FEATURES

3 Archives Law and Machine-readable Data Files: A Look at the United States
   By Thomas Elton Brown

8 An Archivist's Challenges: Adapting to Changing Technology and Management Techniques
   By Donald Fisher Harrison

14 Issues of Privacy and Access
   By Per Nielsen

23 Archives and Dinosaurs
   By Eric Tanenbaum

32 Promoting a Computer Conference, Continued: The Experience of the Association of Public Data Users
   By Patricia C. Becker

35 Topically-focused Data Archives: A New Paradigm for the Codification of Social Science Research
   By Josefina J. Card

44 The Potential for Computer Communications Among ICPSR Representatives
   By Charles Humphrey

DEPARTMENTS

51 Memoriam
FEATURES

3 Archives Law and Machine-readable Data Files: A Look at the United States
By Thomas Elton Brown

8 An Archivist's Challenges: Adapting to Changing Technology and Management Techniques
By Donald Fisher Harrison

14 Issues of Privacy and Access
By Per Nielsen

23 Archives and Dinosaurs
By Eric Tanenbaum

32 Promoting a Computer Conference, Continued: The Experience of the Association of Public Data Users
By Patricia C. Becker

35 Topically-focused Data Archives: A New Paradigm for the Codification of Social Science Research
By Josefina J. Card

44 The Potential for Computer Communications Among ICPSR Representatives
By Charles Humphrey

DEPARTMENTS

51 Memoriam
Editorial Information

The IASSIST QUARTERLY represents an international cooperative effort on the part of individuals managing, operating, or using machine readable data archives, data libraries, and data services. The QUARTERLY reports on activities related to the production, acquisition, preservation, processing, distribution, and use of machine readable data carried out by its members and others in the international social science community. Your contributions and suggestions for topics of interest are welcomed. The views set forth by authors of articles contained in this publication are not necessarily those of IASSIST.

Information for Authors

The QUARTERLY is published four times per year. Articles and other information should be typewritten and double-spaced. Each page of the manuscript should be numbered. The first page should contain the article title, author's name, affiliation, address to which correspondence may be sent, and telephone number. Footnotes and bibliographic citations should be consistent in style, preferably following a standard authority such as the University of Chicago press Manual of Style or Kate L. Turabian's Manual for Writers. Where appropriate, machine-readable data files should be cited with bibliographic citations consistent in style with Dodd, Sue A. Bibliographic references for numeric social science data files: suggested guidelines. Journal of the American Society for Information Science 30(2):77–82, March 1979. If the contribution is an announcement of a conference, training session, or the like, the text should include a mailing address and a telephone number for the director of the event or for the organization sponsoring the event. Book notices and reviews should not exceed two double-spaced pages. Deadlines for submitting articles are six weeks before publication. Manuscripts should be sent in duplicate to the Editor:

Walter Piovesan
Research Data Library
W.A.C. Bennett Library
Simon Fraser University
Burnaby, B.C., V5A 1S6 CANADA
(01)604/291-4349 E-Mail: Piovesan@SFU.MAILNET

Book reviews should be submitted in duplicate to the Book Review Editor:

Kathleen M. Heim
School of Library and Information Science
Louisiana State University
Coates Hall, Room 267
Baton Rouge, Louisiana 70803 USA
(01)504/388-3158

Key Title: Newsletter – International Association for Social Science Information –Service and Technology

ISSN – United States: 0739-1137 Copyright © 1985 by IASSIST. All rights reserved.
Archives Law and Machine-readable Data Files: A Look at the United States

by Thomas Elton Brown
National Archives and Records Administration
Washington, D.C., United States of America

Introduction

In the strict sense of the word, archivists have responsibility for the official records of an organization, in contrast with manuscript curators who collect private documents accumulated by an individual person, or librarians who manage publications. The organizational records which the archivist is to manage may include a variety of materials, including machine-readable data files. How these corporate records — regardless of media — are created, maintained, preserved and accessed is specified in the organization’s official policy statements. Such policies will generally specify who in the organization has responsibility for each of these activities relating to the organization’s official records. When the organization is a government entity, these policies are embodied in the laws or statutes of the government. Such laws are of obvious importance to government employees concerned with records since the statutes specify the basis for the activities relating to records by each agency and its personnel. Individuals wanting information from a government agency should also be aware of these laws because they have direct impact on the accessibility of the information. This paper will review the provisions of the laws relating to archives in the United States, relate them to machine-readable data files in the Federal Government, and then will use the records of the Bureau of the Census to illustrate the legislatively mandated approaches.

Within the United States, the Federal Government primarily controls the creation and disposition of record material through the Federal Records Act of 1950 as amended. This statute defines records as:

- all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States Government under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of data in them. [44 U.S.C. 3301]
One should note that the definition of the records in this statute specifically includes "machine-readable materials."

The Federal Records Act also includes a provision that states:

- The head of each Federal agency shall make and preserve records containing adequate and proper documentation of the organization, functions, policies, decisions, procedures, and essential transactions of the agency and designed to furnish information necessary to protect the legal and financial rights of the Government and of persons directly affected by the agency's activities. [44 U.S.C. 3101]

It is this provision that grants to the head of each agency the authority to determine what records the agency will create. Thus it is the Federal agency that determines what machine-readable information will be collected and processed.

Once an agency has created machine-readable records, they cannot be destroyed without the approval of the Archivist of the United States. If the Archivist determines that the machine-readable record has archival value and should not be destroyed, then disposition of the data involves their transfer to the National Archives for continued preservation.

When will the data be transferred? The timing of the transfer may best be described as a date negotiated between the agency and the Archives. For all records regardless of media, the Archivist:

- may direct and effect the transfer to the National Archives of the United States of records of a Federal agency that have been in existence for more than thirty years and determined by the Archivist of the United States to have sufficient historical or other value to warrant their continued preservation by the United States Government, unless the head of the agency which has custody of them certifies in writing to the Archivist that they must be retained in his custody for use in the conduct of the regular current business of the agency. [44 U.S.C. 2107]

Because of the fragile nature of machine-readable records, a special provision for information on this medium has been added to the regulations which all Federal agencies must follow:

- When the National Archives and Records Service [Administration] has determined that a file is worthy of preservation, the agency should transfer the file to the National Archives as soon as it becomes inactive or whenever the agency can not provide proper care and handling of the tapes to guarantee the preservation of the information they contain. [41 C.F.R. 101-11.411-6]

In addition, the National Archives has the authority to establish the procedures which constitute proper care and handling. [41 C.F.R. 101-36.12]

Access by the public is governed by the Freedom of Information Act, the Federal Records Act, and individual statutes governing specific programs or data collection activities. The Freedom of Information Act generally provides that any person has the right of access, enforceable in court, to Federal agency records except to the extent that such records (or parts of those records) are protected from disclosure by any one of nine exemptions. This statutory guarantee to access Federal information applies equally to all record material — whether in the custody of the creating agency or in the National Archives. Thus the act of transferring the information to the National Archives neither expands or limits the right to access the information. The limitations on access stem not
from the physical location of the material but from the nine exemptions. One of these nine exemptions is "all matters specifically exempted from disclosure by statute." [5 U.S.C. 552] According to the Federal Records Act, all statutory limitations and restrictions on the examination and use of the records while in agency custody are transferred with the records when they go to the National Archives. Again the physical custody of the records does not affect any restrictions on access. These statutory restrictions:

- shall remain in force until the records have been in existence for thirty years unless the Archivist by order, having consulted with the head of the transferring Federal agency or his successor in function, determines, with respect to specific bodies of records, that for reasons consistent with standards established in relevant statutory law, such restrictions shall remain in force for a longer period.[44 U.S.C. 2108]

Thus the statutory restrictions acknowledged in the Freedom of Information Act expire after thirty years unless extended by the Archivist of the United States in consultation with the agency.

The machine-readable records of the Bureau of the Census can serve as an illustration of the management of Federal records even though a specific provision of the Federal Records Act governs access to some records in the Census Bureau. First the Census Bureau determines what material it will collect as part of its census and survey activities. In making its determination of what information to collect and how, the Census Bureau actively seeks advice from a variety of sources including other Federal agencies. The National Archives does not offer advice to the Census Bureau on what questions should be asked or on how the censuses and surveys should be conducted. The National Archives does have the statutory responsibility to "provide guidance and assistance to Federal agencies with respect to ensuring adequate and proper documentation of the policies and transactions of the Federal Government." [44 U.S.C. 2904] With regard to the Census Bureau, the National Archives would interpret this provision as authorizing the National Archives to provide advice on how to document how the census or survey collected the information. It would not include advice on what information the census or survey should collect. However, if the Census Bureau determines that it will collect information which the National Archives determines to have archival value, then the Archives will advise the Census Bureau on how to process and maintain the information to ensure that the information is retained in a format that can be transferred to the National Archives.

Title 13 of the United States Code is the legislation which authorizes the Census Bureau to collect and process its data and imposes three restrictions on the information gathered by the Census Bureau. The Census Bureau may not:

- use the information furnished under the provisions of this title for any purpose other than the statistical purposes for which it is supplied; or

- make any publication whereby the data furnished by any particular establishment or individual under this title can be identified;

- permit anyone other than sworn officers and employees of the Department [of Commerce] or bureau or agency thereof to examine the individual reports. [13 U.S.C. 9]

To comply with these limitations and yet to provide users with needed data, the Census Bureau creates public use files, either extracts or microaggregations. In this way, the Census Bureau can release information which will not
identify a respondent — whether an individual person or economic establishment. The National Archives has the responsibility for determining which information has archival value and which information may be destroyed when no longer needed by the agency. This determination is what the archivist refers to as "appraisal." Such appraisal of machine-readable information is done separately for microdata files with individually identifiable records, for public use extracts, and for microaggregations. The Federal Records Act would normally limit Title 13's restriction on the release of individual information to thirty years unless extended by the Archivist. However a provision in the Federal Records Act stipulates that:

- with regard to the census and survey records of the Bureau of the Census containing data identifying individuals enumerated in population censuses, any release pursuant to this section of such identifying information contained in such records shall be made by the Archivist pursuant to the specifications and agreements set forth in the exchange of correspondence on or about the date of October 10, 1952, between the Director of the Bureau of the Census and the Archivist of the United States... [44 U.S.C. 2108]

The key to this agreement is that:

- after the lapse of seventy-two years from the enumeration date of a decennial census, the National Archives and Records Service [Administration] may disclose information contained in these records for use in legitimate historical, genealogical or other worthwhile research. [H.R. Report 95-1522, August 21, 1978]

The statute that makes reference to this exchange of correspondence also grants the two agencies the authority to amend the agreement, provided that they publicize the change in the Federal Register.

The statutory clause which makes reference to the exchange of letters specifies "census and survey records of the Bureau of the Census containing data identifying individuals enumerated in population censuses". Thus, this statute and the seventy-five year provision apply only to demographic information dealing with individual persons. They do not apply to the economic censuses and surveys which gather information from business establishments.

What laws do apply to identifiable information on business establishments? Under the authority of the Federal Records Act, the Archivist has appraised most of the microdata from the economic censuses and surveys as having sufficient value, primarily for economic time series studies, to warrant continued preservation in the National Archives. However, Title 13 restricts access to this information to Census Bureau employees only. As discussed earlier, the Federal Records Act limits statutory restrictions to thirty years unless extended by the Archivist of the United States. This statute also empowers the Archivist of the United States to direct and effect the transfer of any of these records which not used in the regular, current business of the Census Bureau.

Since such old economic information is not needed in the regular current business of the Bureau of the Census, the agency has agreed to transfer the information to the National Archives when the information is thirty years old. The mere transfer of material to the National Archives for continued preservation does not necessarily mean that the information is available to the public; the National Archives routinely accessioned material to which access is denied for a period of time. Of course, any such restriction on access must be sanctioned by one of the exemptions of the Freedom of Information Act. A statutory restriction can be extended beyond thirty years by the Archivist in consultation with the agency "for reasons
consistent with standards established in relevant statutory law." Because of the permissiveness of this authority, Census and Archives personnel have from time to time discussed when the National Archives would be able to release Census-gathered machine-readable information concerning individual economic entities. To date, however, no agreement has been reached.

This review can allow one to draw some conclusions about records administration within the United States government. The legal provisions which relate to records and archives are "media non-specific" in that the statutes relate to all record material regardless of medium. However, as seen in the 1952 agreement regarding Census material, these policies and responsibilities generally have been developed to deal with human-readable records and have later been applied to all record material. The statutes divided responsibility for the administration of the record material. But in this division of responsibility, the Archivist has significant powers which have an impact on access to the information. The first of these powers is the exclusive authority to sanction the destruction of record material. Obviously, the destruction of a document or a machine-readable data file effectively limits access to it. The Archivist has the authority to direct the transfer of non-current material to his custody after the records are thirty years old. Finally, the Archivist has primary responsibility to determine whether statutory restrictions will be extended past the thirty-year statutory limit. While only having a minimal impact on current or contemporary records, these latter powers can be decisive in determining access to older information. Yet, because of this division of authority, disagreements are possible among those sharing records management responsibilities. Until these differences are resolved, open questions — such as the ones about access to microdata from the economic census — will remain.
Introduction

Over twenty years ago, the National Archives of the United States embraced the concept that automated records were actually records which could be considered permanent within the meaning of the Federal Records Act and set about collecting them. Since then it has confronted problems incident to finding these automated records, acquiring them, preserving them and making them available to the public. Previous papers have discussed access to public automated records in the normal sense; that is, the ability of the researcher to get at them. In this paper I wish, however, to discuss the National Archives' acquisition process as a form of access.

This paper addresses three threats to the acquisition of machine-readable records: the threat of an onslaught of hardware and software incompatibility, the threat of discontinuity within textual records series brought about by end-users with microcomputers and the threat brought about by new management techniques from the Paperwork Reduction Act of 1980. Archivists ought to view these threats as challenges. When overcome, the challenges will have presented the Archives with the opportunity to create a better collection of automated records.

