in communications with the Bureau of the Census, which was due to closing the SIPP ACCESS facility at the public PSL facility and transferring the SIPPRUN database to the Bureau in April 1990. In the next section we look more closely at the subject matter of the communications to understand why there was a substantial increase in 1989 and 1990.

Table 3. Number of Communications by Type of User, April 1985–August 1990

	Calendar Year							
	1985	1986	1987	1988	1989	1990 (actual 9 mon.)	1990[2] (Annual total)	Total[3] Messages
Source of Message[1]								
Analyst	11	23	78	187	215	181	241	708
Census Bureau liaison	0	5	31	45	54	4	5	126
Computer Center	0	0	6	9	13	3	4	31
SIPP ACCESS staff[4]	8	16	25	83	361	288	384	781
	19	44	140	324	643	476	634	1646
Target of Message[5]								
Analyst	6	18	51	80	271	268	357	694
Census Bureau liaison	0	2	6	18	70	8	11	104
Computer Center	0	1	1	4	10	2	2	18
SIPP ACCESS staff	13	23	82	222	292	198	264	830
	19	44	140	324	643	476	634	1646

[1] *User is defined as source (originator) of message. A unique user may have more than one project account; in this case, each user-project is counted as a unique user.*

[2] *January - August 1990 computed on an annual basis.*

[3] *Total includes 1990 computed on three-quarters of a year.*

[4] *SIPP ACCESS staff always counts as one source, even when several project staff members are responsible for communicating as SIPPASSIST.*

[5] *User is defined as target (recipient) of message. The rule of user-project account also holds.*

Subject Matter of the Message

When the project staff first began communicating with research network members in the Spring of 1985, a message was typically very short and its contents related to one subject. By 1990, messages contained as many as six different subjects. Table 4 shows the growth in number of subjects contained in the messages classified by primary subject category for the period April 1985 through August 1990. Included in the SIPP category are data processing decisions made by the Bureau of the Census that significantly affected the construction of public use files and, in particular, data quality. Hardware and remote computing issues are included in the "electronic networking" category.

We see a significant decline in issues related to remote computing. Network access was extended to social scientists by early 1985 at a number of major research institutions, but SIPP ACCESS staff and social scientists had no experience with hardware and communications protocols. Consequently, information about how to use the networks was exchanged, and testing and practice, which could lead to additional communications, took place. After 1986, everyone associated with the SIPP ACCESS research network had become experienced with networking, and we were rarely queried for assistance. Communications about "PSL"(*Physical Sciences Laboratory*) were analysts' informational requests about or problems with administrative matters, including project accounts and rate structure. Analysts corresponded directly with PSL administrators and systems staffs about any questions they had about their projects. We were rarely involved in the discussions, and the small number of questions about PSL would indicate that this aspect of the SIPP ACCESS facility did not create a significant number of problems for analysts.[18]

Table 4. Number of Subjects by Calendar Year, April 1985 -August 1990

	Calendar Year					
	1985	1986	1987	1988	1989	1990
All subjects	19	61	203	474	894	587
database		5	22	22	108	115
electronic network	17	7	16	18	18	1
Ingres RDBMS		5	10	18	57	40
PC-SIPPTEST			11	98	96	8
PSL		4	23	23	41	6
OAS			1	5	9	38
SIPP	1	21	83	181	201	99
SIPP ACCESS	1	15	32	96	222	137
VAX		4	5	13	142	143

We note large increases in two subject areas during 1989 and 1990: the VAX computer and the database. In 1989 and 1990, an overloaded VAX computer system at IRP seriously degraded the environment for analysts. Two-thirds of the communications derived from problems with the IRP VAX computer. Another 25 percent of the traffic related to the VAX came from analysts using the PSL facility who had either forgotten how to use the operating system because they had not used the system in a long time or they had loaded too many tables and filled up the disk where the databases were stored. In 1990, systems changes at PSL impeded access to the databases for several days and accounted for the remaining e-mail about VAX-related matters.

Computational equipment and complex software (the relational database management system) accounted for only a small percentage of communications during any calendar year. This reinforces our observations that the computing environment presented few problems for our analysts.[19] Only in one year did communications about PC-SIPPTEST exceed those about the SIPP database, but this was due almost entirely to communications with our beta sites about testing the PC database, evaluating documentation, and completing a problem set that was part of our evaluation of the PC design and development project.

Overall, the databases, SIPP ACCESS data facility, and the SIPP data accounted for most of the subject matter exchanges between facility analysts and the SIPP ACCESS staff. Communications about SIPP declined from a high of 42 percent in 1987 to a low of 17 percent in 1990. The decline in 1990 can be explained by the fact that in late 1989, network analysts were informed that the SIPP ACCESS project would be terminated, and they were advised to complete their projects by early 1990. Communications related to SIPP ACCESS's information function remained between 20 and 24 percent over the period, except for a decrease to 16 percent in 1987. In 1989 and 1990, a variety of database-related issues accounted for an increase in traffic, both in requests for explanation and assistance and project staff responses. Subjects identified as SIPP ACCESS were almost completely related to our reference provision function. But we received only three inquiries during the entire period from analysts regarding other researchers who were conducting research on the same subject matter as our research network analysts. As might be expected, more than 75 percent of the traffic was accounted for by exchanges between analysts and the SIPP ACCESS staff.

What is the nature of communications about the SIPP panel data? It appears that most of the communications between SIPP ACCESS analysts and project staff concerned the quality of the SIPP data and errors discovered by analysts; the complexity of the SIPP design that made researchers uncertain about how to interpret the data and their output; and definitions of variables that required clarification or were not explained in the Bureau's documentation. A number of communications from analysts described their planned research investigations. The remaining SIPP-related e-mail concerned data processing and coding decisions made by the Bureau and confidentiality issues.

