Union Calendar No. 571

101st Congress, 2d Session - - - - - - - - - - - House Report 101-978

TAKING A BYTE OUT OF HISTORY: THE ARCHIVAL PRESERVATION OF FEDERAL COMPUTER RECORDS

TWENTY-FIFTH REPORT

BY THE

COMMITTEE ON GOVERNMENT OPERATIONS

NOVEMBER 6, 1990.—Committed to the Committee of the Whole House on the State of the Union and ordered to be printed

---

U.S. GOVERNMENT PRINTING OFFICE

34-937 WASHINGTON : 1990

For sale by the Superintendent of Documents, Congressional Sales Office
U.S. Government Printing Office, Washington, DC 20402

COMMITTEE ON GOVERNMENT OPERATIONS

JOHN CONYERS, JR. Michigan, *Chairman*

CARDISS COLLINS, Illinois
GLENN ENGLISH, Oklahoma
HENRY A. WAXMAN, California
TED WEISS, New York
MIKE SYNAR, Oklahoma
STEPHEN L. NEAL, North Carolina
DOUG BARNARD, JR., Georgia
BARNEY FRANK, Massachusetts
TOM LANTOS, California
ROBERT E. WISE, JR., West Virginia
BARBARA BOXER, California
MAJOR R. OWENS, New York
EDOLPHUS TOWNS, New York
BEN ERDREICH, Alabama
GERALD D. KLECZKA, Wisconsin
ALBERT G. BUSTAMANTE, Texas
MATTHEW G. MARTINEZ, California
NANCY PELOSI, California
DONALD M. PAYNE, New Jersey
JIM BATES, California
GARY A. CONDIT, California
DENNIS E. ECKART, Ohio
PATSY T. MINK, Hawaii

FRANK HORTON, New York
WILLIAM F. CLINGER, JR., Pennsylvania
AL McCANDLESS, California
HOWARD C. NIELSON, Utah
RICHARD K. ARMEY, Texas
DONALD E. "BUZ" LUKENS, Ohio
J. DENNIS HASTERT, Illinois
JON L. KYL, Arizona
CHRISTOPHER SHAYS, Connecticut
PETER SMITH, Vermont
STEVEN SCHIFF, New Mexico
CHUCK DOUGLAS, New Hampshire
C. CHRISTOPHER COX, California
CRAIG THOMAS, Wyoming
ILEANA ROS-LEHTINEN, Florida

JULIAN EPSTEIN, *Staff Director*
DONALD W. UPSON, *Minority Staff Director*

GOVERNMENT INFORMATION, JUSTICE, AND AGRICULTURE SUBCOMMITTEE

ROBERT E. WISE, JR., West Virginia, *Chairman*

GLENN ENGLISH, Oklahoma
EDOLPHUS TOWNS, New York
BARBARA BOXER, California
GARY A. CONDIT, California

AL McCANDLESS, California
STEVEN SCHIFF, New Mexico
ILEANA ROS-LEHTINEN, Florida

EX OFFICIO

JOHN CONYERS, JR., Michigan

FRANK HORTON, New York

LEE GODOWN, *Staff Director*
ROBERT GELLMAN, *Chief Counsel*
AURORA OGG, *Clerk*
MONTY TRIPP, *Minority Professional Staff*

LETTER OF TRANSMITTAL

HOUSE OF REPRESENTATIVES,
Washington, DC, November 6, 1990.

Hon. THOMAS S. FOLEY,
Speaker of the House of Representatives,
Washington, DC.

DEAR MR. SPEAKER: By direction of the Committee on Government Operations, I submit herewith the committee's twenty-fifth report to the 101st Congress. The committee's report is based on a study made by its Government Information, Justice, and Agriculture Subcommittee.

JOHN CONYERS, Jr., *Chairman.*

CONTENTS

Union Calendar No. 571

101ST CONGRESS
2d Session

HOUSE OF REPRESENTATIVES

REPORT
101-978

TAKING A BYTE OUT OF HISTORY: THE ARCHIVAL PRESERVATION OF FEDERAL COMPUTER RECORDS

NOVEMBER 6, 1990.—Committed to the Committee of the Whole House on the State of the Union and ordered to be printed

Mr. CONYERS, from the Committee on Government Operations, submitted the following

TWENTY-FIFTH REPORT

BASED ON A STUDY BY THE GOVERNMENT INFORMATION, JUSTICE, AND AGRICULTURE SUBCOMMITTEE

On October 19, 1990, the Committee on Government Operations approved and adopted a report entitled "Taking a Byte Out of History: The Archival Preservation of Federal Computer Records." The chairman was directed to transmit a copy to the Speaker of the House.

I. INTRODUCTION AND SUMMARY

Long after Federal agency records are needed for current operations, they remain useful in other ways.[1] The records are source material for agencies, historians, political scientists, genealogists, and other students of the United States, its Government, its foreign relations, and its people. Federal records document the history and intent of public policy and form the basis of our national history.[2]

[1] See generally "To Establish A National Policy on Permanent Papers." Hearing on H.J. Res. 226 before the Government Information, Justice, and Agriculture Subcommittee of the House Committee on Government Operations, 101st Cong., 2d Sess. (1990) [hereinafter cited as "Hearing"].

The committee wishes to acknowledge the assistance of Stephen B. Gould, Information Science and Technology Analyst, Science Policy Research Division, Congressional Research Service, in the preparation of this report.

[2] The preservation activities of the National Archives and Records Administration have long been of interest to the committee. See, e.g., *National Archives and Records Service Film-Vault Fire at Suitland, MD.* Hearings before a subcommittee of the House Committee on Government Operations, 96th Cong., 1st Sess. (1979); House Committee on Government Operations, *National Archives Film-Vault Fire, Suitland, Maryland, December 7, 1978,* House Report No. 96-574, 96th

Continued

The ongoing computer revolution makes the task of saving Government records for historical purposes more complex. This is the principal subject of this report. Preserving computer records for archival use requires careful advance planning.

A record remains useful only as long as the medium on which its information is stored can be read and understood. There are many examples of historical records that were preserved but became unusable because the ability to read them was lost. Ancient Egyptian hieroglyphics were undecipherable for centuries until the discovery of the Rosetta Stone in 1799. Stonehenge—a standing stone circle on Salisbury Plain in England—is now believed to have been a lunar eclipse calculator. The ability to read this "stone computer" was lost for thousands of years until the middle of this century.[3]

Records can lose utility for other reasons. Most paper manufactured since the mid-19th century has a high acid content. Over time, the acid attacks the cellulose that makes up paper, breaking it into smaller and smaller pieces until it becomes impossible to turn the pages of a book without destroying it.[4] Only heroic and expensive efforts can save a book that is crumbling.[5]

The life span of "modern" acidic paper may be only a few decades. By contrast, the life span of acid free paper can be measured in centuries. Books that are hundreds of years old remain readable while books printed in this century have deteriorated beyond use. The Federal Government is spending millions of dollars to save important documents printed on acidic paper.[6]

The ability of computers to manage information has developed rapidly in recent decades. Electronic information systems have been steadily deployed by the Federal Government to replace and supplement outdated and overwhelmed paper-based records systems. According to one estimate, 75 percent of all Federal transactions will be handled electronically in the year 2000.[7] Computers play a central role in making Federal agencies more productive and in improving the quality, timeliness, and cost-effectiveness of government operations.

Cong., 1st Sess. (1979); *National Archives and Records Service Documents Preservation Program and Trust Fund Operation,* Hearings before a subcommittee of the House Committee on Government Operations, 96th Cong., 1st Sess. (1979); *Government Recordkeeping Requirements and Actions Act of 1981,* Hearings on H.R. 316 before a subcommittee of the House Committee on Government Operations, 97th Cong., 1st Sess. (1981); *Oversight of the National Archives and Records Service,* Hearings before a subcommittee of the House Committee on Government Operations, 97th Cong., 2d Sess. (1982); *National Archives and Records Service and Its Relationship to the General Services Administration,* Hearing before a subcommittee of the House Committee on Government Operations, 98th Cong., 1st Sess. (1983).

In addition, the committee has originated public laws relating to preservation of Federal Government records. See, e.g., Presidential Records Act of 1978, Public Law 95-591, 92 Stat. 2523 (1978); National Archives and Records Administration Act of 1984, Public Law 98-497, 98 Stat. 2280 (1984); Presidential Libraries Act of 1986, Public Law 99-323, 100 Stat. 495 (1986).

[3] See G. Hawkins, "Stonehenge Decoded" (1965).

[4] See generally Office of Technology Assessment, "Book Preservation Technologies" (1988).

[5] The Library of Congress has microfilmed over 250,000 since 1968. Over 20,000 books are being filmed each year. See "Hearing" at 13 (testimony of Dr. James H. Billington, Librarian of Congress).

[6] See generally "Hearing." Dr. Don Wilson, Archivist of the United States, testified that the historical papers from the World War II period are becoming fragile and difficult to use as the fiftieth anniversary of that war approaches. He estimated that the cost of transferring these records to a more stable medium would exceed $71 million dollars. Id. at 30.

[7] J. Bortnick, "CRS Review" 2 (July-August 1990) (Information Technology Revolution).

Increasingly, Federal records with enduring value are created, maintained, and stored in electronic formats.[8] For example, electronic mail is being increasingly used as a means of written communication on administrative and policy matters. Computer systems track projects and manage caseloads. Imaging systems capture paper documents in digital formats for computerized management. These systems reduce the paper flow in Federal agencies.[9]

While the "paperless office" may still be a long way off, computers are changing the nature of Federal records. The current, widespread use of personal computers directly by Federal managers, lawyers, and other professionals is creating a multitude of important records that may never be printed on paper. A study completed in 1989 found that Federal agencies retain most policy documents on paper and might continue to do so for another decade. There is, however, substantial uncertainty about this projection.[10] Regardless, some computer files are multidimensional and cannot be reproduced on paper.