Software and hardware dependency

The first challenge to the National Archives is well publicized and needs no significant introduction in this treatise. The Archivist of the United States, confronted with the research community's complaint that valuable data were being created by Federal agencies without any consideration for their preservation or dissemination to the public, established in the...
1960's the forerunner of today's Machine-readable Branch. This branch was given the task of inventorying Federal data bases and deciding how best they should be preserved for posterity. We accessioned a number of machine-readable data files created in the 1960's. Some of these files were dependent on other outside factors and could not be read on their own. Three examples of software and hardware dependency illustrate our initial problems.

The first example came early in our organizational being. We received over thirty-five machine-readable data systems from the Office of the Secretary of Defense and the Office of the Joint Chiefs of Staff. These systems were encoded in a data base management system called the National Information Processing System (NIPS). They caused serious problems in access and handling and a considerable backlog in the accessioning workload.

NIPS was devised for generalized file handling using languages designed to support user requirements in six components. It afforded any data center the capability of reporting long and involved statistical manipulations on extremely short notice to a variety of users. However, the software was compatible only with IBM computers.

The presence of NIPS files suggested serious difficulties in providing a uniform reference service to researchers and brought up the whole question of software dependent files. To retain the files in NIPS would constitute a precedent. Since researchers by and large preferred to use their own utility software, transportable files would afford a range of options that encoded files would not. Last but not least, maintaining large inventories of software would add to the preservation costs and require more shelf space. For all these reasons we decided to decode the files. It appears now, with hindsight, that, despite the fact that these files were unique and very valuable, we should have insisted that the material be transportable before being accepted by the National Archives.

The second example was the National Archives' accessioning of a microfilm series of records containing pictures of captured North Vietnamese documents. These were filmed in Saigon during the war by the Combined Document Exploitation Center on 94 oversize (13-inch) rolls of 35mm microfilm, each roll 1000 feet long. The documents were on one side of each frame, with digital bar codes on the other side to provide indexing and control information.

Soon after we received the microfilm we discovered to our chagrin that the material was hardware dependent in a system known as "File Search." Four configurations of this machine had been manufactured and sold to Federal agencies in the 1960's. The last model (generation four) had a small computer in it. It could therefore provide a printout by reading the bar code on the film strip, transferring it to magnetic tape, which in turn could be manipulated and dumped on to paper. The machines cost $250,000 new and were used only by military and intelligence agencies, as far as is known.

It was only after this information was made available to us that we discovered that other file systems were known to exist in this environment and were equally unreadable without any machines in existence to retrieve data. These included some important files in the Navy Sea Systems Command (in Arlington, Virginia) and the Navy Oceanographic Command (in Bay St. Louis, Mississippi), including the Defense Intelligence Agency. Recently we have discovered the existence of an intact File Search model in salvage channels. We have requested that it be turned over to the National Archives, and we think we have the technical expertise to restore the model to operating condition.
The third example entailed the 1960 Decennial Census, offered to the National Archives by the Census Bureau in the mid 1970's. These records were created by a UNIVAC II-A computer, of a generation that had been effectively phased out of use in the Federal government after the tapes had been created. It has been reported that once the tapes became available for transfer to the National Archives, only two such machines existed to read them, one in Japan and one in the Smithsonian Institution. Eventually a reasonable approach was agreed to by the National Archives and the Census Bureau, to convert the data into a compatible format, making them available for preservation in our vaults.

These three examples are illustrative of the long term problems created by hardware and software dependency of records created in the 1960's, when computers were maintained in relative isolation from each other. It was a period in which data managers were concerned with the creation and the use of computer products and were by and large ignorant of the long term value of these products as Federal records. It can be said that the letter of the law — the fact that the tapes were handed over to the National Archives — was carried out. The fact that the tapes were unreadable because of software and hardware dependency was a new problem that had never been faced with paper acquisitions. For their part, agencies were understandably reluctant to dispense funds solely for the benefit of depositing these records in the National Archives. Thus reason has had to prevail in our dealings on transfer of the tapes, and no one solution can be applied in all cases.

---

**Small computers and office automation**

The second challenge to the smooth flow of records into the archives stems ironically from the very machine meant to facilitate administrative operations in the modern office. For several years, most federal agencies have been extending the advantages of their word processing pools by placing terminals at the hands of management officials, giving fingertip control to their own records creation. Office automation (AO), more aptly termed "the paperless office", is based on a series of compatible, menu-driven programs utilizing a centralized data base for common shared-use data and unique smaller databases for individual users. These systems have the ability to transfer data and information between data bases through a network or a distributed system.

The advantages of such a system are obvious. Federal managers frequently need information suddenly and immediately, and often the demands for this information come after the staff has left for the day or the weekend. Managers would like the ability to search for the data or reports they need through an indexed automated bibliographic/numeric data base, access and use the appropriate software to perform simple to moderately complex analyses of this data (e.g. forecasts, correlations, etc.), use graphics to illustrate their results, access word processing/office automation tools to produce a memo in the appropriate format, and finally send this report/memo electronically to the recipient's office, all without the necessity of using the phone, typewriter, or staff that are not available.

Keeping all this in the system can cause an archival "log jam". The designers and the users of paperless office systems are frequently ignorant of the paper systems they are replacing and the archival need for intellectual continuity. Outside contractors compound the problem. In

---

1 Committee on the Records of Government, Report, Washington, DC, March 1985, pp. 86-87.
the absence of any other information the hardware and software dependency problem has reemerged in the small computer world, and agencies are finding that transportability cannot cross the boundaries between offices. Software now provides end users with ultimate fingertip access. This allows handcrafted programming and instant manipulative gratification. The same person who creates data on the system can now dispose of it with equal ease. By closing the gap between the user and the machine, the system eliminates the apparent need for the data middleman, to say nothing of the records manager who, under other circumstances, looked after standardized formats, ensured traditional records disposition practices and provided for a continuity of records series in the agency.

Thus the danger inherent in OA is that the practice concentrates on the information as it is used immediately after creation without making a record of actions taken. It is said to parallel the dangers of telephone use when first introduced. With that instrument, managers needed go through no intermediate device for communication. Telephones assured privacy of communication and were sheltered from the public record. The comparison with OA is evident. Just as managers could converse at the push of a telephone button, so they do now with electronic mail. Further, if one of the parties is absent, there need be no callback, because the mail has already been delivered electronically.

Like the telephone, the OA challenge is to find a way to record the communication. With the small computer, software must be devised to ask the user for a determination of the ultimate value of the information before it is ever keyed into the system. This software has been integrated into the planning for OA systems in most Federal agencies. Whether or not it solves the problem in practice remains to be seen.

Information resources management

The third challenge to a smooth transfer of records to the archives now comes in the form of an application of new management techniques to the creation and the use of information within the Federal establishment. This new methodology typically accommodates the reality that government must function with less personnel and with individuals of lesser skill and training by altering the way agency missions are carried out. The Paperwork Reduction Act of 1980 was rightly concerned with a problem that had existed for some time in that the Federal government was preoccupied with the physical problems associated with the large volume of paper records created. The authors of the bill reasoned that managers should have been concerned with how the information was being controlled and how it could be shared with the maximum number of sources. Thus the new law espoused intellectual control vis-a-vis physical control, regardless of the medium on which the information had been stored. In order to do this a number of organizational changes have taken place in Federal agencies, each a bit different from the next, in which an "information Czar" has been placed at the highest levels to control access and dissemination of all information, regardless of the medium. This new arrangement has now been entrenched for four years.

A typical arrangement has been established to combine the former functions of "automation, communications, office automation, records management, publications, audiovisual activities and other information activities, services and facilities." An information management plan is usually mandated beginning with a problem analysis, designing a model information system, constructing the "architecture" which produces a program and provides guidance for a budget request. Under this concept, every information system will have centralized management. The
"single manager" concept has been extended to encompass all information, defined as "... all processes by which the user may receive, display or project desired information... (including) voice, text, graphics, audiovisual, video teleconferences, micrographics, files, records management, optical discs and other forms of published information."³

In many ways, the single manager system makes a lot of sense. The information manager is in a unique position to disseminate information within an agency to avoid duplication of effort — or better, to avoid disparate and conflicting data creation. By being organizationally placed at the highest level, the IRM provides information for important decisions and controls a sizeable portion of the agency’s budget.

Furthermore, the concept will ease the path of liaison between the agencies and the National Archives. As we began to accession records in machine-readable form in the 1960’s, we became increasingly aware of the presence of the data manager as a viable records creator and manager. Between 1961 and 1980, the Machine-readable Branch frequently communicated with the data manager directly when it was not able to get required information any other way. Furthermore, in the first half of this decade, we became more and more concerned with dealing directly with government managers since they were creating (and destroying) information without acknowledging either the Federal Records Act or their agency records administrator. With the advent of the IRM principle, however, the Archives need only deal again with one official, who, if properly briefed on the urgency of the problem, will coordinate the actions of the records manager, the data manager and the end user.

³Draft Army Regulations 25–1 and 25–5, 1984.

Conclusions

Technology has created new solutions to old problems, but in the process has itself created new problems. The archival community is thus confronted with unique challenges to its traditional role as keeper of the records, which requires our attention. Some measures come to mind as actions to stem the tide.

First, the archivist must keep professional pace with the proliferation of computing technology, not only as it is practiced in Federal agencies in this decade, but also as many writers envision that it might be practiced 25 years from now. Reading the literature is not enough. It requires a shrewd selection of educational services and an on-going dialogue with other archivists. This must include at a minimum the study of software, hardware and storage media as trends develop. An archives must be capable not only of receiving machine-readable records in various modes and written on various media, but also of serving its users with a multiplicity of arrangements.

This leads to the second measure. The archivist must determine far enough ahead in time in what mode and on which medium these new records will appear as candidates for acquisition. To do this, archivists must assert their professional needs to the creators of records throughout the life-cycle of the records.

Furthermore, the requirement to deposit tapes and other media in the National Archives should be anticipated and budgeted by Federal agencies.

Third, the archivist must reach end users by some means, to ensure standardization of practices and procedures. It is vitally important to overlay records management practices on the uses and outputs of small computers and of office automation systems. This might include communicating with procurement officers and
IRM officials to standardize hardware and software packages which would be interchangeable within and between Federal agencies.

The fourth, and by no means the least important, point is that the IRM developments in Federal agencies, formed as a result of the Paperwork Reduction Act of 1980, must be influenced by direct communication with archival officials.

IRM managers have been imbued with the immediate needs of the agency information program in mind — the here and now concept. There is always the danger that not enough planning will be conducted for the ultimate fate of records. By the way they maintain certain modes of information, IRM officials can influence the disposition, and in turn, the configuration of future holdings of the National Archives.

**Records Acquisition Liaison for Flow of Records Between The National Archives and Federal Agencies**

- **1961-1965**
  - Archivist
  - Records Administrator
  - ADP Manager
  - END USER

**1981-1985**
- Archivist
- Records Administrator
- ADP Manager
- END USER

**1986**
- Archivist
- Records Administrator
- ADP Manager
- END USER

**1981-1985**
- Archivist
- Records Administrator
- ADP Manager
- END USER

**1986**
- Archivist
- Records Administrator
- ADP Manager
- END USER

*Spring 1986*
Issues of Privacy and Access

by Per Nielsen

Early academic reactions to privacy and access regulations

The IFDO resolutions, August 1978

For a few years in the mid-seventies, members of the international social science community could study the Hessian and the Swedish data legislation practices whilst preparing the viewpoints for which they found it necessary to fight on their national home ground before the enactment of similar privacy legislation. To member institutions of the newly established International Federation of Data Organizations (IFDO), the issues of privacy and access were of central importance, not just as an academic field of study, but as central issues that might represent a threat to the survival of the data organizations – and to a certain extent even of quantitative social research.

Consequently, in August of 1978, IFDO sponsored an International Conference on Emerging Data Protection and the Social Sciences' Need for Access to Data. At this conference, which was hosted by the most experienced and biggest European Data Archive, the Zentralarchiv in Cologne, comparative national status reports were presented from 10 countries. The national reports were sent to the organizers who collected them in a volume of proceedings that was a tangible point of departure for the discussions at the 3-day conference.

The participants invited to the Cologne Conference unanimously adopted three IFDO Resolutions which served the purpose of drawing attention to as many aspects and consequences of data legislation as possible. The IFDO Resolutions are appended to this note as one of the first, outspoken, academic reactions to privacy and access regulations. They are included in the same form as that in which they were presented to the Danish public in the Danish Data Archives (DDA) newsletter.

The Bellagio Principles, 1977

One year prior to the IFDO Conference, a group of social scientists and senior administrators at national statistical bureaux had discussed access to statistical data in Bellagio, Italy. From this event, 18 so-called Bellagio Principles were circulated in the social science community. These principles were considered important because they represented a first compromise between social scientists on the one hand and senior statistics administrators on the other.

1 Presented at IASSIST/IFDO International Conference May 1985, Amsterdam.

2 DDA-nyt 9:12-14, December 1978.
In an explicit statement, the IFDO Conference endorsed the Bellagio Principles, which are reproduced below – again in the same form as that in which they were presented to the Danish public in the DDA newsletter.3

The European Science Foundation statement, 1979–1980

In 1979, a working group of invited specialists in the social sciences, the medical sciences, and administrators from data inspection authorities, tried to reach agreement on a statement which was going to be subject to approval by the assembly of the ESF. During the working group meetings, I felt a peculiar distrust between each group and the other two. The medical experts asserted that their data-handling procedures were safe and felt that it was in the interests of patients (i.e. the public) to supply necessary information to their doctors – without too much interference from the data inspection authority; on the other hand, the medical experts were sceptical about some of the data collection and handling procedures applied by social scientists! The experts with a social science background tended to hold that their own rationale for data collection, as well as their applied data handling routines, were less dangerous to the public than most of the data collection ventures within medical science. And the data inspection authorities felt that both medical and social science projects involving confidential data should be rather rigidly controlled.