We note that the number of subjects communicated between source and target on the SIPP data averaged between a low of 2.3 in 1985 and a high of 5.2 in 1989, with four of the five years averaging somewhat more than three subjects per message. Most questions could be quickly responded to and most problems could be quickly resolved (this was determined by examining the dates of the query and response). While analysts' questions about the meaning of a particular variable or the logical structure of the SIPP survey or confirmation of an error or oddity in the data usually required only a minimal exchange of messages between the source and sender, we did, however, find examples in the e-mail archive of issues that represented a large investment in analyst, project staff, or Bureau of the Census staff time. For example, after loading the data dictionary for the 1985 panel in a new database, quality and related issues accounted for 52 of the 220 subjects recorded for communications between the Bureau and project staffs during calendar year 1989. In 1990, an unusually extended exchange on one section of a supplementary survey accounted for 23 of the 99 subjects recorded in a series of e-mail exchanges between an analyst and SIPP ACCESS project staff. Discoveries made by the analyst could not be explained by the SIPP ACCESS project staff, and the analyst decided to call the Bureau to find out whether the staff could confirm his discoveries.

Improving the Quality of Data

We looked for evidence that electronic communications provided a vehicle for improving data quality. We examined the responses by the Bureau of the Census staff to SIPP problems identified by analysts and SIPP ACCESS project staff and the responses by SIPP ACCESS project staff to database problems identified by analysts or project staff after data became publicly available in the database. We then obtained information on the outcome of the exchange in terms of whether it led to an *immediate* improvement in the quality of data.

We actually profile fewer problems than were uncovered because we exclude two types of communications. The first pertains to a preliminary analysis of the SIPP data prior to their loading and availability in the public database, when, as we have already noted, a great number of data quality problems were first diagnosed and subsequently corrected by the Bureau of the Census. The second pertains to the data structure of longitudinal panel surveys or general problems of survey design, processing or data file design, but about which nothing could be done short of modifying the original SIPP design or modifying the processing system created by the Bureau of the Census, although, as we saw earlier, both design and the processing system created significant problems for analysts. Design issues identified by analysts were nevertheless serious, and their discoveries indicated problems of SIPP data quality.[20] But identifying these problems would not necessarily lead to immediate improvements in data quality during the life of the SIPP ACCESS project. It would therefore be unrealistic to use the outcome of modifying the SIPP design or the infrastructure of the Bureau of the Census as the test for whether improvements in data quality took place.

Instead, our test for whether communication exchanges between analysts and project staff and the Bureau of the Census staff led to improvements in SIPP data quality is much more modest. We use a communication exchange between the project staff and analyst or project staff and Bureau of the Census about a data problem and then examine whether an identified problem was later corrected. If corrected, we conclude that communications contributed to improving the quality of SIPP data.

Table 5 lists by year SIPP data problems that were identified by either the SIPP ACCESS project staff or analysts after the data became publicly available in the SIPP database at the Physical Sciences Laboratory (PSL). A total of thirty-nine problems were identified in the e-mail archive between January 1986 when the relational database containing three interviews became available and June 1990 when the database was officially closed to public access. Twenty-eight of the thirty-nine problems were identified by the SIPP ACCESS staff as a result of their own work. The remaining eleven problems were identified by analysts after they began to extract data for their research problem.

The majority of the identified problems led to the release of new data files by the Bureau within several months. In some data files more problems were revealed, and a second or third version of the data file was subsequently produced. Five problems led SIPP ACCESS to design and redesign database tables in order to eliminate difficulties that we anticipated would be encountered by future analysts. After SIPP ACCESS designed the tables, we received no more communications on the problems that either we or the analyst had earlier identified. This should not be construed to mean that the problem disappeared. Rather it indicates that restructuring the original public use data clarified their meaning, and it was then easier for the analyst to interpret the original data. By applying our accumulated experience with the data, we were able to reconcile oddities or inconsistencies in the original public data files.

Table 6 lists a total of twelve problems over the life of the project that were related to the database created by the SIPP ACCESS staff. The number of problems that derive from work performed by the project staff is surprisingly small.[21] Four problems were identified by the project staff itself, and eight problems were located by analysts after they began their research. The dates of the e-mail communications indicate that we corrected the error within a day or two.

PART FOUR: DISCUSSION AND RECOMMENDATIONS

The 1980s were a period of exponential growth in scientific networks, but our literature review

Table 5. Identification of SIPP Data Problems and Outcomes, Calendar Year 1986 – August 1990

Year	Problem	Identified by	Outcome
1986	• Designated parent code inaccuracies; not all parents & chldren linked	ANALYST	SIPP ACCESS creates parent-child table
	• Entry and exit variables cannot be interpreted	SA	SIPP ACCESS creates database tables
	• Errors identified in 1st interview data file and dictionary	SA	Bureau issues new release
	• Errors identified in 2nd interview data file	SA	Bureau issues new release
	• Errors identified in 3rd interview data file and dictionary	SA	Bureau issues new release
	• Errors identified in 5th interview data file and dictionary	SA	Bureau issues new release
1987	• Error in topcoding for birth date in 2nd interview data file	SA	Bureau issues new release
	• Unknown meaning of imputations & constructed variables	SA	Bureau cannot provide explanation
	• Error in calculating # weeks of employment	SA	Bureau issues technical memo
	• Error in 4th interview supplement data	SA	Bureau issues new release
	• Errors in parent-child linkage variables	ANALYST	SIPP ACCESS creates parent-child table for Phase 2 database
	• Missing poverty threshold variables in 9th interview data	ANALYST	Bureau issues new release
	• Error in recode information for 4th interview supplement	SA	Bureau issues new information
	• Marital status discrepancies	ANALYST	Bureau edits marital status for time-series file
	• Database yields different control counts from Bureau	SA	Bureau corrects its counts