Federal agencies use many different types of software and data formats for electronic data systems. While there is movement toward adoption of common standards, many current data systems are dependent on specific hardware or software. Sometimes, files can be readily converted to a format that uses generic software and standard hardware. When this is possible, specific software and hardware are not needed to ensure long-term access. However, many electronic records remain dependent on specific software or hardware.

The case of the 1960 U.S. Census is illustrative. The type and format of computer tapes used for census returns in the 1960's became obsolete a few years later. There are only two machines in the world that can read the original data tapes from that census. One machine is in the Smithsonian Institution and the other is in Japan.[11]

The 1960 census records were eventually converted to a more standard format.[12] The incident shows how quickly computer records can become obsolete. If a computer record cannot be read, then for all practical purposes, the record no longer exists. Like Stonehenge, it is possible that a computer tape can be seen but not

---

[8] The electronic information "revolution" is creating policy pressures in many different ways. Problems relating to public access and dissemination of Federal information in electronic formats have been well documented in other hearings and reports by the committee. See, e.g., Committee on Government Operations, "Electronic Collection and Dissemination of Information by Federal Agencies: A Policy Overview." House Report No. 99-560, 99th Cong., 2d Sess. (1986). See also "Electronic Collection and Dissemination of Information by Federal Agencies," Hearings before a subcommittee of the House Committee on Government Operations, 99th Cong., 1st Sess. 467 (1985); "OMB's Proposed Restrictions on Information Gathering and Dissemination by Agencies," Hearing before a subcommittee of the House Committee on Government Operations, 99th Cong., 1st Sess. (1985), "Federal Information Dissemination Policies and Practices," Hearings before the Government Information, Justice, and Agriculture Subcommittee of the House Committee on Government Operations, 101st Cong., 1st Sess. (1989).

[9] See National Academy of Public Administration, "The Effects of Electronic Recordkeeping on the Historical Record of the U.S. Government" 23-27 (1989) [hereinafter cited as "NAPA Study"]. The study was conducted for the National Archives.

[10] Id. at 23.

[11] "Report of the Committee on the Records of Government" 9 (1985).

[12] See letter from Kenneth Thibodeau, Director, Center for Electronic Records, NARA, to Robert Gellman, Chief Counsel, Subcommittee on Government Information, Justice, and Agriculture (Aug. 15, 1990) [hereinafter cited as "Thibodeau Letter"] ."[S]ubstantial efforts on the part of the Bureau of Census in the late 1970's solved this problem.").

understood. Rapid innovation in computer technology guarantees that the problems of preserving the utility of machine-readable media will grow.

Of course, not all records of government activities can or should be preserved.[13] The National Archives and Records Administration accessions only about 2 percent of Government records as permanent records for long-term preservation.[14] Since paper has been the principal recording medium for most of the history of the United States, most records in the National Archives today are on paper. Currently, this amounts to approximately 3.25 billion pieces of paper.[15]

Historical documents can have a value and importance that goes beyond their intrinsic content. The actual Declaration of Independence signed in Philadelphia by Benjamin Franklin, Thomas Jefferson, John Adams, and others is a sacred relic carefully preserved and displayed at the National Archives. The original Constitution of the United States is also preserved. These are just two of the billions of pieces of history maintained at the National Archives.

The original Declaration of Independence will always be an important document even though the contents have been widely duplicated. Computer records have no such cachet. They are only worth preserving if meaningful access can be provided in the future and if the utility of the records can be assured. The Federal Government must take steps to identify, preserve, and provide for the practical use of information of historical interest created and stored on computers.

Preserving the existence and utility of electronic records requires more attention from Federal agencies and from the National Archives. The infiltration of the personal computer into government activities has exacerbated the problem. Preservation policies that were adequate for mainframe computer records need to be reevaluated and revised to meet the challenges of the ever-changing computer age. Managing electronic records to ensure long-term availability is the most significant challenge facing the archival community.[16]

II. Findings

1. As a direct result of the increased use of computers—and especially personal computers—by Federal agencies, Federal records of enduring value are increasingly being created and managed in electronic formats. Word processing programs, computer spreadsheets, data base management programs, document imaging, computer-based modeling, electronic mail, and other computer software

---

[13] The Federal Records Acts of 1950 defines Federal 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 information value of data in them." 44 U.S.C. 3301 (1988).

[14] See "Government Recordkeeping Requirements Actions and Act of 1981," Hearings on H.R. 316 before a subcommittee of the House Committee on Government Operations, 97th Cong., 1st Sess. at 125 (1981) (testimony of George Scaboo, Acting Assistant Archivist for Federal Records).

[15] "Hearing" at 25 (testimony of Dr. Don W. Wilson, Archivist of the United States).

[16] National Historical Publications and Records Commission, "Electronic Records Issues: A Report to the Commission" 2 (1990) [hereinafter cited as "NHPRC Report"].

applications are contributing to the growth of Federal electronic records.

2. An increasing number of Government records worthy of preservation for historical purposes will necessarily exist only in electronic formats. The traditional approach of preserving paper documents in a central archives may not be adaptable to electronic records.

3. Simple physical preservation of electronic records may be inadequate to meet archival needs. Electronic records must be accessible and usable in the future or they may not be worth preserving.

4. Preservation of electronic information may require the continued availability of computer software, operating systems, manuals, and hardware, as well as various types of electronic storage media. Continual changes in computer hardware and software and in patterns of computer usage contribute to the complexity of long-term preservation.

5. The development and implementation of standards for computer data files and programs will ease preservation problems, but the rapid development of technology and the slower implementation of standards mean that standards will not solve all electronic preservation problems.

6. NARA's current policies are inadequate to assure the long-term preservation of electronic records. Both evolutionary and revolutionary changes in NARA's policies, methods, and procedures may be needed to respond to the challenge of preserving electronic records.

7. NARA is not currently prepared to accession some computer records created by Federal agencies that will be candidates for preservation in the next few years.

8. Too little attention is being devoted by Federal agencies to planning for the archival needs of information in electronic recordkeeping systems that are in use and being planned.

III. Recommendations

1. The National Archives and Records Administration needs to undertake a thorough review of its role in the long-term preservation of computer records. NARA's policies should address both the preservation of computer records and the practical utility of the records for future users.

2. NARA needs to recognize and take steps to confront the new preservation problems created by computer records and especially by the creation of hardware and software dependent personal computer records. At a minimum, NARA should review the technical characteristics of the electronic recordkeeping systems used within the Federal Government to assess preservation problems.

3. The Federal Government needs to do more long-range planning and preparation for the archival preservation of records of enduring value maintained in electronic formats. This planning must be reflected in selection of software and hardware for electronic recordkeeping systems and in the development and implementation of standards for computer records. NARA and the Office of Management and Budget should take the lead in requiring other agencies to consider long-term archival needs.

4. Federal agencies that use electronic recordkeeping systems should integrate records management and preservation functions with information resources management staffs and activities, particularly in systems design and training. NARA and OMB should take the lead in making preservation a more important part of information resources management.

5. Recordkeeping rules and procedures should be built into major electronic information systems from the outset, with long-term preservation needs a mandatory design consideration. NARA should work with the General Services Administration, the National Institute of Standards and Technology, and OMB to sensitize the computer industry and Federal agencies to the preservation requirements of the Federal Government.

6. NARA and NIST should continue their active involvement in Federal and private sector standard setting activities, including efforts to establish format standards for electronic records that will minimize or eliminate hardware and software dependent computer records.

7. NARA should make recommendations to the Congress for amendments to the law that are needed to reflect the changes in recordkeeping practices resulting from the widespread use of computers in the Federal Government.

8. NARA should respond to the recommendations in this report by submitting a written report to the Congress not later than July 1, 1991.

IV. Discussion

A. Federal Government Use of Computer Technology

The Federal Government is one of the largest users of computer-based information systems in the United States. Federal agencies were among the first institutions to use computers. During the 1980's, Federal agencies have been rapidly deploying mid-sized and personal computers to augment large-scale data processing.[17]

Most Federal agencies are making the transition to computer-based information management technology. In some agencies, the introduction of electronic technology is proceeding quickly, and some Federal offices already have one computer workstation per staff member.[18]

Management of information is a central focus of Federal computing activities.[19] Information collection, cataloging, processing, communication, and storage functions are steadily being automated by Federal agencies.[20] As a direct result of the increased use of com-

---

[17] Obligations for the information technology activities of Federal agencies increased from $9.1 billion in 1982 to $16.7 billion in 1988. Office of Management and Budget, "Curent Information Technology Resource Requirements" (1989). Personal computers are being acquired so rapidly by Federal agencies that it appears to be impossible to obtain current or accurate statistics.

[18] "NAPA Study" at 24. For example, at the headquarters of the Census Bureau, all 4,000 staff members have their own personal computers. Id. at 25.

[19] "The Federal Government is the largest single producer, consumer, and disseminator of information in the United States." Office of Management and Budget, "Management of Federal Information Resources," 50 Federal Register 52730, 52736 (December 24, 1985) (Circular No. A-130).

[20] "The use of up-to-date information technology offers opportunities to improve the management of government programs, and access to, and dissemination of, government information." Id.

puters by Federal agencies, Federal records of enduring value are increasingly being created and managed in electronic formats.