In addition to these disciplinary variations in attitudes, the national differences were more outspoken in the ESF working group than they had been in either Bellagio (where Canada, US, UK, West Germany and Sweden were represented by scientists and statisticians) or in Cologne in which about a dozen countries were involved. Furthermore, it took more than a year to reach agreement on the wording of the final text. After reworking the text as adopted at the conference, a slightly rephrased version was accepted by the ESF Assembly. It is this revised (official) version of the ESF Statement which is appended to this paper.

Reasons behind the diversification in attitudes

I think it is fair to say that three or four major factors caused the change in attitudes to the issues of privacy and access during the last half decade of the seventies – from consensus to a more diversified set of attitudes. First, the various groups of agents became more aware of their group interests in the course of the data legislation process as the latter proceeded in more and more countries. Second, the discussions moved from a level of soft statements towards one of juridical phraseology in sections and subsections. Third, the differences in existing legal conditions between countries (e.g. in such areas as freedom of information) as well as practical set-up (e.g. a tradition for codes of ethics) implied associated differences in the new legislation and in its actual implementation.

This indicates that there is still a lot of research to be done in terms of comparing the conditions for quantitative research between countries as well as following the trend over time within a single country – as practices are defined and acts are amended.

As can be expected, substantial interest is devoted to this issue among social science data "pushers" and "addicts". Since 1977, there has hardly been a conference of any size or generality which has not had issues of privacy and access on its agenda.

Concluding remarks and recommendations

As a convenor of the IFDO/IASSIST 1985 Conference session on Issues of Privacy and Access, I thought that it might be useful to reprint some of these early deliberations, in

---

1 DDA-nyt 9:14–16, December 1978.
order to facilitate discussion along the following lines: what new issues (if any) have entered the debate in recent years, and what is the present-day situation, compared to expectations 5 or 10 years ago.

Finally, I should very much like to see a repetition of the 1978 IFDO conference. Now that most countries have actually been living with enacted privacy bills, a new systematic comparison across countries would prove useful.

---

**IFDO**

International conference on emerging data protection and the social sciences' need for access to data

- Resolutions

In a plenary session the conference unanimously adopted the following three statements.

Social scientists' experiences with data protection.

On the basis of evaluation of developments in data protection within eleven countries, and taking account of the general tendency for legislative measures to have unintended consequences, the conference expresses grave concern about some of the negative impact of data protection laws, regulations, and practices on the social sciences. While we recognize that it is essential to protect the privacy (integrity) of the individual, there is also a need to know and a need to secure the channels through which, under proper safeguards, a reliable and comprehensive understanding of the life situation of individuals and groups of individuals may be obtained.

In the opinion of the conference the need to know and the need to secure a free flow of information constitute the other side of the issue of protecting the privacy of individuals. To a large extent this other side of the privacy issue has not been given due consideration in the process of enacting and implementing data legislation. The conference would like to draw attention to the fact that such legislation can and has become a vehicle for the protection of the vested interests of particularly resourceful groups and organizations, thus contributing toward an infringement of the fundamental rights of other parts of society. It is recognized that the results of significant social research might jeopardize the interests of some of the groups or individuals about whom data are collected. However, it seems important to be sensitive to the possibility that because of this situation data protection measures can be utilized as a shield behind which socially significant issues are excluded from research.

Furthermore, developments in the field of information processing have resulted in very powerful instruments to control individuals and society. In most of the countries represented at the conference data protection laws are used by bureaucracies to monopolize the information necessary for the open discussion of public policies. The data flow among government agencies has increased considerably during the last few years, although data protection has in some cases placed restrictions on this flow. However, researchers often find themselves excluded from the information necessary to enable them to contribute to public discussion by presenting independent opinions. This is especially dangerous in a situation where government policies are based increasingly on large data bases, including microdata.

The conference is of the opinion that these issues have significant political implications and are associated with broad and general notions of the free and unrestricted flow of information in society. They should be given thorough political
consideration in the future development of data legislation and practices.

The conference has learned that with respect to data protection there are significant differences in the situations of the different countries. There are nations that have found an acceptable balance between data protection and access to data for research purposes. On the other hand, there are countries where data flow for research has come nearly to a standstill.

In this situation it is necessary to develop guidelines for a general information policy. A fundamental aim of a modern information policy is to make information gathered by public (and private) institutions more transparent and visible in order to improve democratic control. Within this broader framework, social research must be considered not only as a matter of interest to social scientists, but as part of that system of democratic control.

A first important recognition of these problems at the international level came in 1977, when a group of social scientists and senior administrators of national statistical bureaus discussed the issue and drafted a set of recommendations, which are now known in the international social scientific community as the Bellagio Principles. We endorse these principles. We also hope that the pattern set by the Bellagio conference of joint discussion of common problems between social scientists and governmental officers at all levels will be continued.

In the perspective, the distinction between statistical and administrative data should not be used to make the latter less accessible to researchers. Access to administrative data for scientific purposes should be regulated according to the principle of functional separation of research and administrative data incorporated also in the Bellagio Principles.

The conference wishes to point to the high value placed on freedom of the press. The social science community might be in a better position to improve its services to society if its freedom and rights to do research were secured through similar principles, including the obligation to protect the source of information.

Preservation and accessibility

In addition to these general principles the conference recognized other points of interest for the international development of social research. In particular it recommended:

- that the data relevant to scientific investigations on human affairs should be preserved in readily usable forms;
- that with the sole limits of protection of privacy and confidentiality recognized in the first part of this statement, research data should be openly accessible to social researchers and the general public of all nations;
- that governments should work to eliminate barriers to general access to research data and should take appropriate action to facilitate their use under the principles established by the United Nations charter and incorporated in UNESCO.

Codes of conduct

Finally the conference supported the following recommendations toward the adoption of codes of conduct by social researchers:

- Social scientists collect information from and about individuals for research purposes. In doing so they have traditionally followed certain standards of behaviour: social research is conducted at all times so that no harm should come to
individuals while being subjects of research.

- The current concern to better protect the privacy of individuals makes it necessary to increase awareness of differences between administrative and research uses of information.

- To make this point better understood by the public, governments and researchers, it is recommended that in addition to the existing codes of ethics in various disciplines, codes of conduct should be developed for each research methodology. These can make explicit the rules that are already respected by the professional researcher. Thus, by common practice in survey research, the anonymity of respondents, their right to be informed about the purpose of a study, their right to refuse cooperation at each stage of an investigation, and their right to know the identity of the researchers have been respected.

- The practical ground rules for the responsible research use of personal data will differ with the research method. Each professional specialty should be asked to make its practitioners fully aware of the range of alternative techniques available to implement codes of conduct. For survey research, as an example, such alternatives include randomized response methods, insulated data banks, and appropriate levels of aggregation.

- Codes of conduct should have sanctions so that the public can be assured that such codes of conduct are more than mere declarations.

**The Bellagio Principles**

Excerpted from David H. Flaherty's report.

1 National statistical offices should provide researchers both inside and outside government with the broadest practicable access to information within the bounds of accepted notions of privacy and legal requirements to preserve confidentiality.

2 Legal and social constraints on the dissemination of microdata are appropriate when they reflect the interests of respondents and the general public in an equitable manner. These constraints should be re-examined when they result in the protection of vested interests, or the failure to disseminate information for statistical and research purposes (i.e., without direct consequences for a specific individual).

3 All copies of government data collected or used for statistical purposes should be rendered immune from compulsory legal process by statute.

4 In making data available to researchers national statistical offices should provide some means to ensure that decisions on selective access are subject to independent review and appeals.

5 The distinction between a research file, in the sense of a statistical record (as defined in the 1977 report of the U.S. Privacy Protection Study Commission), and other micro files is fundamental in discussions of privacy and dissemination of microdata. All dissemination of government microdata

---

4 David H. Flaherty, Final report of the Bellagio conference on privacy, confidentiality, and the use of government microdata for research and statistical purposes. Statistical reporter 78-8:274-279, May 1978.
discussed in connection with the Bellagio Principles is assumed to be a transfer of data to research files for use exclusively for research and statistical purposes.

6 There are valid and socially significant fields of research for which access to microdata is indispensable. Statistical agencies are one of the prime sources of government microdata.

7 Public use samples of anonymized individual data are one of the most useful ways of disseminating microdata for research and statistical purposes.

8 Techniques now exist that permit preparation of public use samples of value for research purposes within the constraints imposed by the need for confidentiality. Countries with strict statutes on confidentiality have prepared public use samples.

9 There are legitimate research purposes requiring the use of individual data for which public use samples are inadequate.

10 There are legitimate research uses which require the utilization of identifiable data within the framework of concern for confidentiality.

11 Other techniques of extending to approved research the same rights and obligations of access enjoyed by officers of the government agency need to be considered in terms of better access.

12 There is considerable potential for development of more economical and responsive customized-user services, such as: 1) record linkage under the protection of the statistical office, 2) special tabulations, 3) public use sample for special purposes. Such services must often involve some form of cost recovery.

13 Some research and statistical activities require the linking of individual data for research and statistical purposes. The methods that have been developed to permit record linkage without violating law or social custom regarding privacy should be used whenever possible.

14 Professional or national organizations should have codes of ethics for their disciplines concerning the utilization of individual data for research and statistical purposes. Such ethical codes should furnish mutually agreeable standards of behaviour governing relations between providers and users of governmental data.

15 Users of microdata should be required to sign written undertakings for the protection of confidentiality.

16 Considerable efforts should be made to explain to the general public the procedures in force for the protection of the confidentiality of microdata collected and disseminated for research and statistical purposes.

17 The right of privacy is evolving rather than static, and closely related to how statistics and research are perceived. Therefore, statisticians and researchers have a responsibility to contribute to policy and legal definitions of privacy.

18 Public concern about privacy and confidentiality in the collection and utilization of individual data can be addressed in part as follows:

a. voluntary data collection, whenever practicable,

b. advanced general notice to respondents and informed consent, whenever practicable,

c. provisions for public knowledge of data
uses,

d. public education on the distinction between administrative and research uses of information.

---

**EFS's statement on 'privacy'**

Statement concerning the protection of privacy and the use of personal data for research (adopted by the Assembly of the ESF on 12 November 1980) 5

**Preamble**

The necessity of safeguarding the individual against misuse of his personal data has been repeatedly emphasized, in the last few years, at both the national and the international level. This has been particularly the case in the countries with organizations which are affiliated to the ESF. In Austria, Portugal and Spain data protection is explicitly referred to in the constitution. Specific legislation already exists in Austria, Denmark, France, the Federal Republic of Germany, Norway and Sweden. Draft laws are under consideration in Belgium, the Netherlands and Switzerland, while an official report on the issue has been prepared in the United Kingdom.

There has also been considerable concern with these matters at the international level. The Council of Europe has recently elaborated a Convention for the Protection of Individuals with Regards to Automatic Processing of Personal Data, while the OFCD has prepared a series of guidelines concerning the protection of privacy and the movement of personal data across frontiers. Mention should also be made of the discussions going on within the Commission of the European Communities about a possible directive and of the enquiry carried out by the European Parliament which led to a resolution calling for immediate action.

However, the implementation of data protection laws has led, in an increasing number of cases, to serious restrictions on access to personal data for research purposes. For example, problems connected with the collection and evaluation of information by means of questionnaires, access to information held by public authorities, particularly statistical offices, and the destruction of personal data by such authorities once the purposes for which they were collected have been fulfilled, have been creating considerable concern amongst the scientific community. This led to the drawing up of the Bellagio Principles in August 1977 6 and to an international conference on emerging data protection and the social sciences’ need for access to data which was held in Cologne in August 1978, sponsored by the International Federation of Data Organizations (IFDO). These problems were also discussed at the 10th Colloquy on European Law organized by the Council of Europe at Liege in September 1980.

The ESF fully endorses the necessity of protecting the privacy of the individual. It feels, however, that the attention of the legislators and international bodies concerned should drawn be to the researchers’ case for special conditions for the use of personal data. These should ensure, under proper controls,

---

5 DDA-nyt 18:9–13, sommer 1981.

6 Contained in the Final Report of the Bellagio Conference on Privacy, Confidentiality, and the use of Government Microdata for Research and Statistical Purposes, which was a meeting of representatives of the central statistical agencies of Canada, the Federal Republic of Germany, Sweden, the United Kingdom and the United States held at the Rockfeller Foundation Bellagio Study and Conference Center in Italy, 16–20 August 1977.
access to such data when it is needed for specific research purposes. Accordingly, a group of experts under the chairmanship of Professor S. Simitis, Professor of Civil and Labor Law at the University of Frankfurt and Data Protection Commissioner of the State of Hesse in the Federal Republic of Germany, was set up to draft such a statement. After full discussion and revision within the ESF the following principles and guidelines were adopted by the ESF Assembly at its meeting in November 1980. They are put forward to ensure both the protection of personal data and the necessary access to such data for research purposes.

J. Goormaghtigh
Secretary General
Strasbourg
13 November 1980

Basic principles

- 'Personal data' are, in the context of this document and in accordance with the definition to be found in the Council of Europe's Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data and also adopted by the OECD, any information relating to an identified or identifiable individual.

- Data protection legislation must, in order to fulfill its task, which is to guarantee the respect of privacy, cover all uses of personal data and therefore include its use for research purposes.

- Professional codes of ethics are a complement to legislative measures safeguarding the respect of privacy. The scientific communities concerned should encourage the development of such codes, within the framework of the rules established by the legislator, in order to take into account the specific needs of the different disciplines.

- Freedom of research presupposes the broadest possible access to information. Legislation should, therefore, besides specifying the conditions under which personal data may be used for research, ensure access to the information needed.

- In order to ensure the respect of privacy, research should, wherever possible, be undertaken with anonymized data, following already accepted practices.

- Scientific and professional organizations, together with public authorities, should promote further development of techniques and procedures to secure anonymity. Anonymity should be considered as given, whenever the individual can only be identified with an unreasonable amount of time, cost and manpower (de facto anonymity).