SA = SIPP ACCESS identified problem
ANALYST = analyst identified problem

Table 5. *Continued*

Year	Problem	Identified by	Outcome
1988	• time-series data dictionary error	SA	Bureau issues new release
	• time-series data dictionary error	SA	Bureau issues new release
	• time-series data file error	SA	Bureau issues new release
	• time-series data file error	SA	Bureau issues new release
	• time-series data file error	SA	Bureau issues new release
	• time-series and cross-sectional control counts differ	SA	No explanation
	• Unidentified persons found in time-series data	SA	No explanation
	• New sample units found in time-series data	SA	No explanation
	• Dead people not all properly identified	SA	SIPP ACCESS creates table for Phase 2 database
	• Family record counts error 9th interview	SA	Bureau issues new release
	• Missing households in 8th interview supplement	SA	Bureau confirms errors, but does not reissue file
	• Errors in interview status	SA	SIPP ACCESS creates new table for Phase 2 database
	• New errors 4th interview supplement	ANALYST	Bureau reissues file
	• Documentation and data differ in 5th interview supplement	SA	Bureau issues technical memoranda, new release
	• More errors in 9th interview	SA	Bureau issues re-release
	• Undocumented imputation flags 3rd interview supplement	SA	Bureau provides information
1989	• AFDC income too high	ANALYST	Bureau acknowledges known problem, but cannot correct
	• Error in coding of person identifier for disabled children	ANALYST	Bureau acknowledges unadvertised problems, issues user note
	• Differences between constructed and original variables ("other" income)	SA	Bureau acknowledges, but cannot explain
	• 1985 data dictionary errors	SA	Bureau acknowledges, but cannot correct because original file destroyed
1990	• errors in 4th & 7th interview supplements	ANALYST	Outcome unknown
	• calculation flags (income) construction unidentified	ANALYST	Bureau unable to explain meaning
	• 4th interview supplement missing explanation in data dictionary	ANALYST	Outcome unknown
	• 2 marital status changes in one reference period	ANALYST	Not communicated to Bureau

SA = SIPP ACCESS identified problem

ANALYST = analyst identified problem

showed that we had an inadequate knowledge base to determine whether and in what ways electronic networks had furthered developments in science and technology and contributed to scientific productivity. The objective of this article on electronic communications of the SIPP ACCESS research network was to contribute empirical data about how electronic networks advance the scientific enterprise. We suggested that claims of economic benefits derived from networks could not be supported because of measurement problems associated with the concept of scientific productivity. As such, we examined the intermediate steps in the research activity—the process between inputs and outputs (outcomes) to understand how social scientists and the data producer relied on electronic research networks. We profiled their communications about complex and dynamic statistical data, how often they communicated, and the content of their communications with the SIPP ACCESS project staff.

The first part of these concluding remarks summarizes the direct and indirect benefits of electronic networks for the social science enterprise. The second section discusses the implications of our findings in light of the policy initiatives of the High-Performance Computing and Communications Act of 1991 (P.L. 102–194) and the policy agenda set forth in *Grand Challenges 1993* (1992).

Summary of Findings

We found confirmation for the finding by McClure and associates that electronic networks made an important contribution to understanding and using

Table 6. Identification of Database Problems and Outcomes, Calendar Year 1988–August 1990

Year	Problem	Identified by	Outcome
1987	• Children in food stamp unit deleted in program coverage table	ANALYST	SIPP ACCESS corrects and reissues table
1988	• Table identifying movers does not identify all dead people	SA	SIPP ACCESS designs new Phase 2 tables to identify all dead people, new entrants, and exits from the panel
	• Table of marital status does not identify immediate spouse changes	ANALYST	SIPP ACCESS designs new Phase 2 table for marital spouse changes
	• SIPPEST contains too few cases for small populations	ANALYST	SIPP ACCESS writes user applications program to select sample, but creates no other test database
1989	• Errors in 5th interview supplement child care segment identified	ANALYST	SIPP ACCESS corrects table
	• Phase 2 version of interview status table not installed in PSL database	ANALYST	SIPP ACCESS installs new table
	• Metadata table unreadable	SA	SIPP ACCESS corrects table
	• Interview status table errors (several cases)	SA	SIPP ACCESS corrects table
	• Demographic table errors (several cases)	SA	SIPP ACCESS corrects table
1990	• State unemployment table contains logical error; states do not correspond to sampling frame	ANALYST	SIPP ACCESS corrects table and metadata explanation
	• Errors in 3 time-series tables	ANALYST	SIPP ACCESS corrects tables
	• OAS maintains bad version of a time-series data table	ANALYST	SIPP ACCESS archives corrected table on OAS

A = SIPP ACCESS identified problem

ANALYST = analyst identified problem

data. Although our discussion explicitly highlighted only one noneconomic indicator of performance-effectiveness, improvements in data quality, our profile of electronic communications offered both direct and indirect evidence of other indicators of performance of the electronic network, including efficiency and effectiveness. This evidence includes resource sharing; knowledge circulated to all participants in an information system; and assistance in problems related to the networks, computers, software and the scientific design of SIPP.

The SIPP ACCESS central repository served as a vital communications link for social scientists who were geographically dispersed throughout the United States, Europe, and Australia. This capability could not have occurred outside an electronic network system.

Knowledge of the SIPP data was centrally collated and then analyzed and incorporated into the information system by the SIPP ACCESS project staff. Information obtained from the Bureau's liaison, who was a member of the research network, became part of the knowledge resource of the data facility. The electronic network made it possible to archive discoveries about the data at a central data facility permanently and to make them immediately accessible to everyone.

There were ongoing communication exchanges among analysts, expert, and data producer. Intensive use of the data facility brought more communications with the staff, and the staff communicated more often with the data producer. Through the network, answers to technical and substantive research questions were immediately available from the project staff at SIPP ACCESS, the liaison at the Bureau of the Census.

Analysts' discoveries about data quality could be quickly verified by project staff and communicated to the data producer. Feedback improved the quality of information and increased the stock of knowledge about the SIPP data. Improved data quality, as manifested in corrected SIPP data released by the Bureau, resulted from electronic communications. The rapid discovery, diagnosis, and correction of error could not have taken place without the electronic network.

Individual social scientists did not each obtain and process the SIPP data files of more than 2.2 gigabytes of data. Only one copy of the data resided at a central location, at a reduction in cost to the social science enterprise of perhaps tens of thousands of dollars. The electronic networks made it possible for researchers to access data that were unavailable or too expensive to purchase or process efficiently at their home site. The electronic networks also made it possible to use computational resources that were unavailable at their home institution.

In sum, we found that the LAN and wide area networks supported a cooperative work environment. And more work could be completed than by analysts working alone: from our library of publications and working papers on the SIPP, we know that two-thirds of the corpus of research on SIPP between 1985 and June 1990 was carried out by members of the SIPP ACCESS research network.

On the other hand, our data did not show that the SIPP ACCESS was a catalyst for collaborations with new colleagues. We had few requests to identify others in the network who were conducting research along similar lines. Perhaps the SIPP ACCESS research network was a special case in network use, and there were too few researchers nationwide conducting research on SIPP to confirm that the electronic network created an opportunity to collaborate with new colleagues.[22] It is quite possible, however, that new intellectual networks take much longer to establish than the short period of time that SIPP ACCESS was in existence.