Thousands of computer tapes in the Federal Government contain statistical, scientific, technical, or descriptive information reflecting major programs of the Federal Government. The National Archives and Records Administration retains electronic records that document significant agency missions and programs. These records are preserved for subsequent use by the original agency; other Federal, State, and local agencies; scholars and researchers; and members of the public.[21] Responsibility for appraisal of electronic records is vested in NARA's Center for Electronic Records in the Office of the National Archives.[22]

The following categories and examples illustrate the variety of electronic records that have already been accessioned into the National Archives or are scheduled for accession as permanent records:

Political or judicial data.—Examples are the Immigration and Naturalization Service's Naturalizations File (1971–); and the Bureau of Justice Statistics' Census of Juvenile Detention and Correctional Facilities.

National security and international relations data.—Examples are the State Department's Automated Document System of Central Foreign Policy Files, consisting of both an index for locating documents and the texts of telegrams, memoranda, and correspondence; and the United States Information Agency's General Population Surveys of International Political Issues (1972–).

Data that document military or civilian operations during times of war, civil emergency, or natural disaster.—Examples include the President's Commission on the Accident at Three Mile Island's 1979 Study of Behavioral Effects; and the Joint Chiefs of Staff's Combat Activities File, containing data on flight operations flown in Southeast Asia from 1965 to 1970.

Socioeconomic data.—Examples include the Department of Education's National Evaluation of the Emergency School Aid Act, 1973–1976, the Office of Economic Opportunity's 1970 study of Poverty Neighborhoods in 105 Large Central Cities; the Bureau of the Census' Annual Import and Export Data Banks (1964–), Census of Agriculture (1949–), and Decennial Censuses of Population and Housing (1940–); and the U.S. Environmental Protection Agency's Population at Risk System, containing data related to pollution and health.

Administrative data that have government-wide coverage or significance.—Examples include the Office of Personnel Management's 1980 studies of Whistleblowing and the Federal Employee, and 1981 Survey of the Senior Executive Service; the Equal Employment Opportunity Commission's Equal Employment Opportunity Surveys; and the Federal Awards Assistance

---

[21] National Archives and Records Administration, "Managing Electronic Records" 26 (1990) (Instructional Guide Series) [hereinafter cited as "NARA Guide"].

[22] Letter from Don W. Wilson, Archivist of the United States, to Chairman Bob Wise, Subcommittee on Government Information, Justice, and Agriculture (April 13, 1990) reprinted in "Hearing" at 50 [hereinafter cited as "Wilson Letter"].

Data System, which contains quarterly data about Federal assistance to State, county, and local governments.

Natural resources data.—Examples include the Minerals Management Service's Mineral Availability System, containing types and locations of mineral deposits, and its Royalty Accounting Schedule; and the President's Commission on the Coal Industry's 1979 Survey of Community Conditions in Coal Processing Areas.

Unique and important scientific and technical data.—Examples include data from the National Aeronautic and Space Administration's Tektite I and II undersea living experiments in 1969 and 1970, data from the National Institutes of Health's National Collaborative Perinatal Project, and the Environmental Protection Agency's Pesticide Product Information Data.

Geographic data used to map the surface of the earth, other planetary bodies, or the atmosphere.—An example is the National Oceanic and Atmospheric Administration's Hydrographic Nautical Chart Data Base (1970–).

Automated indices to other permanent records.—Examples include the automated index to the Department of Interior's microfilmed records relating to the Trust Territories for the Pacific Islands; the automated index to the correspondence files of the Secretary of the Air Force; and an automated index to the microfilmed records of the Presidential Commission on the Space Shuttle Challenger Accident.[23]

Many governmental activities produce information that exists exclusively on electronic media. For example, a recent report by the Office of Technology Assessment concluded that electronic media offer the only way to manage the massive volume of Federal scientific and technical information.[24] A single specific illustration underscores the tremendous volume of scientific data: the National Aeronautics and Space Administration has more than 1.2 million magnetic tapes containing space science data.[25]

While statistical, scientific, technical, or numeric information may exist primarily in electronic form, records that reflect final decisionmaking and policy are still mostly retained on paper today. A 1989 study of electronic recordkeeping by the National Academy of Public Administration found that paper is still the dominant medium in the world of Government policy formulation, program development, and decisionmaking.[26]

How much of this paper-based information is duplicated in electronic media is unknown. Some—and perhaps most—of the paper currently being produced has electronic counterparts. Many paper documents created in the last few years for use in agency activities have been created using word processing software on personal computers. Paper may provide a satisfactory long-term storage medium for documents created with word processing programs. Until there

---

[23] "NARA Guide" at 23-25.

[24] Office of Technology Assessment, "Helping America Compete: The Role of Federal Scientific & Technical Information" 2 (1990).

[25] General Accounting Office, "Space Operations: NASA Is Not Property Safeguarding Valuable Data From Past Missions" at 12 (1990) (GAO/IMTEC-90-1) [hereinafter cited as "GAO NASA Report"].

[26] "NAPA Study" at 23.

are more effective ways of preserving, indexing, and accessing electronic copies of documents that can be printed on paper, continued reliance on paper for archival uses may be appropriate.[27]

Reliance on a printed copy of a document created on a computer may result in the loss of some of the document's history. With personal computers, each user becomes an information manager.[28] The author of a document written using a word processor continuously decides which versions of a document will be saved and which will be deleted. The drafts of a document created on a personal computer are more likely to be lost than the drafts of a document that only existed in paper form.[29]

The widespread and growing use of computers—and especially personal computers—has created additional categories of computerized information for which no convenient paper equivalent is possible. Computer software creates pools of information that are either not normally printed on paper or that cannot be printed on paper. Preserving this type of information for long-term use presents a new type of challenge to archivists. New technology—and perhaps even completely new approaches to preservation—will be necessary.

Exclusively electronic records are created by computer-based systems for tracking projects, managing caseloads, and preparing correspondence. Similar records are created by spreadsheet programs like LOTUS 1-2-3, which perform calculations involving numeric data bases and mathematical formulas or models, and by data base programs, which permit the manipulation of interrelationships between large sets of numeric and nonnumeric data. These and similar programs are used widely throughout the Federal Government.

An example of a new category of computerized information is electronic mail. Electronic mail is being increasingly used as a means of written interoffice communication on both administrative and policy matters. The Iran-Contra Affair illustrates the importance of an electronic mail system as a repository of information crucial not only for historians but for current policy, oversight, and investigatory purposes.[30] The incident also illustrates the lack of

---

[27] "In spite of much discussion of the paperless office of the future, most automated systems greatly enhance the capability and the tendency of agencies to produce more information in hard copy formats." M. Hedstrom & A. Kowlowitz, "Meeting the Challenge of Machine-Readable Records: A State Archives Perspective" "Reference Services Review" 31, 33 (1988) (Number 1-2) [hereinafter cited as "Hedstrom & Kowlowitz"].

[28] Id. at 38.

[29] A report prepared for the Florida legislature describes this problem in more detail:

> Most computer applications are, however, designed to meet the needs . . . of a specific set of users, and do not consider requirements of other elements of the organization. As a result, files which are no longer being used by a system may be periodically purged without regard for future use. The paper trail of a first draft, amendments, annotations, and a final draft may simply disappear. The increasing use by state and local governments of microcomputers adds a new dimension to this problem, because, unlike large data centers, there is generally no systemic "back-up" or copying of computer files by users. Not only may records be lost, the rest of the organization may not know of their existence and cannot assess their importance. Electronic mail will also complicate the responsibilities of records managers. There may be no paper records of communications and, therefore, no documentation will exist.

Joint Committee on Information Technology Resources, "Flordia's Information Policy: Problems and Issues in the Information Age" 140 (1989).

[30] For a general review of the Iran-Contra Affair, see "Report of the Congressional Committees Investigating the Iran-Contra Affairs," House Report No. 100-433, 100th Cong., 1st Sess. (1987) [hereinafter cited as "Iran-Contra Report"].

attention that has been paid to the preservation of some computerized records.

During the Reagan administration, the National Security Council and others in the White House used a computer system (known as the "PROF system") [31] to create, store, and retrieve information. The system combined routine office functions, including the ability to send electronic mail and documents to others on the system.[32] Both the Tower Commission and the congressional committees that investigated the Iran-Contra Affair relied heavily upon the information from the PROF system. The Tower Commission described the system's importance:

> The PROF messages were conversations by computer, written at the time events occurred and presumed by the writers to be protected from disclosure. In this sense, they provide a first-hand, contemporaneous account of events.[33]

During the last few days of the Reagan administration, a lawsuit was filed to prevent the destruction of computer tapes from the PROF system.[34] The lead plaintiff was Scott Armstrong, Executive Director of the National Security Archive. In the original complaint, Armstrong alleged that records from the PROF system were about to be destroyed in violation of the Presidential Records Act and other laws.[35]

The principal concern of the plaintiff was that the system contained substantive information that does not exist anywhere on paper. The plaintiff alleged that the only copy of some data was on the computer tapes that the White House was about to erase. Whether the electronic messages in the PROF system were printed on paper and preserved is a factual issue that remains unresolved.

The experience with the White House PROF system illustrates the transitory nature of electronic records. Some paper documents vital to the administrative and historical records were intentionally removed from the files and shredded.[36] Electronic documents were also intentionally deleted from the computer system.[37]

The destruction of historically valuable records—willful or not—is a problem of unknown dimension. It exists for both paper and electronic records. The recovery of some electronic records from the PROF system was possible because system backup tapes still existed. The preparation of backup tapes is a standard management procedure for most computer systems. The White House backup tapes were normally kept for only 2 weeks before being recycled.[38]

---

[31] The acronym stands for "Professional Office System," an inter-company communications system marketed by IBM. The system is also referred to sometimes as the "PROFS" system. See *Armstrong* v. *Bush,* 721 F. Supp. 343 (D.D.C. 1989).