Guidelines

- Any use of personal data for research purposes, irrespective of the aims for which they were or are to be collected, presupposes either the explicit permission of the legislator or informed consent unless the individuals concerned are not identifiable by the receivers.

- There is informed consent when the individuals concerned have been clearly informed:

  a. that the provision of data is voluntary and that a refusal to comply will have no adverse consequences on them;

  b. of the purposes and nature of the research project;

  c. by and for whom the data are being
collected;

d. that the data collected will not be used for any other purpose than research.

- With the approval of the data protection authority, or its equivalent, informed consent is not required in cases where the nature of the research project is such that:

  a. the informed consent of the individual would invalidate important objectives of research;

  b. informed consent could cause mental or physical distress to the individual concerned.

- For the sole purpose of selecting samples for research involving population-based surveys, legislation or other legally acknowledged procedures should permit the use of data concerning name, address, date of birth, sex and the occupation of individuals collected by state agencies for non-research purposes.

- Personal data obtained for research should not be used for any other purpose but research.

- In particular, personal data obtained for research purposes should not be used to make any decision or take any action directly affecting the individual except within the context of research or with the specific authorization of the individual concerned.

- Whenever personal data are used for research, they should not be published in identifiable form unless the individuals concerned have given their consent.

- In the case of personal data used for research, the individual's right to obtain confirmation whether or not data pertaining to him are maintained, to challenge data relating to him and to have data erased, rectified, completed or amended should be limited to other research projects where it is intended that the data be used in an identifiable form.

- The leaders of research projects using personal data should be responsible for ensuring that the necessary technical and organizational measures are taken in order to guarantee the confidentiality and security of the data and for keeping these measures under review in accordance with the latest scientific and technical developments.

- Once the specific research purpose for which personal data have been collected has been achieved, these data should be depersonalized and the necessary measures (e.g. the deposit of identifying code numbers with a central research data archive) should be taken for their secure storage.

- The decision to destroy personal data held by public authorities should only be taken after consideration of their possible future use for research and after consultation with the central data archive or a similar organization.
Archives and Dinosaurs

by Eric Tanenbaum

Introduction

Dinosaurs and social data archives have a lot in common. When both began their existence they had their respective fields pretty much to themselves. Having almost exclusive control over their environment for a long period, dinosaurs and data archives both swept up material whenever possible and, in time, appeared cumbersome and bottom-heavy. From this state both had to confront a changing environment. However, for all the similarities between the two, dinosaurs differ from data archives in at least one important respect — they no longer exist. Thus while it is too late for dinosaurs to learn from data archives, archivists should consider the dinosaurs’ progress if they wish to distance themselves from the dinosaurs’ end. This note suggests how they might do so.

Palaeontologists may differ when assessing the relative weight of specific causes of the dinosaurs’ demise, but there is common agreement that non-adaptation to changing climactic conditions is important in their undoing. In modern terms it could be said that dinosaurs were frozen out by a changing hardware environment. Archives also confront hardware changes, but their impact on archival work is confounded by concurrent software developments.

This paper describes major changes in several areas which affect computerized data archives. On the hardware side, the paper examines improvements in mass storage capacity and the ergonomics of computers (of all sizes). Software developments, in parallel with these hardware changes, encourage new orientations to social information. From among these the paper focuses on "new" database management techniques and electronic publishing — both have implications for archive growth. Changes in hardware and software are combined by improved communication facilities; the catalyst producing the "alloy" lies in the imagination of information analysts (archive users) whose expectations are aroused by these more elementary developments. The paper describes aspects of the agents of change which are germane to the future operation of archives. An integrated systematic approach to the tasks required ensure that future concludes this paper.

1Presented at IASSIST/IFDO International Conference May 1985, Amsterdam. This paper has been previously published in European political data newsletter no. 55:33-44, June 1985.
Mass storage devices

The history of computerized data archives for the social sciences illustrates the evolution of computerized mass storage devices. Cardboard computer cards, or "IBM cards" as they were commonly known, were an early de facto standard medium for data storage. The "data archive movement" of the early 1960's was launched when it was recognized that these cards could be banked centrally for subsequent redistribution to other sites which supported this physical standard.

Although computer cards were reproduced and shipped "by the forest", the medium was not ideal. It is clumsy — cards get dropped, insecure — cards get torn, and expensive — bulk reproduction is a resource intensive activity. It also limited the analyst's access to large volumes of information. Clearly faster forms of "data memory" would yield vast improvements in the kind of service that archives could provide researchers.

The magnetic computer tape offered the medium of distribution that data repositories required. It is not as universal as computer cards, for each brand of computer uses a different mode of tape storage. However, almost all archives have computer software that allows them to read and write tapes written in all formats used in their user constituency. Thus, for example, the British ESRC Data Archive maintains a suite of conversion routines that permits it to transform data from its own in-house standard to any form required by British users.

While magnetic computer tapes gave archives a cheap medium for transmitting subsets of their holdings to analysts working at remote sites, the medium constrains the kind of material that can be accessed. First, in almost all cases, it requires that information be stored as sequential files. This immediately limits the scope of data that can be transmitted to a few discrete chunks, if only because of the effort and skill required to reassemble anything more ambitious at the receiver's end. Second, the medium itself has a small finite capacity. Granted the volume of data that can be stored on magnetic tapes has increased dramatically from the 6.4 megabytes feasible with the earliest tapes to a current 210 megabytes, but still requires six physical tapes to hold the results of the 1981 British population censuses after the data have been subjected to complex compression routine. Operationally, this means that the analyst who wants to select census data from points across the nation is involved in considerable tape manipulation. Third, and finally, tapes, which are volatile, offer poor archival security. Ensuring the physical integrity of a tape-resident database is a labour and time intensive task which a central facility can perform because it can take advantage of economics of scale but which an individual would find restrictive.

For these reasons archivists should welcome the recent emergence of new modes of mass data storage, two of which will be described here as a prelude to a later discussion of how they should be incorporated into archival operations.

Several manufacturers have announced the development of disks that use laser techniques

1A side benefit of this mode of operation, initially designed to cope with the inelegancies of computer manufacturers' whims about tape standards, is that central archives have protected their, and thus their constituency's, data resources by creating a protective buffer between a single in-house standard to which all data are converted and changing external

1'(cont'd) requirements. Thus, when external technological changes occur the entire database can be transformed to the new requirement by a single routine operation which "maps" the old format to the new.

1The comparison is between a 2400' reel recorded at 200 bits per inch ("bpi") and one recorded at 6250 bpi.
to input and output information at extremely high densities onto small robust platters. Thus, for example, one firm's first release promises the storage of one gigabyte (i.e. 1,000,000,000 characters) on a single side of one physical disk. Using the British population census again as an illustration, it ought to be possible to store the entire set of counts on a single disk.

While at their initial release the disks, which cost about £200.00, are somewhat more expensive than the conventional computer tapes required to store a similar amount of information, the radical impact of these new devices will come both because they allow non-sequential access to data and because they are of archival quality, offering a minimum of ten years' secure storage. Data analysts can now realistically contemplate linking large volumes of information from diverse sources in their pursuit of new connections between and among social phenomena.

In response to this facility, archives have to reconsider how they service their constituency. Eventually, archives will have to meet the needs of analysts who have access to mass storage devices by supplying mass data packages. These will likely be based on diverse data sources which might in turn be linked by "discrete", but otherwise broad, "story lines". This orientation to data, for which the Italian and Norwegian data archives' work constructing ecological databases is a precedent, will have to be extended to many areas of social inquiry and will conceivably require a more active intervention in the work of data archives by subject specialists acting in an editorial capacity.

Optical disks, because of their robustness and cheapness, are amenable to distribution in much the same way as traditional magnetic tapes are. Their local use (by independent analysts) is feasible, as the manufacturers of optical disk drives generally use a standard "interface" between computer and drive. Thus, unlike tape equipment, it is possible that this kind of mass storage will soon be available even for desktop "personal" computers.

However, optical disks have value to the archives' own computer installations. For the British Data Archive, it is estimated that over 80% of its files are sufficiently stable to make it sensible to transfer the bulk of its holdings to these devices. This would have the immediate advantage of simplifying internal operating procedures, even if the Data Archive continues to supply most of its users with copies of data files for access on their local machines. However, if one considers another development in technology, the "networking" of computers which permit individuals to address many computers directly from a single site, these storage devices assume a higher profile in the archives' future landscape, because, with a conceptually, if not technically, "simple" modification, they offer an almost limitless volume of fast access data retrieval.

Physically, optical disks resemble long playing gramophone records. Thus, as with gramophone records, these disks can be stored in a machine similar to a "juke box" whereby a would be listener (analyst) can choose any song (data file) that is available within its confines. No human intervention, other than by the "listener", is required. The songs are permanently on-line.

Suggesting a machine that would keep the "Top 40" data files readily accessible to analysts is

---

1 In fact, this is analogous to the approach taken by the British Broadcasting Corporation's Domesday Project, which was described elsewhere during the conference and with which the British ESRC Data Archive is collaborating.

5 The first optical disks on the market offer a "write once, read many times" facility. Thus, for the moment at least, they are best considered devices for storing stable data. Of course, from an archival perspective, the data security offered by a non-erasable device is a bonus to the mass storage capacity.
not fanciful. In fact, at least one optical disk developer (Philips) supplies a "carousel" option for its "Megadoc" system. Although the system is initially directed to the storage of document images, there appears to be no reason it could not be adapted to numerical data bases.

The juke-box approach to data storage is shared with another recently released mass volume device which is based on densely packed cassette-like tape cartridges. Although these are not transportable in the way that optical disks are, they offer much more storage potential and have to be considered a likely enhancement to the hardware offered by a data library service which wishes to support direct access to its holdings by analysts.

As mentioned, the impact of improved inter-computer communication facilities on archiving is considered later in this paper. For now, it is sufficient to note that the potential for "on-line" access to masses of data which is made possible by the two devices just described will encourage social researchers to explore the use of the developing "network" capability, particularly as improved storage capacity is interacting with a radical change in the overall provision of computers themselves. A brief description of the "new ergonomics" of computer use is a useful prelude to a discussion of the effect of networks on archives.

---

**Computers: a changing style**

Traditionally, social science data archives could assume, reasonably, that their catchment area comprised all computer using social researchers. As computer use in social science was intricately linked to a quantitative orientation, computer users were numerate and usually shared a kit of tools that were applied to research tasks. Moreover, the "conventional" computer oriented social investigator, who was most attracted to the "calculative" power offered by computers, was adequately served by the existing provision of computers in research environments. The computer, physically located in a central position in the institution, was fed numeric data, manipulated them, and then supplied the results of the manipulation. Data archives, which were also centrally located, were well-suited to this mode of computer access and in most countries developed strong institutional ties with the providers of computer services used by the research community. In this way, archives could minimize the technical barriers which inhibited researchers' access to their holdings.

The recent growth of desktop computers, cheap enough to be purchased by individuals, threatens the homogeneity of the computer-using community. A cursory glance at the "micro-computer" marketplace suffices to show that the main appeal of these machines is not that they are superior calculators but that they are remarkably sophisticated typewriters which manage to combine a keyboard, an electronic scissors and a truly non-spill gluepot.

More important, though, for archival development, these desktop machines are changing the prevailing view of what constitutes "machine-readable" data. It does not take long with a "word-processor" to recognize that semi-structured textual information often is more easily organized, manipulated and analysed with the help of a computer than it is manually.

Not surprisingly, facilities offered by desktop workstations are also changing the "traditional" computer analyst's orientation to computerized functions. Granted, quantitative analyses of large data sources are still best done by large, central "mainframe" computers, but the post-processing of the results for research reports is now best accomplished with the software (and sometime hardware) facilities offered on micro-computers. Thus, for example,
the survey researcher will continue to manipulate the survey’s data with a large computer to produce summary information. These results will be captured on the machine on which the report itself is composed. While this might be only to avoid re-typing tables or matrices, the analyst will likely also wish to apply micro-computer facilities like graph processors or spread-sheets to the reduced dataset for further refinement.

Both of these instances of expanded computerization demonstrate an increasing integration of information processing which replaces the earlier compartmentalization of computers by specific tasks. Clearly, computer use is no longer the exclusive prerogative of the specialist in quantitative techniques. This has profound implications for data archives.

First, archives are bound to encounter a new community of users who regard them (archives) as just another source of information. These researchers will have been in contact with electronic publications of other types (e.g. bibliographic search services or reference publications) and will not immediately consider the traditional data archive as being in any way different. Nor should they. It would be odd if the oldest purveyors of computerized information could not service the needs of the newest seekers of that kind of material.

Of course, data archives do not generally provide the summary (digested) style of information that most reference seekers want. However, often that information is available to archives who, however, reject it because it is not their normal stock in trade. In the future, if archives are to serve this new market (which will include a significant proportion of their older market) they will have to make this kind of information accessible.

Second, the integration of information handling practised by the archives’ traditional users will affect what these users expect archives to provide. They too will be less inclined to halt their work progress to permit conventional archive practices to work. They will demand more immediate access to these services, requiring that the archives’ input to their information needs be much more transparent. Archives will (and likely should) become visible only when (infrequent) hitches develop which require intervention.

These developing expectations can be traced to the influence of desktop workstations. However their fulfillment can only be realized by archives if the archives have access to the large scale storage capacity described earlier and if the archives can offer access to these facilities to remote users. Communication networks offer the link.

Computer networks

When computers first became available to researchers employed in the British academic sector, they were provided by individual institutions. In time, an informally organized system of resource sharing developed, wherein some institutions assumed a responsibility to service some of the larger needs of institutions in a particular region. By the end of the 1970’s it was likely that a computer user in any given institution would have access to a larger regional centre as well as to a local center. Indeed, communication with the larger machine was often via the smaller local computer.

Although this arrangement allowed researchers to use much more powerful installations than their own institutions could afford to provide had they stayed totally independent, they still offered a limited access route to the country’s entire community of computers. For the British Data Archive this meant that there was little need to develop a facility that enabled direct
enquiries to its holdings by external users — most users continued to depend on a magnetic tape based service and the Archive was best advised to devote its efforts to improving that mode of data dissemination.