We also found that researchers faced technical problems in using the central data facility, particularly network and hardware problems at local sites. However, technical problems with the wide area electronic networks, with the exception of transferring large quantities of data, disappeared for the most part after the networks became more reliable and social scientists became more experienced with the technical aspects or had access to network support staff.

The electronic communications also revealed other, less direct benefits of the research network. First, the communications archive maintained by SIPP ACCESS documented severe problems in the infrastructure of the Bureau of the Census. Although we gave a high rating to the Bureau's responsiveness to the user community's identification of problems, our data show that the Bureau was consistently unable to locate written information about or explain processing decisions. These communications also revealed that the Bureau had a significant organizational problem: it poorly preserved an institutional memory of its own data collection project. Furthermore, the designing of new database tables to ensure that research would not be further constrained by the data structure of the public use files provided additional documentation of problems in the data processing environment of this government agency. This evidence was later used as part of an evaluation of the SIPP by the National Research Council's Committee on National Statistics Subcommittee on the SIPP, when the Committee examined the quality of the Bureau's infrastructure and implications for efficient and effective data production and quality control (Olsen et al., 1991).

Second, the relatively small number of problems identified by analysts (contrasted with the very large number of problems discovered by the project staff prior to loading the database) suggests that the SIPP ACCESS staff played an important role as a gatekeeper for data quality. It also suggests that the expert as gatekeeper in a research network contributes important efficiencies in the science infrastructure. Relatedly, the burden placed on the government agency to provide expertise and technical assistance and to verify discoveries in the data was reduced because the SIPP ACCESS project staff played a critical role as consultant.

Third, although existing patterns of scientific activity by elites and their collaborators were reinforced (as shown by other data in the SIPP ACCESS archive), the project demonstrated that an electronic network can effectively extend opportunities to a new group of researchers. About a quarter of the research output associated with the SIPP ACCESS project derived from researchers at institutions that do not have a history of poverty research. Like Hesse and his colleagues, we found that access to networks is important for investigators at the "periphery" of centers of research.

Fourth, the project demonstrated that problem-solving can be effectively carried out in a research network. A complex scientific design, large set of data, and complicated technologies were mastered without face-to-face or telephone communication. Nearly all our communications with the SIPP ACCESS analysts were conducted through the electronic network, and only occasionally were traditional media used even by analysts not affiliated with the SIPP ACCESS network. Thus, although we have no alternative source of empirical evidence to demonstrate that social scientists have substituted CMC for traditional communications media, by comparing the number of telephone, post, and electronic communications received over the life of the SIPP ACCESS project, we may infer that research network analysts substituted electronic for the traditional communications media. This supports findings from thirteen other studies that found "decreases in written, telephone, travel, and some face-to-face communication after a CMC was implemented" (Rice, 1987, p. 79; see also Rice & Shook, 1988). Our data also corroborate evidence obtained in a study of a private sector firm by Rice and Bair (cited by Rice et al., 1990, p. 31), who found that electronic messaging systems were more likely to be substituted for face-to-face communication when the information could be analyzed (e.g., facts, data).

Recommendations

What policy recommendations can be made based on the evidence that we have an inadequate knowledge base for evaluating how electronic networks advance the scientific enterprise and on the SIPP ACCESS research network experience? These concluding remarks address (1) the need for good data to describe and evaluate the networks' contribution to scientific activity, (2) the importance of expanding expertise required for the design of high-performance computing and communication technologies initiatives to include statisticians and social and behavioral scientists, and (3) the need to evaluate the current allocation of resources for developing the National Educational and Research Network (NREN) infrastructure. These remarks are not, however, intended to represent a coherent agenda, but rather to highlight lacunae in the policy initiatives and suggest that we reconsider the emphasis of our national policy initiatives in high-performance computing and communications.

First, we need to increase our stock of knowledge about the processes and outputs of scientific activity supported by the electronic networks. Simply put, we need to know what we are doing in order to know where we should go. The High-Performance Computing Act of 1991 (HPCA) (P.L. 102-194) and *Grand Challenges 1993: High Performance Computing and Communications* (1992) make no mention of the need to gather data about electronic networks; consequently, no resources are allocated for recordkeeping and evaluation. However, advocates of the importance of electronic networks must devote far more attention to data collection than they have in order to justify continuing or increased levels of funding support. Furthermore, without empirical data on how and in what ways electronic networks are used and the outputs that derive from their use, we will be hard pressed to demonstrate that high-performance computing and communications initiatives have the expected payoff of advancing the scientific enterprise.

A variety of empirical data must be collected to reflect the different properties of the system. For example, at a system level, we need good data on the development and use of the Internet, and we need to institutionalize recordkeeping as part of the Internet's monitoring and maintenance. Lottor's (1992) report on Internet growth between 1981 and 1991 shows the sorry state of recordkeeping. He points out that complete data could not be collected because of a variety of technical problems related to software and hardware, the amount of data collected, procedural problems in identifying server or host sites and transferring information from a site, poor quality and missing data, and costs of downloading information. Lottor concludes his assessment of statistical information by recommending new software "to handle the enormous amount of data collected and expected in the future" and alternate procedures for archiving the data (pp. 7–8). Good data are not just the concern of the Internet systems staff, however. We need more participants in the development of network statistics because of the diversity of stakeholders in the network system. In addition, Lottor's recommendations should be preceded by a scientific design for data collection and evaluation. And the expertise of sampling statisticians should be applied to the statistics-gathering efforts to determine cost-effective ways of data collection.

We need to understand and evaluate the settings in which scientific use of electronic networks takes place. Comparing the amount of research on the scientific enterprise done from the 1960s to the middle 1970s with the significantly reduced output of the next fifteen years, it appears that, with few exceptions, we have given inadequate attention to studying the social organization of science, bibliometric work excluded. Renewed attention and resources should be dedicated to the sociology of science, so that we can achieve a better understanding of the relationship between information technology and the structures and processes of scientific and technical organizations and work groups.

