United States General Accounting Office

GAO

Report to Congressional Requesters

June 2002

INFORMATION MANAGEMENT

Challenges in Managing and Preserving Electronic Records

GAO-02-586

GAO
Accountability★Integrity★Reliability

Highlights

INFORMATION MANAGEMENT Challenges in Managing and Preserving Electronic Records

Highlights of GAO-02-586, a report to Congressional Requesters

Why GAO Did This Study

In the wake of the transition from paper-based to electronic processes, federal agencies are producing vast and rapidly growing volumes of electronic records. The difficulties of managing, preserving, and providing access to these records represent challenges for the National Archives and Records Administration (NARA) as the nation's recordkeeper and archivist. GAO was requested to (1) determine the status and adequacy of NARA's response to these challenges and (2) review NARA's efforts to acquire an advanced electronic records archiving system, which will be based on new technologies that are still the subject of research.

What GAO Recommends

GAO recommends that the Archivist of the United States develop documented strategies to raise awareness of the importance of records management programs and for conducting systematic inspections of these programs. In addition, to reduce risks, GAO recommends that the Archivist reassess the schedule for acquiring the new archival system so that the agency can complete key planning tasks and address IT management weaknesses. In commenting on a draft of this report, the Archivist agreed with our recommendations and offered clarifications, which we have incorporated as appropriate.

What GAO Found

NARA has taken action to respond to the challenges associated with managing and preserving electronic records. In 2001, NARA completed an assessment of the current federal recordkeeping environment. This study concluded that although agencies are creating and maintaining records appropriately, most electronic records (including databases of major federal information systems) remain unscheduled (that is, their value has not been assessed nor their disposition determined), and records of historical value are not being identified and provided to NARA for archiving. As a result, valuable electronic records may be at risk of loss. Part of the problem is that records management guidance is inadequate in the current technological environment of decentralized systems producing large volumes of complex records. Another factor is the low priority often given to records management programs and the lack of technology tools to manage electronic records. Finally, NARA does not perform systemic inspections of agency records management, and so it does not have comprehensive information on implementation issues and areas where guidance needs strengthening. Although NARA plans to improve its guidance and address technology issues, its plans do not address the low priority generally given to records management programs nor the inspection issue.

Recognizing the limitations of its technical strategies to support preservation, management, and sustained access to electronic records, NARA is planning to design, acquire, and manage an advanced electronic records archive; however, this project faces substantial risks. Although the electronic records archive project is in its initial stages, it is already falling behind schedule. Further, to acquire a major system of this kind, NARA needs to improve its information technology (IT) management capabilities, and although it has made progress in doing so, its efforts are not yet complete.

Master Copies of Electronic Records in NARA's Archives

Source: NARA.

This is a test for developing highlights for a GAO report. The full report, including GAO's objectives, scope, methodology, and analysis is available at www.gao.gov/cgi-bin/getrpt?GAO-02-586. For additional information about the report, contact Linda Koontz, 202-512-6240. To provide comments on this test highlights, contact Keith Fultz (202-512-3200) or email HighlightsTest@gao.gov.

Contents

Letter

1

Results in Brief 2
Background 3
NARA Is Responding to Challenges of Electronic Records Management 15
NARA's Effort to Acquire Advanced Electronic Archival System Faces Risks 23
Conclusions 32
Recommendations for Executive Action 33
Agency Comments and Our Evaluation 33

Appendixes

Appendix I: Objectives, Scope, and Methodology 37

Appendix II: Approaches to Archiving Electronic Records Provide Partial Solutions 39

Appendix III: NARA's Electronic Records Guidance Has Evolved 57

Appendix IV: Agencies Are Managing Large Volumes of Important Electronic Records 66

Appendix V: Comments from the National Archives and Records Administration 70

Glossary

75

Table

Table 1: Timeline for ERA Program 25

Figures

Figure 1: Removable Hard Drives and Backup Devices Used by Independent Counsel Staff 7
Figure 2: Master Copies of Electronic Records in NARA's Archives 8
Figure 3: OAIS Model and Its Components 15
Figure 4: Sample of XML Version of State Department Telegram 42
Figure 5: The Long Now Foundation Rosetta Disk Language Archive 51
Figure 6: Internet Archive Collection of Presidential Candidate Web Sites 54
Figure 7: Google's Usenet Archive 55