Recent developments in data communication have changed this aspect of the computer user’s working environment. In many countries, most computers are now functionally no further away from the user than the nearest keyboard. In the British academic sector, for example, the Computer Board-sponsored Joint Academic Network ("JANET") offers researchers in that sector an appropriate inter-computer link which facilitates communications among university computers in Great Britain (as well as with other "networks" in Great Britain and abroad).

JANET, as a communication path, meets the need for a facility that allows one computer to talk to many computers. Its more important contribution, however, is to hide the intricacies of network use behind a facade which makes it simple for computer users to address multiple computers with little more knowledge than that required to use their local computers. It accomplishes this with "protocols" which standardize message transmission between sites. At the level of the network, messages may be commands to join a remote computer site as a "local" user or to retrieve information to the user’s own site.

It does not take too much imagination to see what the effect of this new facility for communication with remote sites might be on the operating procedures of data archives. At the least, many analysts will want to interrogate a catalogue of archival holdings to determine which, if any, files contain information of interest to their projects. However, having located pertinent data sources many will wish to select only those parts that are relevant. They may then be happy to "download" the data to their own installations for analysis but, in theory, they could just as easily (or perhaps even with greater ease) analyse the data at the archive’s site, particularly if the archive had implemented specialized software tools to facilitate secondary analysis.

The last paragraph contains an implicit research agenda of projects that are necessary to build the interface for on-line access to archival holdings. An integrated approach to these essential tasks is mentioned in the conclusion to this paper and so the individual tasks need not be dwelt on here. However, at this point it is appropriate to note that the tasks that face an archive also confront any computerized information utility that wishes to encourage direct access to its wares. As these utilities increase in number and as demand for them grows pressure will develop for a coordinated approach to information management on a national (and possibly international) scale which will transcend particular subject orientations. Data archives will have to join these integrated systems — they should be in the forefront of developments. In any event, whatever their institutional inclinations two related features of the environment in which archives now work, the demand for multiple sources and the acceptance of the "relational model" of data management, will push archives in this direction.

### Multiple sources

For years, advocates of secondary analysis as a research strategy for the social sciences, and thus supporters of social data archives, have argued that only this form of research allowed linkage of diverse data sources which was necessary to fully explore social phenomena. However, the records of data archives’ use patterns suggest that these multiple linkages are rarely made — the majority of secondary analysis seem to be of single data sets.

---

*Spring 1986*
There are several reasons for this. The first might relate to the difficulties entailed in merging large masses of data supplied on magnetic computer tape. As described earlier in this paper, the user of archival material often had to request much more than was needed, largely because there were no facilities available for obtaining the subsets that were really required. At this level, it could be that the potential of direct user-archive access will be sufficient to encourage users to pre-process archival material before analysing it.

However the availability of many different kinds of information which are not only in computerized form but which, often, are in only machine readable form will force, or at least teach, researchers to address multiple sources of information. At the onset many of these will be "reference" works which are of interest because they yield independent "facts". However as experience is gained in locating "facts" from several (many?) places and retrieving them for assembly on a single computer, researchers' perspectives will become more ambitious. They will begin to want the same capability of addressing multiple numeric data files which they will tailor to a form which is adequate for reassembly into a purpose-built whole.

Besides the potential for network access and the increasing availability of multiple sources of computerized information, one more development on the computer landscape will have a great impact on archive users' expectations of the type of service that an information utility should provide. Fortuitously, this development, the widespread acceptance of a relational model of data management, also provides users with the tool necessary to take advantage of multiple sources addressed on-line.

The relational data mode

Further commentators on the penetration of computers into "everyday" life during the 1970's and 1980's will highlight the provision of "easy" database management techniques which permit researchers to take multiple logical perspectives of particular group of data. Although several different modelling strategies are available, the "relational" approach to database management offers the most exciting and attractive prospect for social researchers because, among all the alternatives, the relational model most closely replicates the way analysts think about data. It thus offers a tool for analysts interested in analysing substantive problems rather than itself becoming the goal for which analysts strive.

While it is not possible to delve into the details of the model here, it is worth noting that the model's strategy of simplifying the association between discrete sets of data supports the exploitation of multiple data sources when analysing a phenomenon. Most importantly, from an archive's viewpoint at least, it suggests that only those data that are required must be retained when assembling a file for analysis. As suggested earlier, this runs counter to the conventional archive practice of "user takes all", with its demand that the analyst cope with a massive body of unnecessary data.

Strangely, given the esoteric nature of database management, this is the change which could have the greatest single impact on the demands put to archives in the future. The elegance of the relational approach to data management has attracted many micro-computer program developers. Consequently, social researchers who were first introduced to computers via these machines will have experienced "quasi-"relational management systems and will have grown accustomed to applying their power. Moreover, as micros have (until recently) offered only limited data storage capacity, these
new computer users will have learned to work within the confines of these machines. They will not appreciate that moving to larger machines, as they will do when accessing central information utilities, permits a more relaxed view of data storage.

As these new computer users represent the "growth" potential for data archives, their influence on archival development cannot be ignored. Thus it is appropriate that a description of the impact of "new technology" on archives conclude with a speculative note on the most powerful driving force for change, the new user community.

Changing people

It will be evident from the remarks earlier in this essay about prevailing archival practice that users of social science archives almost invariably came from a small segment of the social science community. Oriented toward "quantitative" social research, they grew up with archives and, like archives, learned to accept — and perhaps even like — the "user hostile" environment in which computer users were expected to work. The ethos of computer use has changed and new entrants will be unaware of the need for a hairshirt. Archivists, who tend to be of the old school, will have to adjust their expectations of users to correspond to what their expanded catchment area expects of them.

This new generation of computer users will treat computers with the same ease as they did typewriters a decade ago. For them, the computer is a general utility for a wide range of tasks, among which is information gathering. People accustomed to interrogating a bank account on line or ordering furniture from a direct access shop, will not consider that assembling cross-national data on the association between class and political participation is sufficiently special to warrant the cumbersome hurdles that now impede access to archival data. The archivist must be aware that barriers which reflect past contingencies will direct a major portion of their user community to other information services which offer more flexible access to social information.

Having said this, it must be recognized that the transition from dinosaur to butterfly will not be an easy metamorphosis. One feasible route toward the changeover is described as a conclusion to this paper.

From dinosaur to butterfly: an easier metamorphosis

There is a danger that the earlier discussion which related technological developments and current practice will leave the mistaken impression that archives are unresponsive to change. In practice, archives have worked to incorporate most technological advances into their operating procedures. In the area of networking, one could cite the EEC-sponsored ACCESS project which is designed to produce a cross-national, integrated, bibliographic, on-line data base. The longstanding development of the CESSDA Study Description Scheme fosters the bibliographic control crucial to the identifying sources of comparable data. There have been many examples of archival use of centralized mass storage facilities — for example, the British Data Archive’s distributed arrangements for the supply of the 1981 population census.

However, for all these individual projects, the breakthrough to a comprehensive information service still seems a remote prospect. Although part of the problem is related to archival practices, a significant share of the difficulties are attributable to more general features of
computer use. As these affect all information providers, the removal of these encumbrances on efficient information access requires the development of an integrated system. It is to this joint effort that archives should devote their resources.

In most countries, the computer user can give an empathetic hearing to the tale of the Tower of Babel. While communication utilities like JANET mask the intricacies of making connections between computers, they do little to improve users'-access to different computer systems. In effect, the computer network gets the user to the computer's door but, in most cases, that door is locked against the user's entry unless the user possesses privileged knowledge, to say nothing of privileges.

Prevailing computer practices, which reflect a period when each institution offered its own computer power and each computer manufacturer devised its own operating system, throw up the greatest barrier to a "butterfly-like" access to information. Until this artificial restriction on computer use is overcome, archives and users will be forced to work in an environment in which flexible approaches to information sources are blocked.

However, the obstacle could be removed with a central computer-based facility, accessible to all by network communications, which shields users from different computer environments and protects the environments from many different users. This facility would offer a classified catalogue of all available information sources in the United Kingdom which contained information about the substance of each source and technical information about access arrangements. More importantly, the user would only use the database for subject searches — the technical information, which would be kept transparent to the user, would be used to "automatically" invoke the dialogue necessary to access the host information sites.

Social data archives should be promoting the development of a utility like this. It requires more than a unilateral venture from any single sector and demands more resources than archives themselves can expend. Social data archives, nonetheless, have a privileged role among information providers for they were among the earliest to be computerized. Thus they offer a rare perspective from which to view the changes described in this paper to those with whom they might cooperate.

A central utility like this would benefit social researchers because the only "new" specialist skill required relates to the bibliographic search procedure, which would be common to all sources. It would reward social science because it would allow the exploitation of technological advances which would otherwise be barred to it. It would be attractive to social information providers because they could work to one common standard. It should appeal to current data archives because it promises to provide the protection against the technological "chill" that spelled the dinosaurs' demise.
Promoting a Computer Conference, Continued:
The Experience of the Association of Public Data Users

by Patricia C. Becker
City of Detroit Planning Department

Following publication of Chuck Humphrey’s article in the Summer 1985 issue of this journal, Judith Rowe suggested that readers might be interested in our experience with a computer conference for the Association of Public Data Users (APDU).

APDU is an organization of organizations, rather than of individuals, bringing together people with an interest in the development and use of public data. Because these activities center around the federal government in Washington, the membership is entirely American. Almost everyone involved in APDU uses demographic data from the census, but many other kinds of data are of interest as well, such as economic data, health statistics, and data on specific populations such as the ageing. Members are also interested in the software packages available for processing these data and, increasingly in recent years, in the potential for the use of microcomputers in their everyday work lives. Cross-cutting all of this is a concern with federal statistical policy.

The APDU electronic conference (or e-conf, as we refer to it) was the brainchild of Ken Riopelle, APDU board member. Ken had previous experience as a conference organizer and promoter for the Mott Foundation, experience that included users signing on from around the country. I was a veteran conference participant, but had never been an organizer. The Wayne State University Computing Services Center in Detroit is the "electronic home" for both Ken and I, so it made sense to set up the e-conf there. The software of choice was CONFER, a sophisticated electronic conferencing package developed for use on the MTS operating system.

The original proposal called for a pilot project, for Board members and a few others. This was a group of about 13 people. We began in the spring of 1984. A project account was established at the Wayne State Computing Services Center, the e-conf itself was created, and sign-on materials were sent to each member of the group.

The initial group was registered into the userdirectory, and materials on the electronic messaging system were provided as well as on CONFER. Since all were coming in on Telenet...
or Autonet, the appropriate phone numbers were provided to each prospective participant. A wallet-sized "1-2-3" crib sheet with sign-on instructions was provided. In addition, a sample session of CONFER was created, printed and reproduced.

How well did it work? Of the initial group of 13 people, seven (including Ken and myself) became active participants. Active is defined, here, as signing on at least once a month. Three board members signed on, joined, but rarely participated; the remaining three never actually joined the e-conf at all.

Two major factors seem to explain non-participation; to some degree they are interactive, one with the other. One was lack of equipment, and the other was a lack of of familiarity with using computer terminals. It appears to be necessary, to maintain active participation, to have a computer terminal available in the office, and to be in a position to use it frequently for other, routine work activities. Most non-participants either had no access to equipment or had access only at home.

However, the seven people who did participate had a great time. Items were entered on internal APDU Board issues, on federal information policy, on software and data access, and on use of the e-conf itself. Enough was going on to keep people interested, so that activity levels did not drop among the seven active participants.

After evaluation of the pilot project, the Board decided to extend the e-conf to the membership. To promote interest, a $20 credit was offered to each organizational member. In the APDU membership structure, each member organization has a primary representative who is responsible for the dues; additional representatives can be added for a small fee.

The $20 credit allowed members to sign up and try out the conference at no cost. Expenditures in excess of $20, however, had to be paid for "up front", so that APDU would not get into financial difficulty. Additional representatives were welcome to participate as well, but were required to arrange for funding from their primary organizational representatives. The Computing Center's accounting system allows us to control the amount of money available to each sign-on ID, so there was no problem managing the accounts.

In January 1985, the entire membership received a mailing which explained the opportunity to participate in the e-conf and included forms for signing up. Unfortunately, the response can best be described as underwhelming. Between the time of mailing and the annual "people" conference in October, only ten sign-on IDs were issued; of these, only three or four became active participants. We picked up a few more as a result of heavy promotion at the conference in October, as well as having added some new board members.

At this point, in January 1986, 37 sign-on IDs have been issued but only 14 can be described as active participants. The number of items has grown to 121 and discussions continue to be lively. A great deal of information is being exchanged. Several specific decisions have been made "electronically". Draft resolutions, letters of comment, and the like, representing proposed positions for APDU to take vis-à-vis the federal statistical establishment, have been put into items to be reviewed by the Board and other interested parties.

And the cost? It appears that 1985 expenditures will be under the budgeted figure of $2500, primarily because there are fewer members than anticipated taking advantage of the $20 credit. By budget category, costs can be broken down as follows:

- E-conf maintenance: disk space, on-line time for organizers, printing manuals, and project accounting. These costs would
have been higher had the organizers not had other Wayne State accounts through which to participate and do some of the maintenance work. $425

- Use of system by executive secretary. $425

- Non-reimbursed use by Board members. Those who are in work situations in which it is difficult to obtain reimbursement are allowed unlimited access at present. This item also includes use of the system by the organizers of the annual conference, for which two people in different cities accomplished most of their work via electronic messaging. $950

- $20 credits (excluding accounts in the previous item). $250

Overall, the consensus of the Board is that the expenditure is worthwhile and justifiable within our total budget scenario. There has been some criticism of the project within the APDU membership, primarily of the cost to the members. Some feel that they cannot afford participation (since $20 really doesn’t go that far on Telenet or Autonet). As in any organization, members belong for different reasons and have different agendas; not all our members find interaction with other members to be a useful expenditure of their time and/or money.