We need to understand and evaluate the cognitive and behavioral aspects of information technology use. Again, *Grand Challenges* is silent about the human aspects of technology initiatives, emphasizing, instead, that barriers to acceptance of new technologies include "high initial cost, inadequate and user-unfriendly software, and lack of standards" (p. 11). Neglected in the policy discussion is how competitiveness and productivity depend on the intellectual content of what is processed by the computer and transmitted across the electronic networks and the methods by which knowledge is transferred. Our project found that prior training and experience in a number of knowledge domains and expert assistance were essential prerequisites for successful use of the electronic network environment (see also Tennant, 1991), but were either unsupported or inadequately supported by the national research and educational establishment. The laudatory goal of improving the capability to access information and databases appears irrelevant if students and scientists are not trained to use information and databases and to make informed judgments about their quality. Successful transfer of technology depends, we would argue, on good scientific and technical knowledge and training. As such, attention must be equally devoted to pedagogy and learning, as well as technology.

Second, we need to rethink the allocation of resources for developing the infrastructure for high-performance computing and communications. The emphasis of national policy initiatives has been on the physical infrastructure of national and international networks. However, the SIPP ACCESS project found severe problems in the infrastructure of the analyst's *home* institution, including primitive telecommunications and computational facilities at both the university and departmental levels which impeded or entirely prevented access to the SIPP ACCESS data facility. We also found that resources for building local infrastructure were differentially distributed. Although 131 people attended the training workshops, only 16 percent went on to use the complete panel database; graduate students and scholars from small institutions were unable to profit from a nationally supported experiment funded by the National Science Foundation. The majority of the analysts who used SIPP ACCESS and subsequently produced the most research output were members of elite research institutions with a well-developed scientific infrastructure.

Another emphasis of the policy initiatives is the provision of technological resources for a new generation of engineers and computer scientists. The initiatives exclude support for the social and behavioral sciences. But our project found that social and behavioral scientists lacked adequate resources for computation, research, and access to information and data. The benefits of research networks appeared to be differentially allocated, even at major research institutions, with physical, biological, and computer scientists profiting from better access to a wide array of intellectual and other resources.

The policy initiatives emphasize improvements in access to information and data resources, but both HPCA and *Grand Challenges* are silent with regard to the development of high-quality information and data and the expertise to maintain them and provide assistance in their use. Yet it is the content and quality of what is being transmitted through the high-capacity networks that will ultimately determine what scientific advances take place. As our project found, without the human resources to maintain, evaluate, and provide assistance in the use of the information and data, these resources will be unused, underutilized, or improperly used (see also National Research Council, 1988). The comments of Cinkosky and his colleagues (Cinkosky et al., 1991, p. 252) about their work on Genbank, the national repository for nucleotide sequence data, reinforce our concerns about this omission in the policy initiatives. They note that:

> Our experience has consistently indicated that the largest source of errors in nucleotide sequence data is transcription errors in the creation of figures for printed publication. Further, most journals and their reviewers tend to concentrate on the quality of the scientific reasoning embodied in the paper, typically leaving the data essentially unreviewed. We, on the other hand, specialize in the evaluation and analysis of nucleotide sequence data; the (increasingly automatic) checks that we perform on sequence data are far more extensive than most reviewers would have the time to perform.

The High-Performance Computing Act of 1991 and *Grand Challenges 1993: High Performance Computing and Communications* set out an important and ambitious agenda to improve the nation's infrastructure for computation and communication through technical and engineering efforts. However, the policy should be evaluated in view of the need for empirical data on the properties of the infrastructure. The policy should also be evaluated for its emphasis on the technical aspects of national infrastructure development. Specifically, we first need to acknowledge the importance of empirical data for evaluating the proposed initiatives. Second, we need to recognize that improvements in infrastructure development derive from activity that takes place in social organizations. Social and behavioral scientists can contribute both to the scientific design of data collection and the study of the human properties of the infrastructure. As the physicist Philip Morrison has stated, "The enterprise of science leads directly to the necessity of the social and behavioral sciences" (Adams et al., 1982, p. vi). Third, policy initiatives must address support for disciplines other than engineering and computer science. Fourth, the information component of the infrastructure must receive more support, including the generating, maintaining, and enhancing of information and data resources. Fifth, we need to direct and increase resources to local infrastructures if we are to create a viable national research and educational network.

Notes

1. References are intended to be illustrative, not exhaustive of the literature on the subjects. Our emphasis is on empirical research and not theorizing. Thus, although a great deal has been written on how computer technology is altering the structure of organizations, references to this literature are not included if data analysis was not reported by an author. For a synthesis of the extensive literature, see McClure et al., 1991.

2. See, for example, Allen, 1970; Bavelas, 1950; Braam et al., 1991; Crane, 1972; Erbadi & Utterback, 1984; Fischer, 1973; Hagstrom, 1970; Hargens 1975; Hummon & Doreian, 1989; Price, 1970; Utterback, 1971.

3. Alice Robbin and Martin David (1992) were co-principal investigators and co-directors of the SIPP ACCESS project. Between Spring 1984 and December 1991, the SIPP ACCESS project at the University of Wisconsin-Madison was supported in part by the National Science Foundation (SES-8411785, SES-8921213, and SES-8701911), the Sloan Foundation (B1984-25 and B1987-46), the Social Science Research Council, the Bureau of the Census (through the National Science Foundation), and the University of Wisconsin-Madison. Administrative support was provided by the Institute for Research on Poverty and Center for Demography and Ecology (P30 HD05875). We are indebted to our first NSF program manager Murray Aborn for his wisdom, foresight, and nurturing of this project in the years before the NREN was created.

4. On the need for systematic thinking by members of the information system about the processes involved in the data delivery system and creating new knowledge, see David, 1980, 1985, 1991; and Robbin, 1981a, 1981b, 1983, 1986.

5. We exclude from our discussion the long delay in preparing an extract of data, which also impedes user understanding of data.

6. The project director informally communicated this finding to the author.

7. A host of other methodological issues are identified with the study of organizational communication which are not addressed here (for one discussion, see Monge et al., 1984).

8. Other noneconomic indicators of the "output of science" utilized by David and Robbin (1992) included data products, client use of the facility, publication products, and conceptual and methodological applications. These outputs are the subject of forthcoming papers on measuring the performance of an information system for complex data.

9. For the role of gatekeeper, see, for example, Allen, 1970; Katz & Tushman, 1979; Tushman, 1977; and Tushman & Katz, 1980.

10. When it became possible to move data archived on the Optical Archive Store (OAS) at PSL reliably over the LAN, UW campus researchers no longer needed to establish an account at PSL, but could carry out database work at the Institute for Research on Poverty (IRP). As such, the project maintained databases at both PSL and IRP, although it had never been envisioned that the IRP development databases would serve IRP researchers. At that time, distributed database management systems were still in the design and testing stages.