[32] Operated on a mainframe computer, the PROF system tapes may be easier to identify and preserve on magnetic tape. An electronic mail system operated by personal computers networked together may present more difficult preservation problems.

[33] "Report of the President's Special Review Board" at III-1 (1987) ("Tower Commission").

[34] *Armstrong* v. *Reagan,* Civ. A. No. 89-142 (D.D.C.). See *Armstrong* v. *Bush,* 721 F. Supp. 343 (D.D.C. 1989).

[35] The claims under the Presidential Record Act, Federal Records Act, and Freedom of Information Act were dismissed. The claim under the Administrative Procedure Act was not dismissed. *Armstrong* v. *Bush,* 721 F. Supp. 343 (D.D.C. 1989).

[36] See, e.g., "Iran-Contra Report" at 307.

[37] See "Computer Detective Found Crucial Data," Washington Post, February 28, 1987, at A10; *Armstrong* v. *Bush,* 721 F. Supp. 343 (D.D.C. 1989) (describing the backup process for PROF tapes). See also "Iran-Contra Report" at 307 note 2.

[38] See *Armstrong* v. *Bush,* 721 F. Supp. 343, 345 (D.D.C. 1989).

The preservation of some backup tapes with vital Iran-Contra information was apparently accidental. There was no plan to preserve the backup tapes for long-term archival purposes.

The PROF notes lawsuit is still pending. Its relevance here is to highlight the critical importance of records that may reside in computer systems supporting routine office functions. It also exemplifies the transient nature of some computer records, especially those in an electronic mail system. Not all such records will be as crucial to American history as the PROF system at the White House during the Reagan Presidency. Many agency electronic mail records, like many paper records, will not be worthy of preservation at all. But those records worthy of preservation may only exist in electronic form, and methods for identifying and preserving them are needed.

The White House PROF system, like many electronic systems, contain records that change constantly. The archival preservation of this information presents a special challenge. There is no "final" product that can be transferred to the National Archives when records are no longer needed. Further, no single snapshot of the contents of the system will be complete, representative, or reflective of the way in which the information was used or changed. For such records, the traditional approach to preservation will not save all important features of the records. Other ways of meeting archival requirements may be needed.

Findings: As a direct result of the increased use of computers—and especially personal computers—by Federal agencies, Federal records of enduring value are increasingly being created and managed in electronic formats. Word processing programs, computer spreadsheets, data base management programs, document imaging, computer-based modeling, electronic mail, and other computer software applications are contributing to the growth of Federal electronic records.

An increasing number of Government records worthy of preservation for historical purposes will necessarily exist only in electronic formats. The traditional approach of preserving paper documents in a central archives may not be adaptable to electronic records.

Simple physical preservation of electronic records may be inadequate to meet archival needs. Electronic records must be accessible and usable in the future or they may not be worth preserving.

B. SOFTWARE AND HARDWARE DEPENDENCE

1. Data Files

At the most basic level, electronic records consist of bits (binary digits) and bytes (a set of eight binary digits) that represent all data as some combination of ones and zeros. The bits and bytes are the lingua franca of the digital computer. Standard code sets translate the bits into "alphabets" that people recognize. Two alphabets are common: the American Standard Code for Information Interchange (ASCII) and the Extended Binary Coded Decimal Interchange Code (EBCDIC).[39] Computer software is a tool that provides

[39] "Wilson Letter," "Hearing" at 59.

people with the ability to read, understand, and manipulate the bits and bytes used by computers.

When use of automated data bases by Federal agencies began, the data bases were generally equivalent to simple, sequentially organized lists of information. Many early data bases contained only economic or scientific statistics. These compilations of information consisted of data types that appeared in every record and that could be represented in fixed length fields. Most of these data bases were supported through programs written in high level languages such as COBOL or Fortran. Transferring these files to the National Archives is a matter of selecting the desired information, making sure it is in a generally usable form such as ASCII, and locating all pertinent documentation.[40]

As data base design and implementation advanced, more data types began to be included, many of which could no longer be represented in fixed length fields. Instead of files with the same number of fields in each record and the same number of bytes in each field, data bases began to contain variable length fields and records. Fields were also of variable occurrence; that is, a specific field might appear once in most records, three or four times in some records, and not at all in other records.[41] Computers also developed the ability to manipulate graphics, voice, and other new types of data.

The structure of data bases has also changed. Sequential flat files have been replaced by hierarchical, networked, and relational data base structures. In flat files, information in a record is related to other information in that record only. In more sophisticated data bases, records require information from other records to be understood. Understanding the data is now often no longer enough. Understanding the relationships between data is equally important. Sometimes, the relationships are not necessarily part of the data base proper. They might be spelled out in software or in a data dictionary.

As data bases developed, specialized data base software made it easier to enter, manipulate, and retrieve data. These data base programs do not necessarily keep data in simple formats as did standard high level programming languages.[42] Each data base program may have a proprietary means of holding data and information about relationships between data.[43]

The movement in the Federal Government and elsewhere is clearly toward using more sophisticated software to create large and complex data structures. This software expands the functional capabilities, flexibility, and ease of use of information management

---

[40] National Institute of Standards and Technology, "Framework and Policy Recommendations for the Exchange and Preservation of Electronic Records" at 2 (1989) (National Computer Systems Laboratory) [hereinafter cited as "NIST Report"].

[41] Id. at 2-3.

[42] Id. at 3.

[43] Use of the term "proprietary" in this report refers to the practice by computer software and hardware makers of using their own unique technology to create a functional device. The detailed characteristics of such technology may be a trade secret, copyrighted, or patented, and may not necessarily be shared with users or competitors. Some data formats developed by one software publisher have become de facto industry standards. See, e.g., "Wilson Letter," "Hearing" at 50 (The computer budget game accessioned by NARA from the National Economic Commission runs under LOTUS 1-2-3. NARA contends that this does not present a problem "in the short term" because LOTUS is so widespread).

systems.[44] Transfer of data in electronic form, therefore, may no longer be a simple task of cleaning up and copying sequential flat files and transferring them with their documentation to the National Archives.

Data bases may need to be converted from proprietary data formats and preserved with substantial information about relationships between data in the data base. When such conversion is possible, it may require considerable processing of data files, with attendant decisions about what format should be used for long-term storage of the information. It is far from clear whether NARA or the originating agency will do this work, how conversions will be coordinated, and when or even whether it will be done.[45]

Much as data bases have grown from flat files with simple character representation, text documents have grown from ASCII files created by text editors to document representations created on word processors and desk-top publishing systems. Simple text editors create what are essentially strings of ASCII characters using only the few formatting symbols (carriage return, line feed, form feed, tab) defined in ASCII. As word processing software and hardware entered the market, formatting functions such as enhanced text (boldface, italics, oversized characters, etc.) began to be used to improve the appearance of printed documents. These print enhancements are not defined in ASCII and therefore their implementation and representation vary from software package to software package.

The most recent generation of word processing software creates even more complex documents. These word processors create compound documents that contain not only text, but graphics such as charts and other image files. Expanded print enhancements such as automatic footnote numbering and placement have shifted the text portion of the document even further from being a simple sequential representation of characters. With many word processors, the relationship of a footnote to the text is maintained uniquely. The layout is only defined at display or printing using proprietary software to convert the logical relationship of the footnote to the text to a layout relationship of the footnote to a place on a page. Text documents, in a manner analogous to data bases, are coming to rely on relationship information to represent a document. Compound documents, documents with charts and graphics, and documents constructed from relationships cannot be represented in the more universally recognizable ASCII code.[46]

The ultimate technique for decomposition of documents into relationships and data is the so-called "virtual document" in which the document is stored electronically as a set of relationships. At output, the document is assembled from multiple sources following the system's proprietary instructions. At the point when a document becomes a virtual document, the distinction between a data base and a document is very narrow if it exists at all.

---

[44] "NIST Report" at 3.

[45] NARA recently commissioned a study of data bases maintained by Federal agencies. The study will be conducted by the National Academy of Public Administration and will include an assessment of the extent of technological dependence among Federal data bases of enduring value.

[46] "NIST Report" at 4-5.

This does not necessarily mean that documents themselves will cease to exist. Virtual document systems can define a way of explicitly capturing documents. One such method is printing the document on paper. However, it does mean that if NARA chooses to accession an entire document-generating data base system—rather than selected documents—preservation may pose unusually complex problems. As a result, transferring the resulting electronic documents may be just as complex as transferring relational data bases.[47]

2. Hardware and Software

The tendency of computers to create records dependent on specific hardware or software presents additional difficulties for archivists. For example, in New York State, nearly three-quarters of mainframe-based information systems create data in software dependent formats.[48] Dependence is not only a problem with computer records, but with microfiche, videotape, and other media as well.

Some problems associated with hardware dependence diminish or disappear when a standardized process is involved. Reading a microfiche card may not be possible with the naked eye, but there should always be a magnification method even if it is somewhat cumbersome. But if a file is recorded on a magnetic disk and there are no disk drives that can read disks of that size, then the information on the disk is lost.

Rapid changes in technology and unique applications have already led to significant problems in hardware dependence for computer-based files. A new generation of software specifically designed to operate on the latest generation of hardware cannot be operated on earlier models. These trends are certain to continue.[49]

The type of hardware available dictates the media used for storage of records. While magnetic computer tape has been the most popular mass storage medium used by Federal agencies, other media have been used in the past or are emerging as candidates for widespread future use. The earliest storage media, punched cards and paper tapes, were augmented and then supplanted by magnetic tapes and disks. Fixed hard disks and flexible magnetic disks of various sizes are currently being used for short-term storage. The data storage densities of all magnetic media have been increasing steadily. Further increases can be expected.[50]

Many Federal agencies are currently experimenting with optical disks, primarily the 12-inch WORM (Write Once-Read Manytimes) and 4.75-inch CD-ROM (Compact Disk-Read Only Memory).[51] Opti-

---

[47] Id. at 4.