There also remains a significant problem with equipment access – several members who wish to participate have been stymied by the lack of a terminal, or a good modem and communication software. This problem should decrease over time, as more and more organizations acquire microcomputers. We are planning to have an on-line demonstration of the system at our next annual conference (scheduled for October 1986), to promote interest in the system. We are also including a column entitled "From the E-Conf" in our monthly printed newsletter, both to provide information for those who are not conference participants and to encourage them to join.

There is another participation problem of a different kind: some participants sign on regularly, but rarely contribute any responses. This is the reverse of the "habitual commentator" described in the Humphrey article\(^2\). Two factors are at work here: one technical and one involving personalities. Many participants sign on and simply "dump" the conference activity to disk, to be read later, or to a printer, without actually reading it while they are signed on. This has the negative effect of discouraging responses, since they must sign on again to enter them. The other factor is, as Humphrey\(^3\) described, the "implicit norm to say nothing." However, I think this is less a problem in our particular conference than it might be in others. The fact that most of the participants have met each other "in person" at the annual conference helps – we generally know the people to whom we are talking electronically.

All in all, APDU rates the e-conf a success and it has become an integral part of the organization’s functional mechanism. We would like to see greater participation and will continue our efforts in that direction. What we have, though, is a system that works – people are communicating, and that’s what an organization is all about.\(^*\)

\(^2\) ibidem

\(^3\) ibidem
Topically-focused Data Archives: A New Paradigm for the Codification of Social Science Research

by Josefina J. Card
President, Sociometrics Corporation
3191 Cowper Street
Palo Alto, California 94306
Tel. (415) 321-7846

The "information explosion" has become a distinguishing feature of modern science. Both the published scientific literature and its supporting data files continue to grow at unprecedented rates. More than ever, it has become important that efficient ways be found to store available information on a given topic and then retrieve relevant portions of that information as they are required. In the 1970's enormous progress was made in the development of procedures to store and retrieve bibliographic information. The DIALOG, ERIC, and MEDLARS databases are but a small sample of the growing number of computerized bibliographic databases available to social scientists. Less significant progress has been made in the development of analogous procedures to store, catalog, and retrieve elements common to the numeric (or raw data) information underlying the published studies. Enormous productivity and cost savings could result from such development. For little additional cost relative to the data collection costs already incurred, a substantively-focused data archive with indexing capabilities could: accelerate the growth and dissemination of scientific knowledge about a topic of contemporary interest; encourage corroboration and replication of newly reported findings; provide policymakers and practitioners with a larger scientific base on which to build their work; and stimulate investigations by new investigators without access to the substantial funds required for new data collection. This paper introduces the Data Archive on Adolescent Pregnancy and Pregnancy Prevention (DAAPPP), to illustrate features of an emerging information resource: the special-purpose social science data archive.

The accumulation of knowledge about human reproduction, coupled with the development of relatively safe, effective, and inexpensive contraceptive methods, has made it possible for human beings to seize control of their biological destinies, and to plan the size and spacing of their families. Differences continue to persist,
however, in the degree to which various groups of people have been able to avoid unplanned and unwanted pregnancies. Rates of unplanned and unwanted pregnancies are higher in the developing than in the developed world. In a given country, young unmarrieds and the socially and economically disadvantaged have generally been found to be more vulnerable. The rate of out-of-wedlock pregnancy and childbearing among U.S. teenagers is among the highest in the world. DAAAPP was established by the U.S. Office of Population Affairs of the Office of the Assistant Secretary for Health to encourage the conduct and dissemination of research on these important social issues.

DAAAPP identifies, selects, acquires, and archives the most valuable databases dealing with U.S. adolescent fertility and U.S. family planning. Database identification refers to the systematic identification of all machine-readable data sets capable of addressing these topics. Database selection refers to the selection, from the identified universe, of the most outstanding data sets to include in the collection. Technical quality, substantive scope, and policy relevance are considered simultaneously by a National Advisory Panel of scientists in making selection decisions. Database acquisition refers to obtaining selected data sets from their holders. The raw data, the documentation, and completed reports and publications are all acquired. Database archiving refers to the processing and documentation of acquired data sets by archive staff, so that standardized, easy-to-use products are produced and disseminated. DAAAPP then makes its data and documentation publicly available, for the cost of reproduction. The following products are now publicly available for each of the 45 data sets currently in DAAAPP (see Table 1):

- A computer tape for use with mainframe computers with two machine-readable files:
  * the raw data;
  * SPSS-X program statements to convert the raw data to an SPSS-X system file (SPSS is an acronym for the widely-used Statistical Package for the Social Sciences).

- Floppy diskette(s) for use with microcomputers (in either 360-kilobyte or 1.2-megabyte format), with two machine-readable files:
  * the raw data;
  * SPSS/PC program statements to convert the raw data to an SPSS/PC microcomputer system file.

- A printed and bound user's guide, with five standard sections:
  * an overview of the original purpose for which the data were collected, and a description of the file's processing history;
  * a description of the machine-readable files available for the data set;
  * a categorization of the variables included in the data set by their topic and type, followed by a listing of all variables, sorted by topic and type;
  * a report on the completeness and quality of the data;
  * a bibliography of representative publications based on the data set.

- The codebook and instrument from the original investigation, where available.

Users of statistical package programs know that part of the routine procedure in the development of system files for analysis is the assignment of names and labels to variables in the file. We use the output of this routine procedure—lists of variable names and values—in an innovative way to give the archive indexing capabilities.

Each variable in DAAAPP is given an eight-character name for use with SPSS-X or SPSS/PC, to standardize variable names across all files, and to provide the user with quick reference to certain useful information about the
variable. Characters 1–2 encode the variable's TOPIC, the main subject matter of the variable. Character 3 encodes the variable's TYPE, further classifying the subject matter into one of the many variable types commonly used by social scientists. Characters 4–5 are a reference to the DATA SET ID, indicating the original source of the data. Characters 6–8 contain the VARIABLE SEQUENCE NUMBER, indicating the sequential position of the variable within the source data set. Table 2 contains the list of TOPICS, Table 3 the list of TYPES.

Definitions of each topic and type have been developed that provide an inter–rater categorization reliability of over 90%. The list of DATA SET IDs is shown in Table 1. Each list can be altered easily to suit archives focused on other substantive topics. The variable naming scheme encodes information both on what each variable has in common with other variables in the archive (its TOPIC and TYPE), and what is unique to the variable (its SEQUENCE NUMBER with a given DATA SET ID).

DAAPPP staff members have written a simple computer program that uses the TOPIC and TYPE characters of the variable names as input, to produce a matrix that depicts, at a glance, the topical emphasis of each data set. For example, DAAPPP data set no. 2 is the 1976 U. S. National Survey of Young Women (John Kantner and Melvin Zelnik, John Hopkins University, principal investigators). Table 4 contains the Topic–by–Type Matrix for this data set. The matrix allows the user to see at a glance where the "areas of richness" of the data set lie. For example, we can see in the last column of Table 4 that this particular data set has a total of 386 variables; the data set is rich in information on family characteristics (142 variables), contraceptive information (56 variables), and child–bearing related information (42 variables). There are seven items relating to abortion, the first topic in the alphabetically–ordered topic list. The first row of Table 4 shows that all seven of these variables are attitudinal.

Information of the type contained in Tables 4 and 5 can be extremely helpful in ascertaining whether a given data set can be used to answer a particular research question. It is important to note that virtually no extra processing time, beyond the routine procedures used by any social scientist to create an SPSS–like set–up for his data set, is required in order to produce and display the information.

When variable names and labels from all the data sets in DAAPPP are used as input, the same program provides a matrix and variable listing that depicts the State–of–the–Archive. The 45 data sets currently in the DAAPPP collection contain 14,216 variables, characterized as shown in Table 6. While a listing of these 14,216 variables is too long to print here, such a list exists, is publicly available (for the cost of reproduction), and is updated quarterly.

Although the DAAPPP project is by no means over (the collection is currently growing at the rate of about five data sets per quarter), we see from Table 6 that there appears to be relatively little empirical data on important topics such as sexually transmitted disease and substance abuse (in the context of adolescent pregnancy studies). At the end of the DAAPPP contractual period in September 1987, we will be in a position to evaluate the amount and types of information available on adolescent pregnancy, pregnancy prevention, and family planning, and to identify significant gaps.

Social science archival data can be used in many different ways: (1) for secondary analysis (the analysis of data for purposes other than those for which the information was originally collected); (2) for meta analysis (the analysis of data common to a number of data sets to investigate similarities and differences in the patterning of relationships); (3) for longitudinal analysis of panel data (such as that found in DAAPPP Data Sets 20–24, the National
Longitudinal Survey of Youth); (4) for cross-sectional trend analysis of related surveys (such as analysis of trends in information found in DAAPPP Data Sets 11-18, the 1977, 1980, and 1982 Current Population Surveys); (5) for provision of contextual variables to add to an individual-level data file (for example, one could add all or part of the information contained in DAAPPP Data Set 8 on State Policy Determinants of Teenage Childbearing to one's individual-level data file to study the additive and interactive effects of individual versus environmental factors in producing fertility-related behavior); (6) for derivation of comparison group data against which to compare data from clinic patients or service program participants; and (7) for instructional purposes, as an exciting aid in the teaching of statistics and research design.

It is our hope that those interested in studying problems of adolescent pregnancy and family planning will use DAAPPP, and that the DAAPPP experience will be helpful in stimulating and facilitating the formation of other, special-purpose data archives containing the best scientific data on important issues facing us all.

Acknowledgements

The Data Archive on Adolescent Pregnancy and Pregnancy Prevention is funded by Contract 282-84-0083 between the Office of Population Affairs, Office of the Assistant Secretary for Health, and Sociometrics Corporation (J. J. Card).
Table 1

| Data Set Id | Data Set Name (Investigators)                                                                                                                                                                                                 |
|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 01          | 1971 U.S. National Survey of Young Women: Selected Variables (M. Zelnik & J.F. Kantner)                                                                                                                                             |
| 02          | 1976 U.S. National Survey of Young Women (J.F. Kantner & M. Zelnik)                                                                                                                                                                |
| 03          | Project TALENT: Consequences of Adolescent Childbearing for the Young Parents' Future Life, 1960-1974 (J.J. Card)                                                                                                                      |
| 04          | Detroit Mother-Daughter Communication Patterns: Mother File, 1978 (G.L. Fox)                                                                                                                                                       |
| 05          | Detroit Mother-Daughter Communication Patterns: Daughter File, 1978 (G.L. Fox)                                                                                                                                                      |
| 06          | Philadelphia Collaborative Perinatal Project: Economic, Social, and Psychological Consequences of Adolescent Childbearing, 1959-1965 (J. Marecek)                                                                                           |
| 07          | Nashville General Hospital Comprehensive Child Care Project, 1974-1976: Selected Variables (H> M > Sandler)                                                                                                                       |
| 08          | State Policy Determinants of Teenage Childbearing, 1979 (K.A. Moore)                                                                                                                                                              |
| 09          | 1980 U.S. Survey of Services Provided by Adolescent Pregnancy Programs (JRB Associates)                                                                                                                                             |
| 10          | 1982 Evaluation of DAPP Adolescent Pregnancy Programs (M. Burt)                                                                                                                                                                      |
| 11          | 1980 U.S. Current Population Survey: Selected Variables -- Women (Bureau of the Census)                                                                                                                                              |
| 12          | 1980 U.S. Current Population Survey: Selected Variables -- Men (Bureau of the Census)                                                                                                                                               |
| 13          | 1980 U.S. Current Population Survey: Selected Variables -- Children (Bureau of the Census)                                                                                                                                             |
| 14          | 1982 U.S. Current Population Survey: Selected Variables -- Women (Bureau of the Census)                                                                                                                                              |
| 15          | 1982 U.S. Current Population Survey: Selected Variables -- Men (Bureau of the Census)                                                                                                                                               |
| 16          | 1982 U.S. Current Population Survey: Selected Variables -- Children (Bureau of the Census)                                                                                                                                              |
| 17          | 1977 U.S. Current Population Survey: Selected Variables -- Women (Bureau of the Census)                                                                                                                                              |
| 18          | 1977 U.S. Current Population Survey: Selected Variables -- Men (Bureau of the Census)                                                                                                                                               |
| 19          | First U.S. Health and Nutrition Examination Survey (HANES), 1971-1975 (National Center for Health Statistics)                                                                                                                       |
| 20-24       | National Longitudinal Study of Youth (NLSY), 1979-1982: Selected Variables (Waves 1-4), and Supplementary Variables (Ohio State University)                                                                                                       |
| 24          | 1981 U.S. Survey of Title X - Funded Family Planning Clinics (R. Herceg-Baron)                                                                                                                                                       |
| 26          | 1982 National Survey of Family Growth (NSFG), Cycle III -- Women Aged 15-44 (National Center for Health Statistics)                                                                                                                   |
| 27          | 1982 National Survey of Family Growth (NSFG), Cycle III -- Women Aged 15-44 (National Center for Health Statistics)                                                                                                                   |
| 28          | 1979-1980 U.S. Survey of Unmarried Women Under 18 in Family Planning Clinics (A. Torres)                                                                                                                                             |
| 29          | Effects of Organized Family Planning Programs on U.S. Adolescent Fertility (J.D. Forrest)                                                                                                                                                |
| 30          | Johns Hopkins Study of Repeat Adolescent Pregnancy, 1976-1982 (J.B. Hardy)                                                                                                                                                           |
| 31          | 1972-74 Ventura County of Unmarried Pregnant Women aged 13-20 (M. Eisen)                                                                                                                                                             |
| 32          | 1982 San Jose, California Study of Adolescent Perinatal Risk Behavior (P.A. Hensleigh & N. Moss)                                                                                                                                        |
| 33          | 1981-1982 Evaluation of DAPP Adolescent Pregnancy Programs: Individual Level Data I (M.R. Burt)                                                                                                                                        |
| 34          | 1981-1982 Evaluation of DAPP Adolescent Pregnancy Programs: Individual Level Data I (M.R. Burt)                                                                                                                                        |
| 35          | 1979-1981 Philadelphia Study of Psychological Factors Associated With Adolescent Fertility Regulation -- Females (E.W. Flaherty & J. Marecek)                                                                                              |
| 36          | 1979-1981 Philadelphia Study of Psychological Factors Associated With Adolescent Fertility Regulation -- Males (E.W. Flaherty & J. Marecek)                                                                                                 |
| 37-38       | The National Survey of Children, 1976 (Child Trends, Inc.)                                                                                                                                                                              |
| 39          | Florida-Puerto Rico Study of Adolescent Pregnancy and Neonatal Behavior, 1978 (B.M. Lester)                                                                                                                                            |
| 40          | Maricopa County, Arizona Study of Child Maltreatment Risk Among Adolescent Mothers, 1976-1978 (F.C. Bolton, Jr.)                                                                                                                        |
| 41          | 1955 Growth of American Families: Married Women (A. Campbell, P.K. Whelpton, & J.E. Patterson)                                                                                                                                         |
| 42          | 1955 Growth of American Families: Single Women (A. Campbell, P.K. Whelpton, & J.E. Patterson)                                                                                                                                           |
| 43          | 1960 Growth of American Families (A. Campbell, P.K. Whelpton, & J.E. Patterson)                                                                                                                                                        |
| 44          | 1979 U.S. National Survey of Young Women (M. Zelnik & J.F. Kantner)                                                                                                                                                                   |
| 45          | 1979 U.S. National Survey of Young Men (M. Zelnik & J.F. Kantner)                                                                                                                                                                     |
Table 2