11. By "project administration," we mean those messages about, for example, project planning, proposal writing, funding support, training workshops, and communicating with our advisory board and UW institutional personnel. Also excluded is e-mail that provides a detailed historical record of purchasing, archiving, and preliminary analysis and documentation of findings for the SIPP public use files. A significant number of exchanges occurred because of many problems with the SIPP data that surfaced from the time an order was placed to purchase them from the Bureau of the Census and when they were actually loaded into the relational database management system.

12. We actually maintained e-mail through December 1991, but the period September 1990 through December 1991 is not included in this present discussion.

13. See Chapter Four in David & Robbin (1992) for a full discussion of sampling and coding decisions, data quality, and findings.

14. An important innovation in data user services took place within the Bureau of the Census after we had submitted our proposal to the National Science Foundation in August 1984. The Bureau designated a SIPP staff member as liaison to the research community and liaison to the SIPP ACCESS project. The Bureau's decision immediately altered existing conditions in the data delivery system that had, in part, led to our proposal to the National Science Foundation. One consequence was that SIPP ACCESS would not be the focal institution for information provision and user assistance; it would share this responsibility with the Bureau. A second consequence was that the completeness of recordkeeping on communications about SIPP depended on recordkeeping inside the Bureau. If the liaison at the Bureau of the Census failed to record information about SIPP ACCESS researchers, we would have an incomplete picture of communications flows in the SIPP ACCESS information system.

15. The author was responsible for the coding.

16. Initially, researchers located at institutions outside the University of Wisconsin-Madison connected to PSL via remote log-on and modem. The Internet did not become available until 1989. BITNET with computer-to-computer e-mail was not available to most of our analysts as late as 1987. A very small amount of the traffic recorded as originating from the LAN actually derived from use of the Internet after early 1989.

17. SIPPTEST was a test database that contained 2 percent of the sample units (and associated entities) and more than 20,000 variables, resided online 24 hours a day, and was used interactively to learn about the SIPP data and the Ingres RDBMS to test hypotheses and to create and debug command files. SIPPRUN was the complete panel database of more than seven million observations and all the variables. SIPPRUN tables were stored offline on the Optical Archive Store (OAS) and retrieved independently by analysts upon demand; work in the SIPPRUN database was carried out in batch mode. PC-SIPPTEST was a microcomputer version of the SIPPTEST database.

18. On the other hand, we cannot be absolutely certain that project account administration created few or no problems for most researchers or that communication between the PSL staff and analysts was minimal. The PSL staff and analysts circulated e-mail to us only when there was a problem. Nevertheless, we heard few complaints from researchers and therefore conclude that researchers encountered few problems using the PSL computer facility.

19. We exclude from this analysis communications about the utility of the RDBMS environment for work unrelated to the SIPP data. Nevertheless, a complete analysis of project communications shows that "spillover" effects occurred as a result of the innovation diffusion (see David & Robbin, 1992, Volume One, Chapter Five).

20. These included, for example, high rates of missing data; imputation bias for missing data; discrepancies in enumeration of membership in program units; greater between-wave than within-wave changes; left and right censoring; inconsistencies in personal identifiers; poor retrospective data in supplements to the core data; large attrition in the low-income and minority population; skip patterns that created confusion about or inability to select a correct population; sample reductions that created too few cases for longitudinal research; and unavailable tax data due to confidentiality restrictions. A design issue that surfaced repeatedly throughout the history of the SIPP ACCESS project was the staggered interviewing schedule, which created significant problems in linking cases for studying change.

21. On the basis of our prior experience with data, we would have expected many more mistakes. We do expect, however, that more problems will be discovered through more extensive use of the data by the research community.

22. Another explanation is that the poverty researchers were also served by two other data distributors: the Bureau of the Census and the Inter-university Consortium for Political and Social Research. As such, they could have sought information about research in progress from other sources.

References

Adams, R. M., Smelser, N. J., & Treiman, D. J. (Eds.). (1982). *Behavioral and social science research: A national resource.* Part I. Washington, DC: National Academy Press.

Allen, T. J. (1970). Roles in technical communication networks. In C. E. Neson & D. K. Pollack (Eds.), *Communication among scientists and engineers* (pp. 191–208). Lexington, MA: Heath Lexington Books.

American Institute of Physics. (1990). Progress on AIP Study of Research Collaborations. *AIP History Newsletter, 22*(2), 1–2.

Andrews, F. M. (Ed.). (1979). *Scientific productivity: The effectiveness of research groups in six countries.* New York: Cambridge University Press.

Archer, N. P. (1990). A comparison of computer conferences with face-to-face meetings for small group business decisions. *Behavior & Information Technology, 9*(4), 307–317.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *Journal of the Acoustical Society of America, 22*(6), 725–730.

Bikson, T. K., & Eveland, J. D. (1990). The interplay of work group structures and computer support. In J. Galagher, R. E. Kraut, & C. Egido, C. (Eds.), *Intellectual teamwork: Social and technological foundations of cooperative work.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Braam, R. R., Moed, H. F., & van Raan, A. F. J. (1991). Mapping of science by combined co-citation and word analysis, I. Structural aspects. *Journal of the American Society for Information Science, 42,* 233–251.

Cinkosky, M. J., Fickett, J. W., Gilna, P., & Burks, C. (1991, May 31). Electronic data publishing and Genbank. *Science, 252,* 1273–1277.

Clark, N. (1986). Tables and graphs as language. *Proceedings of the 18th Symposium on the Interface* (pp. 83–89). Fort Collins, CO. Washington, DC: American Statistical Association.

Comer, D. (1983). The computer science research network CSNET: A history and status report. *Communications of the ACM, 26,* 747–753.

Crane, D. (1972). *Invisible colleges.* Chicago, IL: University of Chicago Press.

Danowski, J. (1982). Computer-mediated communication: A network-based content analysis using a CBBS conference. *Communication Yearbook, 6,* 905–924. Newbury Park, CA: Sage Publications, Inc.

Danowski, J. (1988). Organizational infographics and automated auditing: Using computers to unobtrusively gather as well as analyze communication. In G. M. Goldhaber & G. A. Barnett (Eds.), *Handbook of organizational communication* (pp. 385–434). Norwood, NJ: Ablex Publishing Corporation.