[48] "Hedstrom & Kowlowitz" at 34. About half of these require software developed by the agency for a specific application. Id. There are no comparable statistics for Federal records created on either mainframe or personal computers. See "Hearing" at 45.

[49] "Report of the Committee on the Records of Government" at 86-87.

[50] B. Bennington, "Beyond FTS2000: A Program for Change," at 165-166 (1989) (Board on Telecommunications and Computer Applications, National Research Council) [hereinafter cited as "Bennington"].

[51] See generally "Federal Information Dissemination Policies and Practices," Hearings before the Government Information, Justice, and Agriculture Subcommittee of the House Committee on Government Operations, 101st Cong, 1st. Sess. (1989) (testimony of Jerry McFaul, Computer Scientist, Office of Scientific Information, U.S. Geologic Survey, and Chairman, SIGCAT, Special Interest Group on CD-ROM Applications and Technology).

cal disks that can be repeatedly written on and erased may be available soon. CD-OM can store the equivalent of 300,000 text pages, while a two-sided WORM disk can store the equivalent of 1,200,000 text pages. The theoretical limits of optical disk storage densities are many times greater and are expected to increase steadily.[52] The useful life of optical disks has not been fully evaluated.

The problems of hardware and software dependence are well illustrated by personal computers. There are several "flavors" of personal computer in common use today (e.g., IBM and Macintosh), and software and data from one type of computer may not be usable on another type of computer. Even within a single family of computers, different models may not be fully compatible.[53] In this respect, computers are like automobiles. All automobiles have engines, transmissions, and tires, but the parts from one will not necessarily fit on another.

Another layer of complexity results from the operating system that is essential to the functioning of a personal computer. The operating system (such as DOS, UNIX, or OS/2) is the set of computer instructions that controls the basic functions of the computer hardware and creates the environment in which programs can be executed. It establishes the interface between the hardware and an application program. Programs that will run successfully under one operating system will typically not run at all under other operating systems. Programs that run under one version of an operating system may not run under a different version of the same operating system. In addition, the development of networks of personal computers linked together creates another computing environment. Computer networks have developed rapidly in the last few years.

Yet another set of problems derives from constant change and progress in end user software like word processing programs and spreadsheets.[54] For example, text files created by one word processing program are not likely to be fully compatible with text files created by another word processing program.[55] Further, files from one version of a program may not be usable by other versions of the same program.[56]

Thus, it is possible to have a fully operational personal computer, operating system, software, and data files without being able to use

---

[52] "Bennington."

[53] A personal computer is not a single item of hardware. It consists of many components, including processors; memory chips; circuit boards; monitors; keyboards; and data storage devices such as hard disks, floppy disks of different sizes, and tapes. Not all components are necessarily compatible with each other or with any given software program or operating system.

[54] One commentator suggests that much of the early history of the personal computer has already been lost after only 13 years. Computer software and hardware of relatively recent vintage are being touted as a possible "collectors items." See J. Dvorak, "Collecting History", 7 "PC Magazine" 71 (Jan. 12, 1988); J. Dvorak, "Collecting History, Part 2", 7 "PC Magazine" 71 (Oct. 11, 1988).

[55] A recent review of word processing software commercially identified at least 55 commerically available word processors. 8 "PC Magazine" 98 (Nov. 28, 1989). NARA does not know and does not believe that it is important to know how many different word processing programs and systems are in use in the Federal Government. See "Wilson Letter," "Hearing" at 45.

[56] Both the importance and growth of personal computer software are underscored by the establishment in 1988 of a Library of Congress Machine-Readable Collections Reading Room in 1988 concentrating on microcomputer materials. The Library is collecting software, operating systems, manuals, and a selection of basic hardware. See letter from James H. Billington, Librarian of Congress, to Chairman Bob Wise, Subcommittee on Government Information, Justice, and Agriculture (April 30, 1990), reprinted in "Hearing" at 101-106.

any two of them together. Given the rapid development of personal computer software and hardware, the problem of maintaining compatibility is not trivial. These problems are the driving force behind the "open systems" movement. Open systems is a conceptual framework for standards that permit communication among different computer systems.

3. Documentation

Documentation about the organization, coding, and contents of electronic records is needed to read files and retrieve information.[57] Documentation is also necessary to operate computer hardware and software. The absence of documentation prevent the use of fully functional hardware, software, or data files.

The need for documentation presents problems for archivists. Over time, documentation can become separated from the data, and this can make it difficult or impossible to gather sufficient information to evaluate the records.[58] NARA reports that it frequently receives incomplete or incorrect documentation for computer files:

> In 1989, NARA located several hundred reels of tape from the Department of Health and Human Services stored in the Washington National Records Center. Although we explored several different paths, we were not able to locate any technical documentation for any of the files.
>
> Temporary commissions created to address problems of major national importance frequently leave behind electronic records which are documented poorly or not at all. We have had to reject data files from the National Commission on Marijuana and Drug Abuse, the Public Land Law Review Commission, the President's Commission on School Finance, the National Commission on Consumer Finance and others because of inadequate documentation.
>
> The Combat Area Casualty file . . . identifies fatalities, POWs and MIAs from the war in Vietnam. When DOD first transferred this file to the National Archives, the record layouts and code tables contained errors and omissions. It required several attempts to get complete and correct documentation for this file.[59]

4. Standards

A countervailing force to the problems associated with software and hardware dependence and incompatibility is the growing demand both in Government and the private sector for what is generically known as "open systems." Open systems allow interconnection and interoperation of heterogeneous computers so that users are not constrained to use software or hardware all produced by one vendor. Open systems overcome the problem of compatibility between different hardware and software. The primary means of achieving open systems is through standards. This allows a

---

[57] "Thibodeau Letter."
[58] "Hedstrom & Kowlowitz" at 31.
[59] "Thibodeau Letter."

vendor to build an interface between a product and the standards, rather than multiple interfaces for each possible system that might be connected to it.[60]

For example, the Standard Generalized Markup Language (SGML) is a text markup standard intended to establish a consistent set of codes for labeling the key elements of a document. These elements include such things as chapter titles, paragraph indentations, tabular presentations, etc., that cannot be described using ASCII. Parts of a document are marked by tags, and the logical relationship of the parts is recorded. Specific tag sets are developed for various applications. By using tags to identify parts of documents, text created on one computer system can be transferred and processed on another without the use of the original software or hardware. The tags do not include presentation instructions so that printing or display format decisions are made by the final user.[61]

Text markup standards allow documents to be described in a way that is independent of any computer system, any input or output device, any application, and any system-specific character encoding. This permits a document to be moved from one system to another without any loss of data or formatting instructions.[62]

The movement toward open systems is international. The Open System Interconnection reference model has been developed and sponsored by the International Standards Organization and the International Telegraph and Telephone Consultative Committee of the International Telecommunications Union. The OSI reference model is supported by the European Computer Manufacturers Association, the Institute of Electrical and Electronics Engineers, the American National Standards Institute, and the National Institute of Standards and Technology.[63]

The Federal Government is involved at several levels in the development and implementation of computer standards. The Brooks Act requires the Secretary of Commerce to promulgate standards and guidelines for Federal computer systems. These standards are required to be based on standards and guidelines developed by the National Institute of Standards and Technology.[64] The Paperwork Reduction Act requires the Director of OMB to develop a program to enforce Federal information processing standards.[65] OMB Circular A-130 contains general instructions to agencies to ensure computer system compatibility.[66]

The development of standards for computer applications has the potential to significantly reduce the problems associated with the archiving of electronic records.[67] Because technology develops rap-

---

[60] "NIST Report" at 5-6.

[61] Id. at 15.

[62] R. Snyder, "SGML: What It is and Its Benefits", 1 "Federal Data Report" 9, 10 (April 1990).

[63] D. Hack, "Telecommunications and Information Systems Standardization—Is America Ready?" 15-17 (1987) (Congressional Research Service) (87-458 SPR). See generally Department of Commerce, "Government Open Systems Interconnection Profile" (1988) (Federal Information Processing Standards Publications 146).

[64] 40 U.S.C. § 759(d)(1) (1988).

[65] 44 U.S.C. § 3505(3)(A) (1988).

[66] Office of Management and Budget, "Management of Federal Information Resources," 50 Federal Register 52730, 52737 (December 24, 1985) (Circular No. A-130).

[67] The creation of standards is a complex and often controversial endeavor, and it can be difficult to achieve consensus around any given standard. This point is well-illustrated by the acronym "SGML." One search uncovered 12 different interpretations for the acronym. Id. at 13.

Continued

idly and standards develop slowly, archival needs will not be met by simply waiting for the development and implementation of standards.

Also, it will take a considerable period for adopted standards to be implemented on the hundreds of thousands of Federal computers in use. This means that, at best, there will be a large volume of electronic records created by the Federal Government during many past and future years of nonstandard computer operations. Since some of these records will certainly be worthy of historical preservation, NARA will need to develop an approach to their preservation and use. Because new technology is rapidly being adopted by agencies, the problem of nonstandard records will not disappear over time as standards are developed. A generic approach to the long-term preservation of nonstandard computer records will be needed.

Findings: Preservation of electronic information may require the continued availability of computer software, operating systems, manuals, and hardware, as well as various types of electronic storage media. Continual changes in computer hardware and software and in patterns of computer usage contribute to the complexity of long-term preservation.