| Code | Topic                          | Code | Topic                          |
|------|--------------------------------|------|--------------------------------|
| AB   | Abortion                       | MP   | Marriage patterns              |
| AC   | Agency Character               | MH   | Mental health                  |
| AD   | Adoption                       | ME   | Meta level                     |
| AG   | Age                            | NU   | Nutrition                      |
| BF   | Biological function            | OC   | Occupation and development     |
| CB   | Childbearing                   | OT   | Other                          |
| CR   | Childrearing                   | OW   | Out-of-wedlock parenthood      |
| CL   | Clinical activities            | PE   | Personality                    |
| CM   | Communication                  | RA   | Race/ethnicity                 |
| CN   | Contraception                  | RC   | Recreation                     |
| CI   | Crime                          | RL   | Religion                       |
| ED   | Education                      | RS   | Residence/Location             |
| FH   | Family and household           | SE   | Sex education characteristics  |
| FS   | Friends and social             | SX   | Sexuality activities           |
| GR   | Gender and gender              | SD   | Sexually transmitted role disease |
| GC   | Guidance and                   | SA   | Substance abuse counseling     |
| HL   | Health                         | UN   | Undocumented                   |
| IN   | Intellectual function          | WF   | Wealth, finances, and material things |
| IV   | Interview                      |      |                                |

Table 3

| Code | Topic                          |
|------|--------------------------------|
| A    | Attitudes                      |
| B    | Behavior                       |
| C    | Cognitions                     |
| E    | Emotions                       |
| H    | History                        |
| I    | Intentions                     |
| M    | Motivations                    |
| O    | Other                          |
| P    | Program/Policy                 |
| R    | Reasons                        |
| S    | Status                         |
| T    | Traits                         |
| U    | Undocumented                   |
| X    | Meta                           |
| Y    | Aggregate                      |
| Z    | Household                      |
Table 4

OVERVIEW OF CONTENTS, THE 1976 NATIONAL SURVEY OF YOUNG WOMEN

| TOPIC                      | ATTITUDE | BEHAVIOR | COGNITION | HISTORY | INTENTION/MOTIVATION | OTHER | REACTION | STATUS | META | TOTAL |
|----------------------------|----------|----------|-----------|----------|----------------------|-------|----------|--------|------|-------|
| AGE                        | 0        | 0        | 0         | 2        | 0                    | 0     | 0        | 0      | 0    | 4     |
| MARRIAGE                   | 1        | 1        | 1         | 25       | 2                    | 0     | 0        | 0      | 151  | 142   |
| RELIGION                   | 1        | 0        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 1     |
| EDUCATION                  | 1        | 0        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 1     |
| OCCUPATION                 | 1        | 0        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 1     |
| OTHER                      | 0        | 0        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 2     |
| RACE ETH                   | 0        | 1        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 3     |
| RELIGION                   | 1        | 1        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 4     |
| OCCUPATION                 | 1        | 0        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 1     |
| FED EDUC                   | 1        | 0        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 1     |
| CIVIL RIGHTS               | 1        | 2        | 0         | 0        | 0                    | 0     | 0        | 0      | 0    | 4     |
| TOTAL                      | 24        | 7        | 9         | 144       | 3                    | 1     | 20       | 186    | 18   | 375   |
### Table 5

**List of Variables, by Topic.**

**National Survey of Young Women**

**Page 1 of 10**

| NEWID | OLDID | TYPE | LABEL |
|-------|-------|------|-------|
| ABAD2110 | ATTITUDES | ABORT OK IF THE WOMAN HAD BEEN RAPED |
| ABAD2110 | ATTITUDES | ABORT OK FOR VERY YOUNG PERSON |
| ABAD2112 | ATTITUDES | ABORT OK IF PG ENDANG WOMANS HEALTH |
| ABAD2113 | ATTITUDES | ABORT OK IF CHILD BORN DEFORMED OR MENTLY DEFIC |
| ABAD2114 | ATTITUDES | ABORT OK IF THE WOMAN COULDN'T AFFORD IT |
| ABAD2115 | ATTITUDES | ABORT OK ANY REASON IMPORTANT TO HER |
| ABAD2116 | ATTITUDES | VIEWS ABOUT HAVING AN ABORTION |

| NEWID | OLDID | TYPE | LABEL |
|-------|-------|------|-------|
| ADAD2099 | ATTITUDES | IF UNABLE HAVE WANTED CHIDLREN, WLD ADOPT? |
| ADAD2100 | ATTITUDES | WOULD R ADOPT CHILD INSTEAD OF HAVING OWN |

| NEWID | OLDID | TYPE | LABEL |
|-------|-------|------|-------|
| AGHO2021 | HISTORY | YR OF BIRTH |
| AGHO2022 | HISTORY | MONTH OF BIRTH |
| AGSO2003 | STATUS | AGE |
| AGSO2061 | STATUS | SCREEN--AGE |

| NEWID | OLDID | TYPE | LABEL |
|-------|-------|------|-------|
| BFHO2132 | HISTORY | AGE LIST PERIOD |

| NEWID | OLDID | TYPE | LABEL |
|-------|-------|------|-------|
| CBAD2101 | ATTITUDES | IDEAL AGE FOR A GIRL TO HAVE 1ST BABY |
| CBAD2135 | COGNITIONS | KNOW WHEN PREG IS MOST LIKELY TO OCCUR |
| CBHO2006 | HISTORY | PREGNANCY STATUS AT MARRIAGE IND |
| CBHO2230 | HISTORY | EVER BEEN PREGNANT |
| CBHO2232 | HISTORY | NUMBER OF PREVIOUS PREGNANCIES |
| CBHO2234 | HISTORY | OUTCOME OF 1ST PG AT MARRIAGE IND |
| CBHO2237 | HISTORY | WHAT FIRST PREGNANCY THINK GOOD CHANCE |
| CBHO2238 | HISTORY | YR OF OUTCOME 1ST PG |
| CBHO2239 | HISTORY | MONTH OF OUTCOME 1ST PG |
| CBHO2240 | HISTORY | AGE AT OUTCOME 1ST PG |
| CBHO2242 | HISTORY | OUTCOME 2ND PG |
| CBHO2243 | HISTORY | WHAT 2ND PREGNANCY THINK GOOD CHANCE |

---

*Spring 1986*
### Table 6

**THE STATE-OF-THE-ARCHIVE (9/85)**

| Topic | Type | ATTITUDE | BEHAVIOR | COGNITIVE | EMOTIONS | HISTORIC | INITIATION | INFORMATION | OTHER | POLICY | READINESS | STATUS | STYLIST | TRAITS | UNCONGRUENT NOTA | AISSISTIVIZATION | RESISTANCE | TOTAL |
|-------|------|----------|----------|------------|----------|----------|------------|-------------|-------|--------|-----------|--------|---------|--------|-----------------|----------------|------------|-------|
| ADOPTION | 24 | 3 | 56 | 0 | 74 | 3 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 21 | 18 | 0 | 217 |
| ADULT EDUC | 0 | 0 | 0 | 0 | 45 | 0 | 0 | 0 | 24 | 1 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 212 |
| ADOPTION | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 24 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 22 |
| AGE | 0 | 0 | 0 | 0 | 55 | 0 | 0 | 0 | 0 | 25 | 0 | 15 | 0 | 0 | 0 | 8 | 27 | 0 | 212 |
| BIG PICT | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 43 | 0 | 0 | 475 |
| CHILDREN | 15 | 0 | 11 | 0 | 23 | 0 | 70 | 79 | 58 | 101 | 78 | 149 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1762 |
| CHIC | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| CLINICAL | 0 | 0 | 0 | 0 | 853 | 0 | 0 | 0 | 0 | 85 | 1 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| DEMOCR | 22 | 0 | 107 | 0 | 26 | 33 | 25 | 0 | 0 | 25 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 251 |
| GUNMETAL | 10 | 0 | 81 | 0 | 34 | 1 | 169 | 32 | 0 | 56 | 15 | 3 | 0 | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 1771 |
| CHILDREN | 15 | 0 | 29 | 0 | 25 | 0 | 5 | 12 | 32 | 36 | 52 | 7 | 94 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 626 |
| EDUCATION | 19 | 0 | 20 | 0 | 2 | 16 | 155 | 26 | 18 | 0 | 136 | 60 | 179 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 742 |
| FAMILY CAFE | 38 | 0 | 107 | 0 | 43 | 14 | 119 | 11 | 25 | 0 | 95 | 0 | 1080 | 3 | 0 | 20 | 1 | 0 | 0 | 0 | 1447 |
| FRIENDS | 23 | 0 | 76 | 0 | 2 | 71 | 0 | 9 | 0 | 0 | 0 | 3 | 82 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 308 |
| GUIDANCE & COUNSEL | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| GENDER | 57 | 0 | 11 | 0 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| HEALTH | 0 | 0 | 0 | 0 | 6 | 0 | 129 | 0 | 0 | 0 | 26 | 2 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| INTELLIGENCE | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 24 | 0 | 19 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 172 |
| INTERVIEW | 2 | 0 | 24 | 0 | 12 | 0 | 2 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 62 |
| MNA | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| META-LEAD | 0 | 0 | 1 | 0 | 1 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| MARRIAGE | 17 | 0 | 4 | 0 | 2 | 372 | 17 | 0 | 0 | 23 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 152 |
| NUTRITION | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| OCCUPATIONAL | 42 | 0 | 28 | 0 | 20 | 0 | 67 | 69 | 0 | 62 | 0 | 161 | 263 | 0 | 0 | 0 | 2 | 23 | 0 | 0 | 1572 |
| OTHER | 0 | 0 | 5 | 0 | 7 | 0 | 6 | 0 | 0 | 0 | 1 | 0 | 22 | 0 | 0 | 0 | 7 | 2 | 0 | 0 | 0 | 0 |
| OUT-HELPS | 4 | 0 | 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PERSONALITY | 14 | 0 | 44 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| PAGE SMS | 0 | 0 | 0 | 0 | 0 | 0 | 16 | 0 | 0 | 22 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| EDUCATION | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| RELIGION | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| RIOT RESEARCH | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| RISK | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SET BACK | 20 | 0 | 29 | 0 | 14 | 0 | 92 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 265 |
| SOCIAL ROLE | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SOCIAL STATUS | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| SOCIAL SUPPORT | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| TOTAL | 448 | 0 | 985 | 0 | 507 | 189 | 5031 | 237 | 706 | 2 | 3101 | 545 | 2052 | 158 | 0 | 17 | 595 | 241 | 3 | 1474 |

*Spring 1986*
The Potential for Computer Communications Among ICPSR Representatives

by Charles Humphrey
Computing Services
University of Alberta

Introduction

Availability of computer-based communication, especially electronic mail and computer conferencing, has become commonplace on most North American campuses. Through such technology, staff at many universities can now communicate with their colleagues at other institutions as easily as they do with those on their own campus. The promise of computer communications lay in facilitating scholarly and professional exchanges which are immediate, easy, inexpensive and widespread.

However, even though this technology has become extensively accessible, obstacles do exist which impede its application. A recent study revealed that among the important factors determining the usage of computer conferences was the prominence of a terminal within a person's immediate work environment. The most active members of the conference of study were those who regularly used a terminal during their daily routine and whose equipment permitted the use of packet-switching networks.

On the basis of these findings and with the advent of the Consortium Data Network (CDNet), a survey was conducted of the participants attending the 1985 biennial meeting of the Official Representatives (OR's) to the Inter-university Consortium for Political and Social Research (ICPSR).

A survey of official representatives

Immediately prior to the ICPSR business meeting, a questionnaire was distributed which focused on two topics. The first of these dealt with the availability and use of terminals in the

---

1. Charles Humphrey and Wendy Watkins, "DataLink: A Computer Conference for Canadian Data Libraries and Archives," Report to the Social Sciences and Humanities Research Council of Canada, p. 24ff.
2. Terminal is used here to refer to the same range of I/O devices that the term workstation has come to denote, which covers everything from teletypes to visual display units to microcomputers. However, since items in the questionnaire referred to terminals, we will continue to use this term below.
3. Ninety-one questionnaires were collected from the 125 representatives at the meeting, providing a 73% response rate. Two factors, however, must be considered when generalizing from this poll. First, 30% of the participants were substitutes for Official Representatives. Thus, generalized comments from this sample encompass more than just OR's. Secondly, only 46% of the ICPSR membership were in attendance.
OR's workplace; while the second sought some indication of the scope of experience that representatives have had with computer communications. This paper reviews specific characteristics of the ORs and examines how prepared this group is to avoid or overcome the impediments to active computer-based communications.