David, M. H. (1980). Access to data: The frustration and utopia of the researcher. *Review of Public Data Use, 8*(4), 327–337.

David, M. H. (1985). *The language of panel data and lacunae in communication about panel data* (CDE Working Paper 85–20). Madison, WI: University of Wisconsin Center for Demography and Ecology.

David, M. H. (1991). The science of data sharing: Documentation. In J. Sieber (Ed.), *Sharing social science data: Advantages and challenges* (pp. 91–115). Newbury Park, CA: Sage Publications, Inc.

David, M. H., & Robbin, A. (1981). The great rift: Gaps between administrative records and knowledge created through secondary analysis. *Review of Public Data Use, 9*, 153–166.

David, M. H., & Robbin, A. (1992). *Building new infrastructures for the social science enterprise. Final report to the National Science Foundation on the SIPP ACCESS project.* Madison, WI: University of Wisconsin.

Dolby, J., Clark, N., & Rogers, W. H. (1986). The language of data: A general theory of data. *Proceedings of the 18th Symposium on the Interface* (pp. 96–103). Fort Collins, CO. Washington, DC: American Statistical Association.

Erbadi, Y. M., & Utterback, J. M. (1984). The effects on communication on technological innovation. *Management Science, 30*(5), 572–585.

Fischer, C. (1973). Some social characteristics of mathematicians and their work. *American Journal of Sociology, 78*, 1094–1118.

Galagher, J., Kraut, R. E., & Egido, C. (Eds.). (1990). *Intellectual teamwork: Social and technological foundations of cooperative work.* Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Gerola, H., & Gomory, R. E. (1984, June 6). Computers in science and technology: Early indications. *Science, 225*, 11–18.

Gould, S. (1990, January 11). *High performance computing: An overview.* (CRS Issue Brief). Washington, DC: Congressional Research Service, Library of Congress.

Grand challenges 1993: High performance computing and communications, The FY 1993 U.S. research and development program [to supplement the President's Fiscal Year 1993 Budget]. (1992). Washington, DC: Committee on Physical, Mathematical, and Engineering Sciences; Federal Coordinating Council for Science, Engineering, and Technology; and the Office of Science and Technology Policy.

Grote, G., & Baitsch, C. (1991). Reciprocal effects between organizational culture and the implementation of an office communication system: A case study. *Behavior & Information Technology, 10*(3), 207–218.

Hagstrom, W. O. (1965). *The scientific community.* New York: Basic Books, Inc.

Hagstrom, W. O. (1970). Factors related to the use of different modes of publishing research in four scientific fields. In C. E. Neson & D. K. Pollack (Eds.), *Communication among scientists and engineers.* Lexington, MA: Heath Lexington Books.

Hargens, L. L. (1975). *Patterns of scientific research: A comparative analysis of research in three scientific fields.* Washington, DC: American Sociological Association.

Hesse, B. W., Sproull, L., Kiesler, S., & Walsh, J. P. (in press). Returns to science: Networks and scientific research in oceanography. *Communications of the ACM.*

Hiltz, S. R. (1984). *Online communities.* Norwood, NJ: Ablex Publishing Corporation.

Hiltz, S. R., & Johnson, K. (1990). User satisfaction with computer-mediated communication systems. *Management Science, 36*(6), 739–764.

Hiltz, S. R., Scher, J., & Turoff, M. (1978). Computerized conferencing in problem solving and decision making. *The information age in perspective. Proceedings of the 39th ASIS Annual Meeting, 15*, 159–161.

Hiltz, S. R., & Turoff, M. (1985, July). Structuring computer-mediated communication systems to avoid information overload. *Communications of the ACM, 23*, 680–689.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*, 39–63.

Jennings, D. M., Landweber, L. H., Fuchs, I. H., Farber, D. J., & Adrion, W. R. (1986, February 28). Computer networking for scientists. *Science, 231*, 943-950.

Johansen, R., DeGrasse, R., & Wilson, T. (1978, February). *Group communication through computers, Vol. 5, Effects on working patterns.* (Report R-41). Menlo Park, CA: Institute for the Future.

Katz, R., & Tushman, M. (1979). Communication patterns, project performance, and task characteristics: An empirical evaluation and integration in an R&D setting. *Organizational Behavior and Human Performance*, 23, 139-162.

Kerr, E. B., & Hiltz, S. R. (1981). *Computer-mediated communication systems: Status and evaluation*. New York: Academic Press.

Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, *39*, 1123–1134.

Lea, M. (1991a). Computer-mediated communication, de-individuation and group decision-making. *International Journal of Man-Machine Studies*, *34*, 283–301.

Lea, M. (1991b). Rationalist assumptions in cross-media comparisons of computer-mediated communication. *Behavior & Information Technology*, *10*(2), 153–172.

Lottor, M. K. (1992, January). *Internet growth (1981 – 1991)*. (Memorandum to the Network Working Group). Menlo Park, CA: SRI International.

Markoff, J. (1990, September 2). Creating a giant computer highway. *The New York Times*, Sec. 3, pp. 1, 6.

Markoff, J. (1992, January 1). A network of networks that keep scientists plugged in. *The New York Times*, p. 21.

McClure, C. R., Bishop, A., & Doty, P. (1990, January 30). *Electronic networks, the research process, and scholarly communication: An empirical study with policy recommendations for the National Research and Education Network*. Unpublished manuscript, Syracuse University, School of Information Studies.

McClure, C. R., Bishop, A., Doty, P., & Rosenbaum, H. (1991). *The National Research and Education Network (NREN): Research and policy perspectives*. Norwood, NJ: Ablex Publishing Corporation.

Monge, P. R., Farace, R. V., Eisenberg, E. M., Miller, K. I., & White, L. L. (1984). The process of studying process in organizational communication. *Journal of Communication*, *34*, 22–43.

National Research Council, Committee on Basic Research in the Behavioral and Social Sciences, Commission on Behavioral and Social Sciences and Education. (1988). *The behavioral and social sciences: Achievements and opportunities*. Washington, DC: National Academy Press.

National Research Council, Panel on Information Technology and the Conduct of Research, Committee on Science, Engineering, and Public Policy. (1989). *Information technology and the conduct of research: The user's view*. Washington, DC: National Academy Press.