The development and implementation of standards for computer data files and programs will ease preservation problems, but the rapid development of technology and the slower implementation of standards mean that standards will not solve all electronic preservation problems.

C. NARA PRESERVATION STRATEGY

The National Archives and Records Administration has 20 years of experience with the accessioning of electronic records. NARA's approach has been to preserve electronic records in a format that can be read by any current computer rather than to preserve the specific hardware and software used in creating the records.[68] It is not clear, however, that this approach fully reflects the complexities of current and future government computer use.

The Archivist of the United States reports that 8,000 data files have been accessioned and that none of the files is hardware dependent.[69] Only 2 percent of the files required special processing to deal with unusual or obsolete formats. The files were processed using current high level software, and there was no need to rely on the original operating system or data base management software used to create the records.[70]

The Archivist concedes, however, that the files accessioned to date reflect an older, simpler norm of data processing.[71] For most

---

There are, of course, other document description standards besides SGML. In its response to questions, NARA referred to Office Document Architecture, Office Document Interchange Format, and Standard Page Description Language. "Wilson Letter, Hearing" at 45. Which standard or standards will actually come into common use in the Federal Government is unclear at this time.

[68] "Wilson Letter," "Hearing" at 47.

[69] But see text accompanying note 91.

[70] "Wilson Letter," "Hearing" at 45. For a discussion of NARA physical preservation policies, see id. at 47.

[71] Id. at 45.

of its short history, the modern electronic computer has been a mainframe operated by a high priesthood of computer specialists. Until the early 1980's, most data processing applications involved batch processing of data files on mainframe computers. Data flowed into and out of the computer through a few choke points, and magnetic tape was a universal storage medium. Since only older records tend to reach the Archives, the electronic records in the Archives are magnetic tapes, mostly from mainframe computers.

The personal computer revolution of the 1980's changed both the role and the accessibility of the computer. By 1990, the personal computer had developed so rapidly that current desktop models can exceed the capacity of state-of-the-art mainframe computers from the 1970's.[72] Low prices meant that personal computers became ordinary tools for office workers. Programs developed for the personal computer offer a powerful set of tools that permit workers who are not computer specialists to manipulate large data bases.

The change in computers and in the way that computers are used means that there will be a change in the type of electronic records that will begin to reach the Archives in the next few years. The Archivist recognizes that the procedures used to provide access to older computer files are not adequate for the more complex data architectures that exist today.[73] Despite this recognition, the National Archives continues to believe that long-term access to electronic records of Federal agencies can be routinely provided in a way that is not dependent on the specific computer hardware and software used to create the records.[74] Several reasons are cited for this assessment.

NARA is operating under the general premise that the archival uses of computer records will be relatively simple. Under this approach, NARA needs only to preserve the records and provide access to them. There is no need to provide the capability to create and update records, to assure data quality, or to provide specific outputs to meet the needs of current agency business.[75] This means that NARA does not necessarily need to maintain fully functional word processing, data base, or other software to support data files that have been accessioned.

NARA's view of the needs of future users of archived electronic records is too shortsighted. For some types of documents, providing access may be sufficient. However, simple access will be inadequate for many of the complex data bases and spreadsheets that are being created today on computers. Future users will need to be able to access and manipulate data in order to analyze use, test alternate hypotheses, and review assumptions of the author.

---

[72] A 1980 computer textbook describes large-scale computers as having a word length of 32 bits, a maximum memory size of 8.4 million bytes, and requiring special space and air conditioning. See E. Award, "Business Data Processing" at 88–89 (1980). In 1990, personal desktop computers are widely available with 32 bit word lengths and a memory size well in excess of 16 million bytes. Laptop computers are also available with some capabilities that exceed those of the 1980 mainframe computer. See 9 "PC Magazine" 118–119 (Mar. 13, 1990).

[73] "Wilson Letter," "Hearing" at 48.

[74] Id. at 44–45.

[75] Id.

In a recent article, two New York State archivists described why preservation of data and the ability to manipulate the data is important:

> In appraising electronic records, archivists must consider which formats are most likely to meet future research needs. One important criterion is whether published summary reports will meet research needs, or whether researcher will want to manipulate the data in ways that were not envisioned by the original creators of the data. If new types of analyses are envisioned, archivists are likely to recommend preservation of the disaggregated data in an electronic format in lieu of or in addition to hard copy output reports.[76]

There are other problems associated with maintaining the long-term utility of records. How will the researcher of the future be able to ascertain what records are stored on the magnetic tapes or disks at a future archives? Indexing and retrieving information from electronic media are not simple problems. Computer tools to accomplish these functions are likely to provide assistance, and computerized search techniques may actually make some aspect of the research process easier. NARA is skilled in accomplishing these tasks for paper documents and should develop similar skills for electronic records.

Also, NARA's assumption that computer records can be preserved and accessed on a long-term basis without some hardware or software support is shaky at best. There are some computer records that cannot be accessed or used effectively without the software that created them.[77]

NARA has identified two general instances when it is necessary to preserve application software. The first is when the software itself has evidential value for documenting policies, procedures, and decisions. The second is when the software is essential to the interpretation of the contents of a record.[78]

For the second category, NARA recognizes that it may be necessary to run the software to access the underlying records. The interrelationship between software and data is illustrated by geographical information systems:

> [G]eographical information systems (GIS) are typically used to produce maps. A GIS contains data and software which translate the data into cartographic formats. It is necessary to run the software to produce the cartographic displays. Such displays realize the information content which is virtual in the system, but which cannot be tapped without appropriate software.[79]

The common computerized spreadsheet program offers a second example of how program and data can be inextricably intertwined:

---

[76] "Hedstrom & Kowlowitz" at 33-4.

[77] See the discussion of hardware and software dependence in Part IV, Section B of this report.

[78] "Wilson Letter," "Hearing" at 48-49.

[79] Id. at 49.

> A second example in which software is necessary to interpretation of the records occurs when some of the data content of the record is actually an extension of the software, e.g., it is not pure data but provides instructions to the software regarding how to process the data. A specific example would be a financial model in the form of a LOTUS 1-2-3 worksheet . . . The worksheet contains not only pure record data, but certain spreadsheet cells contain formulas which instruct the spreadsheet software how to calculate or otherwise derive the contents of those cells from other data in the spreadsheet. Other cells may contain "macro instructions" which are a programming language in their own right. These instructions direct the software in such things as migration through the spreadsheet, solicitation of user input, and alternate (conditional) processing of certain portions of the spreadsheet data.[80]

Maintaining the ability to run software over long periods is a very demanding task. One solution is to preserve the data files, the software, as well as the necessary hardware. For a variety of reasons, NARA is reluctant to adopt this approach.

From an archival perspective, which must address timeframes of 100 years or more, preservation of specific hardware and software may well be technologically and economically unfeasible. To try to preserve the original technology from many different agency systems which generate archivally valuable electronic records would be expensive.[81]

The long-term preservation of operating computer hardware is not a simple task. Engineers, programmers, and operators may be required to maintain and use old computers. Anyone who has struggled to maintain the usability of current computer hardware and software can appreciate the difficulty of maintaining obsolete technology for indefinite periods. Preserving operating systems presents similar problems.[82]

Another aspect of the problem is the availability of spare parts. Computer hardware has a notoriously short life span in the marketplace. It is impossible, for example, to find today a commercial source for magnetic core, which was the standard computer memory as recently as 20 years ago.[83]

NARA is correct in its view that preserving actual computer hardware and software technology is not the most desirable way to preserve the utility of electronic records on a long-term basis. NARA seems to prefer the idea of preserving the records alone in a

---

[80] Id. at 49.

[81] Id. at 46.

[82] "Preservation of operating systems software would depend on the preservation of hardware to which the operating systems are tied. Applications level software in general is not bound to specific hardware or operating systems. This is especially true for software written in standard, nonproprietary languages. However, what is generally true is often false in particular cases. Even standard computer languages evolve over time and there is not always complete compatibility between new and previous versions. Major software manufacturers often develop and sell unique, proprietary implementations of standard languages. Specific applications software is often designed to take maximum advantage of the specific hardware and operating system used by the agency for which the programs are written. Because of factors such as these, there is a high probability that software preserved in its original stage will not run on the advanced computers of future generations." Id.

[83] Id. at 46.

software-independent, hardware-independent format. This "records-only" approach would be less complicated and less expensive. The problem is that it will not ensure the meaningful preservation of all important electronic records.

NARA itself recognizes that the "records only" approach has shortcomings and that "there probably will be cases where it is not possible to preserve functionality without preserving technology." [84] NARA's view that the problems presented by data files created by applications programs will be "exceptional" [85] is simply too limited.

NARA has not, however, fully considered the sweeping effects of the computer revolution on the operations of the Federal Government. For example, NARA does not believe that it is necessary to know the variety of word processing software being used by Federal agencies. The reason cited is that all word processors can produce an ASCII version of a document.[86] While this is true, the limitations of ASCII make it a poor choice for an archival document preservation standard. ASCII cannot preserve all basic features of ordinary documents, such as underlining, italics, footnotes, etc.[87]

Not surprisingly, NARA has not yet accessioned any significant volume of word processing files and does not anticipate that this will become a sizeable requirement for some years.[88] This may be a reasonable assessment, but a large and growing percentage of Government records is being created by word processing programs on personal computers equipped with floppy disks. The archival problems that will result from the use of word processors by Federal agencies must be addressed soon.