The availability and use of computer terminals

Lack of convenient access to a terminal does not appear to be a major problem for the vast majority of ICPSR representatives. Nearly 80% of the respondents have immediate access to a terminal at work (see Table 1), and 48% have terminals both at home and work. Only 14% do not have any convenient access, and of these thirteen respondents, eleven have a terminal available to them either on the same floor as their office or on another floor in the same building.

Having a terminal at your fingertips does not necessarily ensure use of the device. However, as shown in Table 2, a clear relationship between access and use does exist in this data. Those with a terminal immediately available to them during the workday report the highest usage rates. Examining the breakdown across the categories of access, the proportion of those using a terminal several times a day declines monotonically as one moves from those with the highest degree of immediate access to those with no terminal directly available.

The obvious conclusion is that most respondents make use of the equipment that they have. However, this is not necessarily the most significant conclusion. More important is the summation that a computer terminal is an integral tool in the work routine of a large majority of the ICPSR representatives. Over 70% indicated that they make use of a terminal throughout their workday.

Important characteristics for computer communications

A few special features are desirable for the effective use of computer mail or conferencing systems. One feature is the capability of placing a call and making a connection with a central computer system and its mail or conferencing software. This type of terminal connection usually is supported by a modem attached to a standard telephone outlet. Such a configuration permits a user to call either their local computer system or a packet-switching network through which a myriad of computer systems are available. Some terminals, however, are directly wired to a central computer. In such instances, the use of packet-switching networks is dependent upon a call-out facility on the mainframe. Regardless of whether the terminal connection is through a modem or a mainframe call-out facility, the most flexible situation for the user is to be able to logon to the computer system housing the mail or conferencing software.

Of the survey respondents, 44% have a terminal with dial-out capabilities at work (see Table 3). When those who have a terminal at home only are included in the group with dial-out capability, the overall percentage increases to 54%. Furthermore, a call-out facility was present on the central computer systems of over 70% of the respondents. These figures reveal that a majority of the respondents have available some form of call-out facility which would permit them to connect to the Consortium network.

Another characteristic which encourages the use of computer communications is the availability
of a full-screen editor. The backbone of computer communications is the typed word, and the ease with which text can be entered and modified significantly influences the amount of text contributed. Just as was the case with dial-out facilities, respondents seem to have ready access to full-screen editors whether at home or work (see Table 4). Eighty-eight percent of those with terminals at work have such an editor; 82% with terminals at home also have one available.

Experiences with computer communications

Two-thirds of the respondents reported that they had made use of at least one of the three communication methods — electronic mail, computer conferencing and networks (see Table 5). Nearly half (47%) had experience with more than one of these methods. In fact, those saying that they had used both networks and electronic mail constituted the largest single group (27%). Considering the three electronic media separately, 58% of the respondents noted some experience with electronic mail; 57% had used a network; only 20% had tried a computer conference. In terms of overall exposure, one in five indicated experience with all three methods.

Experience with these communication methods clearly varied by type of terminal access. Eighty-one percent of those who have a terminal both at work and home have had experience with at least one of the three methods (see Table 6). Antithetically, 70% of those without immediate access to a terminal indicated that they had no experience with any of the three communication methods. The difference between these two groups accentuates the gap that exists between those who have a terminal at their fingertips and those who do not.

In comparing the remaining two groups, the percentage of those having worked with at least one of the communication methods was virtually the same, 68% for those with a terminal at work only and 67% for those with a terminal at home only. The experience levels of these two groups are much closer to the group with terminals at both work and home. One interesting difference is that a higher proportion of those with only a terminal at home had tried two or more of the communication methods.

An indication of the extent to which these three types of communication have been incorporated into the work routines of the respondents is shown in Table 7. Nearly one-third of those making use of electronic mail check it on a daily basis and over 60% access it twice a week or more. Similarly, 22% of those belonging to a computer conference use that medium daily, while only 10% of network users use the network that frequently. Electronic mail is clearly leading the way among these three methods of communication; and with the introduction of electronic mail service between universities, daily use of electronic mail will undoubtedly increase. Its popularity is exemplified by the fact that 63% of those who use electronic mail daily also reported using BitNet, which is an inter-university electronic mail service.

---

4 Respondents may have been confused about the difference between a carrier network such as Telenet and an application network such as BitNet. The former is a service which allows one to dial a local telephone number and to connect as a remote terminal to a computer system, while the latter type of network refers to special application software making use of packet-switching technology to transmit information between sites. The item in the questionnaire was supposed to identify those who had experience with a carrier network.
Conclusion

Given both the availability of terminals to ICPSR representatives and their experiences with computer communications, what are the implications for the Consortium Data Network (CDNet)? The profiles described above point to a couple of possibilities. First, slightly more than half the respondents possess the proper mix of both equipment and experience, thus making the likelihood that they will use CDNet very high. Fifty-two percent of the respondents reported immediate access to a terminal and indicated experience using a network. This is a significant group, since access to CDNet depends upon a remote terminal connection through a packet-switching network such as Telenet. Secondly, an additional 12% have both a terminal available and some experience with electronic mail or computer conferencing. Assuming that some experience with either of these media develops skills that are easily transferable to the use of networks, this group should also readily use CDNet. Thus, 64% of the respondents appear to possess essential equipment and skills to use CDNet without major obstacles.

An additional 21% of the respondents have ready access to terminals but no experience with the three methods of computer communication. Consequently, this group faces the task of learning some new computing skills. An important factor in this regard will be motivation. Motivating people to use any of the three communication media, even when they already possess the necessary skills, is in itself a challenge. Thus, initiating a service such as CDNet is further complicated by the need to motivate first time users to acquire the additional skills. No data was collected in this survey to indicate directly how significant a factor motivation will be.

Factors other than the computing skills and motivation levels of OR's will also influence the future use of CDNet. Certain environmental factors, such as past demands for ICPSR services and the vitality of the member university's research community, will contribute to usage patterns. These factors have not been examined here. Rather, attention has been focused on a few known obstacles to the use of computer communications. As CDNet swings into production, the importance of these and other factors should become evident.

[ Editor's note: The following is reprinted from ICPSR's Guide to resources and services 1985-1986, p24. "Testing of a new remote service, Consortium Data Network (CDNet), is currently underway and should be available in the fall. This new service is aimed initially at ICPSR Official Representatives. CDNet will provide access to an on-line searchable version of the holdings section of the Guide, an on-line data and codebook ordering facility, an interactive message and conferencing facility as well as access to statistical software for analysis of ICPSR holdings. A data base containing information about each item in a large subset of the studies available through the ICPSR is also being produced for inclusion in CDNet. Connection to CDNet will be available through the Autonet, Telenet and Tymnet public data networks." ]
Table 1

Access to a Computing Terminal

| Terminal at Home | Terminal at Work Have Access | Terminal at Work Don't Have | Home Total |
|------------------|------------------------------|-----------------------------|------------|
|                  | n %                          | n %                         | n %        |
| Have Access      | 43 48%                       | 9 10%                       | 52 58%     |
| Don't Have       | 25 28%                       | 13 14%                      | 38 42%     |
| Work Total       | 68 76%                       | 22 24%                      | 90 100%    |

1 Percentages are based on the total number of respondents in the overall table.
2 One questionnaire was excluded from analysis since information was provided for only one of the items.

Table 2

Frequency of Terminal Use by Location & Access to Terminal

| Frequency of Use | Location of Immediate Terminal Access | Totals |
|------------------|---------------------------------------|--------|
|                  | Access at Work and Home | Access at Work Only | Access at Home Only | No Immediate Access | n % |
|------------------|--------------------------|---------------------|---------------------|---------------------|-----|
| [1]              | 36 84%                   | 18 78%              | 6 67%               | 1 8%                | 61 71% |
| [2]              | 3 6%                     | 2 9%                | 3 33%               | 0                   | 8 9  |
| [3]              | 2 5%                     | 3 13%               | 0                   | 4 34%               | 9 10% |
| [4]              | 0                        | 0                   | 0                   | 2 16%               | 2 2  |
| [5]              | 2 5%                     | 0                   | 0                   | 4 34%               | 6 7  |
| [6]              | 0                        | 0                   | 0                   | 1 8%                | 1 1  |
| Total            | 43 100%                   | 23 100%             | 9 100%              | 12 100%             | 87 100% |
| N.A.             | 0                        | 2                   | 0                   | 1                   | 3     |

1=Several Times a Day  2=Once a Day  3=Couple of Times a Week  4=Once a Week  5=Couple of Times a Month  6=Infrequently
| Table 3 | Type of Access to Dial-out Facilities |
|---------|-------------------------------------|
| **Is Dial-out Available Through ...** | **Both Terminal at Work & via Mainframe Computer** |
| **Answer** | **A Terminal at Work** | **A Mainframe Computer** | **n** | **%** | **n** | **%** |
| Yes | 40 | 44% | 64 | 71% | 29 | 32% |
| No | 25 | 28 | 15 | 17 | 29 | 32 |
| N.A. | 25 | 28 | 11 | 12 | 32 | 36 |

| Table 4 | Availability of a Full Screen Editor |
|---------|-------------------------------------|
| **Does Your Terminal Allow Full Screen Editing?** |
| **Answer** | **Terminal at Work** | **Terminal at Home** |
| | **n** | **%** | **n** | **%** |
| Yes | 58 | 88% | 41 | 82% |
| No | 8 | 12 | 9 | 18 |
| N.A. | 24 | 40 |

| Table 5 | Use of Electronic Mail, Computer Conferences, and Networks |
|---------|----------------------------------------------------------|
| **Experience with ...** | **n** | **%** |
| Only Networks | 9 | 10% |
| Only E-mail¹ | 11 | 12 |
| Networks & CC¹ | 1 | 1 |
| Networks & E-mail | 24 | 27 |
| Networks, CC & E-mail | 17 | 19 |
| None | 28 | 31 |
| Total | 90 | 100% |

¹ E-mail=Electronic Mail  CC=Computer Conference
Table 6

Computer Communication Experience by Access to a Terminal

| Type of Access | Access at Work & Home | Access at Work Only | Access at Home Only | No Immediate Access | Total |
|----------------|-----------------------|---------------------|---------------------|---------------------|-------|
|                | n %                   | n %                 | n %                 | n %                 |       |
| [1]            | 6 14%                 | 1 4%                | 0                   | 2 15%               | 9 10% |
| [2]            | 7 16%                 | 4 16%               | 0                   | 0                   | 11 12 |
| [3]            | 0 1%                  | 1 4%                | 0                   | 0                   | 1 1   |
| [4]            | 12 23%                | 6 24%               | 4 45%               | 2 15%               | 24 27 |
| [5]            | 10 20%                | 5 20%               | 2 22%               | 0                   | 17 19 |
| [6]            | 8 19%                 | 8 32%               | 3 33%               | 9 70%               | 26 31 |
| Total          | 43 100%               | 25 100%             | 9 100%              | 13 100%             | 90 100% |

1=Network Only 2=E-mail Only 3=Network & Computer Conference 4=Network & E-mail 5=Network, E-mail & CC 6=None

Table 7

How Often Computer Communications Are Used

| Use Rate       | Electronic Mail | Computer Conferences | Networks |
|----------------|-----------------|----------------------|----------|
|                | n %             | n %                  | n %      |
| Daily          | 16 32%          | 4 22%                | 5 10%    |
| Twice Weekly   | 15 30%          | 2 11%                | 11 23    |
| Once a Week    | 6 12%           | 3 17%                | 8 16     |
| Twice Monthly  | 6 12%           | 2 11%                | 7 14     |
| Infrequently   | 7 14%           | 7 39%                | 18 37    |
| Total          | 50 100%         | 18 100%              | 49 100%  |
Herbert Hyman died on December 19, 1985 of cardiac arrest. He was in China where he had travelled to speak at a conference on "Uses of Sociology in Developing Countries." The 67 year-old sociologist was a specialist in survey research and was credited with having developed the science of polling in the 1930's. A former president of the American Association of Public Opinion Research (AAPOR), he was the author of a classic book on polling, *Survey Design and Analysis*, which is still widely used today. In addition he wrote three other books. One of them, published in 1972, *Secondary Analysis of Sample Surveys: Principles, Procedures and Potentialities*, is one of the seminal works in this field. It is a book which should be on the shelves of every data archive and every data library.

Hyman received all of his post-secondary education at Columbia University and taught there from 1951 until 1969, rising during those years from assistant professor to department chairman. He left Columbia to become University Professor at Connecticut Wesleyan, from which he retired in 1984. A festschrift is being prepared in his honor.

Hyman was a fine scholar and a kind and gentle man. He was adored by his students and loved and respected by his colleagues. His death is a loss to all of us.
IASSIST

MEMBERSHIP AND SUBSCRIPTION FEES

The International Association for Social Science Information Services and Technology (IASSIST) is a professional association of individuals who are engaged in the acquisition, processing, maintenance, and distribution of machine readable text and/or numeric social science data. The membership includes information systems specialists, data base librarians or administrators, archivists, researchers, programmers, and managers. Their range of interests encompasses hardcopy as well as machine readable data.

Paid-up members enjoy voting rights and receive the IASSIST QUARTERLY. They also benefit from reduced fees for attendance at regional and international conferences sponsored by IASSIST.

Membership fees are:

REGULAR MEMBERSHIP: $20 per calendar year

STUDENT MEMBERSHIP: $10 per calendar year

Institutional subscriptions to the QUARTERLY are available, but do not confer voting rights or other membership benefits.

INSTITUTIONAL SUBSCRIPTION: $35 per calendar year
(includes one volume of the QUARTERLY)

Ms. Jacqueline McGee
The Rand Corporation
1700 Main Street
Santa Monica, California
U.S.A. 90406
(213) 393-0411