Olsen, R. J., David, M. H., & Sheets, C. T. (1991, May). *The operation of SIPP and implications for the future*. Draft. Washington, DC: National Research Council, Committee on National Statistics, Subcommittee on SIPP.

Olson, M. H., & Bly, S. A. (1991). The Portland experience: A report on a distributed research group. *International Journal of Man-Machine Studies*, *34*, 211–228.

Pappa, M. J. (1990). Communication network patterns and employee performance with new technology. *Communication Research*, *17*, 344–338.

Pappa, M. J., & Tracy, K. (1988). Communicative indices of employee performance with new technology. *Communication Research*, *15*(5), 524–544.

Price, D. J. deSolla. (1970). Citation measures of hard science, soft science technology, and non-science. In C. E. Nelson & D. N. Pollock (Eds.), *Communication among scientists and engineers* (pp. 3–22). Lexington, MA: D.C. Heath.

Rice, R. E. (1982). Communication networking in computer-conferencing systems: A longitudinal study of group roles and system structure. *Communication Yearbook 6* (pp. 925–944). Beverly Hills, CA: Sage Publications, Inc.

Rice, R. E. (1987). Computer-mediated communication and organizational innovation. *Journal of Communications*, *37*, 65–94.

Rice, R. E., & Associates. (Eds.) (1984). *The new media: Communication, research and technology*. Newbury Park, CA: Sage.

Rice, R. E., & Case, D. (1983). Electronic message systems in the university: A description of use and utility. *Journal of Communications*, *33*, 131–152.

Rice, R. E., Grant, A, E., Schmitz, J., & Torobin, J. (1990). Individual and network influences on the adoption and perceived outcomes of electronic messaging. *Social Networks*, *12*, 27–55.

Rice, R. E., & Love, G. (1987). Electronic emotion: Socioemotional content in a computer-mediated communication network. *Communication Research*, *14*, 85–208.

Rice, R. E., & Shook, D. E. (1988). Access, usage of, and outcomes from an electronic messaging system. *ACM Transactions on Office Information Systems*, *6*, 255–276.

Rice, R. E., & Torobin, J. (1986). Expectations about the impacts of electronic messaging. *Proceedings of the 49th ASIS Annual Meeting, 23,* 271–276.

Robbin, A. (1981a). Technical guidelines for preparing and documenting Data. In R. F. Boruch, P. Wortman, & D. Cordray (Eds.), *Reanalyzing program evaluation: Policies and practices for secondary analysis of social and education programs* (pp. 84–143). San Francisco: Jossey-Bass, Inc.

Robbin, A. (1981b). Strategies for improving utilization of computerized statistical data by the social science community. *International Journal of Social Science Information Studies, 1,* 89–110.

Robbin, A. (1983). Strategies for increasing the use of statistical data. (Occasional Papers Series). Urbana, IL: University of Illinois Graduate School of Library and Information Science.

Robbin, A. (1986, May 14–18). *Data rich but communication poor.* Paper presented at the meeting of the American Association of Public Opinion Research, St. Petersburg Beach, Florida.

Ryscavage, P. (1987). *SIPP: Filling data gaps on the poverty and social welfare fronts.* (Working Paper No. 8705). Washington, DC: Bureau of the Census.

Smilowitz, M., Compton, D. C., & Flint, L. (1988). The effects of computer-mediated communication on an individual's judgment: A study based on the methods of Asch's social influence experiment. *Computers in Human Behavior, 4,* 311–321.

Sproull, L. (1986). Using electronic mail for data collection in organizational research. *Academy of Management Journal, 29,* 159–169.

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science, 32,* 1492–1512.

Sproull, L., & Kiesler, S. (1991a). Computers, networks and work. *Scientific American,* 265(3), 116–127.

Sproull, L., & Kiesler, S. (1991b). *Connections: New ways of working in the networked organization.* Cambridge, MA: MIT Press.

Steinfield, C. W. (1986). Computer-mediated communication in an organizational setting: Explaining task-related and socioemotional uses. *Communication Yearbook 9* (pp. 777–804). Beverly Hills, CA: Sage Publications, Inc.

Stohl, C., & Redding, W. C. (1987). Messages and message exchange processes. In F. M. Jablin, L. L. Putnam, K. H. Roberts, & L. W. Porter (Eds.), *Handbook of organizational communication: An interdisciplinary perspective* (pp. 451–502). Newbury Park, CA: Sage Publications, Inc.

Tennant, R. (1991). Internet basic training: Teaching networking skills in higher education. *Electronic Networking Research, Applications, and Policy, 1* (2), 37–46.

Tombaugh, J. W. (1984). Evaluation of an international scientific computer-based conference. *Journal of Social Issues, 40*(3), 129–144.

Turoff, M. (1989). An anatomy of a computer application innovation: Computer mediated communications (CMC). *Technological Forecasting and Social Change, 36,* 107–122.

Turoff, M., & Hiltz, S. R. (1978). User behavior patterns in the electronic information exchange system. *Proceedings of the ACM, 8,* 659–665.

Turoff, M., & Hiltz, S. R. (1982). The electronic journal: A progress report. *Journal of the American Society for Information Science, 33*(4), 195–202.

Tushman, M. L. (1977). Special boundary roles in the innovation process. *Administrative Science Quarterly,* 22, 587–605.

Tushman, M. L., & Katz, R. (1980). External communication and project performance: An investigation into the role of gatekeepers. *Management Science, 26*(11), 1071–1085.

U.S. Congress. 102nd Congress. (1991, December 9). *High performance computing act of 1991.* Public Law 102–194. U.S. Code, 105 Stat. 1594.

U.S. Congress. Office of Technology Assessment. (1986, April). *Research funding as an investment: Can we measure the returns?—a technical memorandum.* (OTA–TM–SET–36). Washington, DC: U.S. Congress, Office of Technology Assessment.

Utterback, J. M. (1971). The process of innovation: A study of the origination and development of ideas for new scientific instruments. *IEEE Transactions on Engineering Management, EM–18*(4), 124–131.

Weedman, J. (1991). Task and non-task functions of a computer conference used in professional education: A measure of flexibility. *International Journal of Man-Machine Studies,* 34, 303–318.

Williams, F., Rice, R. E., & Rogers, E. (1988). *Research methods and the new media.* New York: Free Press.