There are other consequences of the personal computer revolution that NARA has yet to confront. For example, NARA currently requires that computer files transferred for preservation be written on half-inch magnetic tape in one of several standardized formats.[89] This is a workable approach for mainframe computers that can create copies of records on magnetic tape. But few, if any, personal computers can produce magnetic tape. The Archives will not accept for permanent preservation electronic data stored on any medium other than computer tape. Thus, the portable media used regularly by personal computers—floppy disks—will not be accepted.[90]

It is not surprising that NARA has encountered compliance problems with its current requirement that electronic records be transferred in a format which does not require specific software or hardware to read the files. Several examples provided by NARA illus-

---

[84] Id. at 47.
[85] Id. at 48.
[86] Id. at 45.
[87] NARA is evaluating the adequacy of ASCII and other standards for preserving the content of electronic documents. Id. at 45.
[88] Id. at 45.
[89] Id. at 80. Files must be written in EBCDIC or ASCII, without internal control characters, on 7 or 9 track open-reel magnetic tape, recorded at 800, 1,600, or 6,250 bytes per inch, and blocked no higher than 30,000 bytes.
[90] The congressional investigation of the Iran-Contra Affair was aided by the recovery of documents from word processing diskettes gathered from staff offices of the National Security Council. See "Iran-Contra Report" at 689-90.

trate the problems of hardware and software dependence, obsolescence, and lack of documentation:

The National Military Command Center Information Processing System (NIPS) files from DOS constitute the largest group of system dependent records we have received. NIPS was a data base management system which ran on IBM mainframes. It is no longer supported. NIPS permitted unusual record structures which now make it difficult to process the data. NARA has used both its own staff and contractors to "de-NIPS" these files. We still have 156 reels in NIPS format. Even when the NIPS control codes are removed, the files remain difficult to process because of the unusual file structures. The Agent Orange Task Force reported to us that they did not succeed in using either the NIPS or "de-NIPSed" version of the Herbicide File, which they acquired to analyze the impact of Agent Orange.

Bureau of the Census files prior to 1989 threaten to eclipse the NIPS problem. The Bureau reported to us in May that they have over 4,000 reels of tape, containing permanently valuable data, which are difficult, if not impossible to use because they are in CENIO (Census Input/Output) format or because the files have been compressed on an ad hoc basis.

The United States Railway Authority Case Tracking/Document Management System (1920–1981) was created using the BASIS data base management system, which is a product of Batelle Laboratories. Unfortunately, this data base was created with an early version of BASIS. The current version of BASIS cannot read the older format. This case is especially difficult because BASIS stores data using a proprietary algorithm which scatters individual records in an undetermined and unrecoverable fashion. The format in which the data was received by NARA was even worse. Through expert sleuthing, a NARA contractor determined that the files were in a format used by DEC for backing up hard disks, on a device rather than file basis. Our inability to read this data is especially frustrating because it constitutes the only finding aid to a large volume of paper records in the National Archives.

In 1988, NARA reviewed data from 600 United States Information Agency studies which had been transferred. Besides identifying numerous studies which lacked adequate documentation, the review identified 249 studies with data in multipunch format, which is a survivor of the days when all data were stored on Hollerith cards. These files will have to be reformatted to be used. Another 21 files are in software dependent formats.

The data base containing testimony and diary summaries from the Watergate affair, which NARA received from the Senate Select Committee and the Office of the Watergate Special Prosecutor, requires BIBSYS software to access the data.[91]

None of this discussion is intended as specific criticism of past NARA policies for the accessioning of records on magnetic tape.

[91] "Thibodeau Letter."

But there are no new policies that respond adequately to recent information technology developments. The long lag time between the creation of records and NARA's acquisition of those records allows an opportunity for NARA to update its policies. NARA must make better use of that opportunity.

NARA's general perspective on electronic records issues appears too heavily based on NARA's experience with paper documents and magnetic tape. NARA was established to accept paper documents from agencies for long-term preservation, and it generally does this job well. This traditional approach was translatable to magnetic tapes produced by mainframe computers. NARA established standards, and the agencies have complied, albeit not completely. Preserving paper and tape, indexing the records, and providing for access and use present few of the difficulties of modern computer records.

NARA policies for electronic records appear to be an extrapolation from its current policies for paper and magnetic tape. NARA recognizes that there are new problems presented by electronic records, but these problems are explained away or minimized rather than confronted. NARA has not fully recognized either the scope or the depth of the problems presented by electronic records.

The "conventional wisdom" [92] for preserving electronic records is now out of date. New techniques are needed to cope with the dependence of electronic records on software and hardware.[93] The computer has revolutionized the way that documents are created, the nature of those documents, and the media on which information is recorded. Information technology that is 20 years old is already obsolete. Today's information technology is likely to be obsolete in a few years. One archivist describes the challenge of technological change this way:

> The problem with preservation of electronic media is that the media can easily be preserved longer than the capability of reading the signals recorded on them. Magnetic and optical media for recording of sound, image and data are all subject to market forces and technological change which are occurring at a rate that requires us to continuously recopy media to newer physical and logical formats in order to preserve access.[94]

The National Archives must adjust its policies and its mission not just to today's technology but to the fact that changes in technology will occur constantly and rapidly in the future. NARA needs to take a much longer term view of its functions at the same time as it responds more effectively to current problems. The computer revolution may require an archival revolution as well.[95]

---

[92] "Hedstrom & Kowlowitz" at 34.

[93] Id.

[94] D. Bearman, "Archival Methods," 3 "Archives and Museum Informatics Technical Report" 21 (Spring 1989).

[95] One suggestion is to replace the traditional archival concept of "permanent retention" in favor of a management formula based on the continuing value of the records. See id. at 17-27. This idea is based on the presumption that changing information technology will impose a continuing requirement to transfer data from one medium to another at least once a decade. Id. at 21. The committee expresses no opinion at this time, but the debate over technology and the role of an archives is welcome.

NARA is taking some useful steps. It has recognized that a promising long-term solution to the archival problems arising from computer records is the development and implementation of standards [96] for electronic records.[97] NARA also reports the development of an agenda for standards development and promotion, including continuing evaluation of emerging standards, collaboration with NIST, providing input to Open Systems Interconnection implementors workshops and other relevant standards implementors workshops, participation in the development of DOD's Computer-assisted Acquisition Logistics Systems, participation in American National Standards Institute standards development committees, and collaboration with State archives and professional archivists associations in promoting standards for electronic records.[98]

There is some evidence, however, that the interests of the archival community have not been represented in standards setting activities.[99] Further, a preliminary investigation by NIST has shown that no single standard will be sufficient for use by NARA for either data bases or documents.[100] Also, since the adoption of standards typically lags behind the use of new technology, NARA should not use the lack of standards as an excuse for refusing to consider the utility of new technology. For example, agencies are making increasing use of optical media, but NARA will not accept optical disks at this time.

NARA's task is further complicated because it has had little input into the design of the electronic recordkeeping systems selected by Federal agencies. NARA must live with decisions made years earlier. Lack of consultation during the planning of electronic record systems may make it difficult, expensive, or impossible to preserve records years later.

Even if a standard format for long-term storage can be identified, a computer information system may still create records that require conversion to permit accessioning. For example, the output of different word processing programs may require different conversion techniques. It is not clear how the responsibility and cost for this conversion will be allocated between the agencies and NARA. It does not appear that these problems have been fully considered.[101] A better approach may be to find a solution that meets archival needs at the earliest possible point in the life cycle of records so that additional processing is minimized or eliminated altogether.

---

[96] See part IV, section B-3 of this report.

[97] "The development and implementation of standards for electronic files would certainly assist NARA in preserving electronic records. In 1987, NARA signed an interagency agreement with the National Institute of Standards and Technology (NIST) to obtain the assistance of the National Computer Systems Laboratory in evaluating national and international standards for electronic records transfer and preservation, including standards applicable to data bases, text documents and technical documents containing drawings, graphics, and other content, in electronic form." "Wilson Letter," "Hearing" at 52.

[98] Id. at 53.

[99] "NHPRC Report" at 6.

[100] Id. at 12.

[101] NARA has suggested that cost and complexity of preservation may result in the destruction of records that might otherwise be accessioned. "Wilson Letter," "Hearing" at 47. Since unlimited preservation funds are not available, this could be a necessary policy in some cases. This should neither please historians nor provide an excuse for the wholesale destruction of classes of records.

The task of preserving Federal electronic records would be simplified if recordkeeping rules and procedures approved by NARA were built into major electronic information systems from the outset, with long-term preservation needs a mandatory system design consideration. The personal computer revolution makes centralized control of computer records more difficult. Unless archival considerations are taken into account at the earliest possible stage, many computer files may disappear without ever being reviewed for historical value.

It also would help considerably if Federal agencies integrated records management and preservation functions with their information resources management staffs and activities, particularly in electronic information systems design and training. Today, the two functions are too often carried out in isolation from each other.[102]

The Office of Management and Budget has emphasized both the importance of compatibility and of total life cycle costs when agencies acquire automated information technology.[103] There is no specific OMB guidance that refers to archival problems. In any event, it is not clear that OMB's guidance is having much influence in this area.

Finally, even if suitable standards for computer records can be developed and implemented soon—and the complexity of the task suggests that it will take a considerable amount of time—there will still be a large body of records created in the pre-standard era that may require preservation. Some of these records may be addressed by conversion software developed after standards have been established. It is unclear, however, that NARA's needs will be fully met either by the commercial marketplace or by the adoption of standards. NARA may need to pursue conversion of available records before time and technological developments make it impossible. For some older Government computer records, NARA will have to take the lead in finding effective long-term preservation methods.

Findings: NARA's current policies are inadequate to assure the long-term preservation of electronic records. Both evolutionary and revolutionary changes in NARA's policies, methods, and procedures may be needed to respond to the challenge of preserving electronic records.

NARA is not currently prepared to accession some computer records created by Federal agencies that will be candidates for preservation in the next few years.

Too little attention is being devoted by Federal agencies to planning for the archival needs of information in electronic recordkeeping systems that are in use and being planned.

---

[102] There is evidence that the same problems exist at the State level. Solutions being pursued at the State level include recognition of the life cycle of records and integration of data processing, information resources management, and archival functions. See, e.g., Kentucky Information Systems Commission, I "Managing Information Resources For Kentucky" 37-41 (1990) ("Information resources planning is facilitating the integration of records management and data management function, or at least bringing the need for integration into focus."); "Hedstrom & Kowlowitz" at 38 ("Developing a comprehensive records management program to handle modern public records is a challenge for agency staff, records managers, and archivists.") See also Joint Committee on Information Technology Resources, "Florida's Information Policy: Problems and Issues in the Information Age" (1989).

[103] Office of Management and Budget, "Management of Federal Information Resources," 50 Federal Register 52730, 52737 (December 24, 1985) (Circular No. A-130).

Recommendations: NARA needs to undertake a thorough review of its role in the long-term preservation of computer records. NARA's policies should address both the preservation of computer records and the practical utility of the records for future users.

NARA needs to recognize and take steps to confront the new preservation problems created by computer records and especially by the creation of hardware and software dependent personal computer records. At a minimum, NARA should review the technical characteristics of the electronic recordkeeping systems used within the Federal Government to assess preservation problems.

The Federal Government needs to do more long-range planning and preparation for the archival preservation of records of enduring value maintained in electronic formats. This planning must be reflected in selection of software and hardware for electronic recordkeeping systems and in the development and implementation of standards for computer records. NARA and the Office of Management and Budget should take the lead in requiring other agencies to consider long-term archival needs.

Federal agencies that use electronic recordkeeping systems should integrate records management and preservation functions with information resources management staffs and activities, particularly in systems design and training. NARA and OMB should take the lead in making preservation a more important part of information resources management.

Recordkeeping rules and procedures should be built into major electronic information systems from the outset. Long-term preservation needs should be a mandatory design consideration. NARA should work with the General Services Administration, the National Institute of Standards and Technology, and OMB to sensitize the computer industry and Federal agencies to the preservation requirements of the Federal Government.

NARA and NIST should continue their active involvement in Federal and private sector standard setting activities, including efforts to establish format standards for electronic records that will minimize or eliminate hardware and software dependent computer records.

NARA should make recommendations to the Congress for amendments to the law that are needed to reflect the changes in recordkeeping practices resulting from the widespread use of computers in the Federal Government.

NARA should respond to the recommendations in this report by submitting a written report to the Congress not later than July 1, 1991.

D. RELATED RECORDS MANAGEMENT PROBLEMS

General records management shortcomings have some bearing on the ability of the Federal Government to preserve electronic records. For example, general Federal recordkeeping practices, whether on paper or in electronic formats, have been found to be inadequate by several independent groups. In 1985, the Committee on Records of Government found that "the condition of Federal executive branch records is, with rare exceptions, deplorable."[104] The

[104] "Report of the Committee on the Records of Government" at 9.

Records Committee found that executive branch records have overwhelmed the system and resources assigned to manage them. A 1989 study of electronic recordkeeping in the U.S. Government by the National Academy of Public Administration also found that "[t]here is a general lack of knowledge about, and interest in, recordkeeping in general."[105] These problems are troubling to the committee, but they go beyond the scope of this report.

Other uncertainties are associated with the scheduling of records for long-term preservation. Federal records laws provide that agency records no longer needed for the transaction of current business be evaluated against records schedules that define whether they need to be preserved.[106] Records with long-term value are ultimately transferred to NARA.

Studies have shown, however, that this procedure is often not implemented. Agency records managers are often assigned to an administrative level in their organizations too low for them to obtain access to important documents in other offices. As a result, "records that document the major plans, programs, policies, and technical achievements or failures of agencies frequently do not get identified, processed, and preserved."[107]

With electronic records, the situation is further clouded by a widespread lack of recognition that formal recordkeeping requirements even apply.[108] For example, the "Model Framework for Management Control Over Automated Information Systems," jointly issued by the President's Council on Management Improvement and the President's Council on Integrity and Efficiency in 1988, did not recognize that the Federal Records Act applies to automated information systems. This guidance did not include records retention and disposition in its model framework, except records covered by the Privacy Act of 1974.[109]

The question of whether information in electronic formats is a "record" at all has been a recurring issue in several different contexts.[110] One court case specifically involving the preservation of electronic records is pending.[111]

---

[105] "NAPA Study" at C-2.

[106] 44 U.S.C. § 3301 (1988).

[107] Joint Committee on Archives of Science and Technology, "Understanding Progress as Process" (1983) (Society of American Archivists).

[108] It is surprising that there is a wide variance in the capabilities of agencies to deal with issues surrounding electronic recordkeeping. There are some agencies that "remain barely aware that the electronic age is upon them." See "NAPA Study" at 24.

[109] "Wilson Letter," "Hearing" at 51-52.

[110] See, e.g., the opening statement of Chairman Bob Wise at "Federal Information Dissemination Policies and Practices," Hearings before the Government Information, Justice, and Agriculture Subcommittee of the House Committee on Government Operations, 101st Cong, 1st. Sess. (1989) ("[C]hanges in technology are making our information dissemination laws obsolete. Most of our statutes were written in an era when all information was on paper. Today, more information is kept in electronic formats, and this raises new problems that existing laws do not resolve").

See also the testimony of Patti A. Goldman, Public Citizen Litigation Center, regarding the applicability of the Freedom of Information Act to electronic records. Id. at 474-503. This issue was also discussed in an earlier committee report. See Committee on Government Operations, "Electronic Collection and Dissemination of Information by Federal Agencies: A Policy Overview," House Report No. 99-560, 99th Cong., 2d Sess. 32-33 (1986).

See "NHPRC Report" at 5 ("[A]rchivists must examine and refine the legal definition of a record in the context of electronic technology").

[111] See text accompanying note 34.

The physical state of records at the time of transfer to the National Archives relates to another set of problems.[112] While NARA often accessions records earlier than 30 years, according to law, accession of documents cannot be compelled before the end of the specified 30 year period. For electronic records this could be disastrous. The National Institute of Standards and Technology (NIST) estimates the longevity of modern magnetic tape to be about 20 years under ideal storage conditions. The longevity of other electronic media, like floppy disks, or media improperly stored, may be considerably less.[113] NARA recently revised its regulations covering electronic records.[114]

Magnetic computer tape has emerged over the past two decades as the prevalent mass storage medium used for both intermediate and archival data storage by the Federal Government. With storage at 6,250 bits per inch, a computer tape holds information equivalent to 75,000 pages of text. Magnetic tape is durable, portable, and inexpensive, but it must be properly managed, maintained, and stored to prevent the loss of recorded signals.[115] Data deterioration can easily be caused by physical damage such as mishandling, contamination, and poor storage.[116] Recommended procedures are recopying of tapes every 10 years, "exercising" of stored tape by rewinding annually, and separate storage of backup copies. With proper maintenance, data stored on magnetic tapes can be preserved indefinitely.[117]

There is evidence that many Federal agencies are not adequately protecting the media on which their electronic records are stored from deterioration. For example, the General Accounting Office recently released a report documenting extensive problems at NASA in managing, documenting and maintaining computer tapes containing data from various space exploration missions.[118]

The National Archives will not officially accept records under its jurisdiction unless they have been arranged and described. With electronic records, agencies are required to transfer files both with adequate and accurate documentation and in a form that is hardware independent. The systematic identification and processing of

---

[112] NARA's maintenance requirements for magnetic computer tape can be found at 36 CFR § 1234.28(g) (published at 55 Federal Register 19220 (May 8, 1990). See also 36 CFR § 1228.188 (1989).

[113] "NAPA Study" at 44-45.

[114] See 55 Federal Register 19216 (May 8, 1990) (final rule amending 36 CFR Part 1234 on Electronic Records Management). Identical rules were also issued at the same time by the Information Resources Management Service of the General Services Administration. See id. at 19221 (final rule amending 41 CFR Part 201-45 on Electronic Records Management).

The new electronic records rules include some useful requirements. For example, the rules requires that official file copies of text documents on electronic media shall "provide a standard interchange format when necessary to permit the exchange of documents on electronic media between agency computers using different software/operating systems and the conversion or migration of documents on electronic media from one system to another." 36 CFR § 1234.22(a)(3). It is far from clear, however, that agencies will be able to carry out this requirement. Agencies can avoid it by determining that the interchange is not necessary. NARA expressly declined to prescribe the use of any specific standard interchange format. 55 Federal Register 19217 (May 8, 1990).

[115] "GAO NASA Report" at 9.

[116] "Report of the Committee on the Records of Government" at 29.

[117] Proper storage can present an immense task. NASA has over 1.2 million reels of magnetic tape. "GAO NASA Report" at 3-4. The equipment and personnel requirements to maintain these tapes are equally large. These requirements raise a serious question about the practical value of magnetic tape as a long-term preservation media.

[118] Id.

records of value from the vast collections of executive documents in Federal records centers can be both labor intensive and expensive. Where agencies are not willing to commit the resources necessary for any required "preprocessing" of records required by NARA, transfer is often postponed and records continue to be inaccessible.[119]

The challenge confronting NARA is to define a means of preserving in a nonproprietary, standard manner, data bases and text information including both raw data and relationship information so that electronic records transferred to the Archives can become software/hardware-independent.[120]

○

[119] "Report of the Committee on the Records of the Government" at 24.
[120] "NIST Report" at 6.