Abbreviations

ASCII	American Standard Code for Information Interchange
DARPA	Defense Advanced Research Projects Agency
DOD	Department of Defense
EAST	Examiners Automated Search Tool
ERA	Electronic Records Archive
GAO	General Accounting Office
GIS	Geographic Information System
GRS	General Records Schedule
GSA	General Services Administration
HTML	Hypertext Markup Language
HUD	Housing and Urban Development
IG	Inspector General
IT	information technology
NARA	National Archives and Records Administration
NASA	National Aeronautics and Space Administration
OAIS	Open Archival Information System
OMB	Office of Management and Budget
PMO	program management office
POP	persistent object preservation
PTO	U.S. Patent and Trademark Office
SAS	State Archiving System
SF	standard form
VERS	Victorian Electronic Record Strategy
WEST	Web Examiner Search Tool
XML	Extensible Markup Language

G A O
Accountability * Integrity * Reliability

United States General Accounting Office
Washington, D.C. 20548

June 17, 2002

The Honorable Stephen Horn
Chairman, Subcommittee on Government Efficiency,
Financial Management and Intergovernmental Relations
Committee on Government Reform
House of Representatives

The Honorable Ernest J. Istook, Jr.
Chairman, Subcommittee on Treasury,
Postal Service and General Government
Committee on Appropriations
House of Representatives

Agencies are increasingly moving to an operational environment in which electronic—rather than paper—records provide comprehensive documentation of their activities and business processes. Although this transformation has improved the way federal agencies work and interact with each other and with the public, it has also created the new challenge of managing and preserving vast and rapidly growing volumes of electronic records. Because these records document essential government functions and provide information necessary to protect government and citizen interests, their proper management is essential for ongoing government activities; further, the preservation of significant documents and other records is crucial for the historical record.

Overall responsibility for the government's electronic records lies with the National Archives and Records Administration (NARA), which carries out a dual mission for the nation: oversight of records management, which governs the life cycle of records (creation, maintenance and use, and disposition), and archiving, which is the permanent preservation of documents and other records of historical interest. In carrying out these missions, NARA and agencies use a process known as scheduling to assess the value of records and determine their disposition.

The challenges associated with managing and preserving electronic records have long been recognized throughout government. Because of concern about these issues, you requested that we review electronic records management and preservation activities at NARA. Our objectives were to

- determine the status of NARA's efforts to respond to governmentwide electronic records management problems and the adequacy of its planned actions and

- assess NARA's efforts to acquire an archival system for electronic records.

As part of our assessment of NARA's efforts to acquire an electronic records archiving system, you also asked that we identify alternative technologies under consideration for the long-term preservation of electronic records.

To address our objectives, we reviewed applicable guidance and other documentation; surveyed NARA's appraisal archivists working with federal agencies; reviewed records management activities and obtained the views of record managers in selected federal agencies managing large volumes of electronic records; and reviewed legal challenges to federal electronic recordkeeping practices. We reviewed agency and contractors' documentation for the electronic records archive program and assessed NARA's effort to develop or enhance its information technology capabilities. Further details on our objectives, scope, and methodology are provided in appendix I.

Results in Brief

NARA has taken action to respond to the challenges associated with managing and preserving electronic records. In 2001, NARA completed an assessment of the current federal recordkeeping environment; this study concluded that although agencies are creating and maintaining records appropriately, most electronic records (including databases of major federal information systems) remain unscheduled, and records of historical value are not being identified and provided to NARA for preservation in archives. As a result, valuable electronic records may be at risk of loss. Part of the problem is that records management guidance is inadequate in the current technological environment of decentralized systems producing large volumes of complex records. Another factor is the low priority often given to records management programs and the lack of technology tools to manage electronic records. Finally, NARA does not perform systematic inspections of agency records and records management programs, and so it does not have comprehensive information allowing it to identify records management implementation issues and areas where its guidance needs to be strengthened. NARA plans to improve its guidance and to address technology issues. However, NARA's plans do not address the low priority

generally given to records management programs nor the issue of systematic inspections.

Recognizing the limitations of its technical strategies to support preservation, management, and sustained access to electronic records, NARA is planning to design, acquire, and manage an advanced electronic records archive (ERA); however, this project faces substantial risks. NARA is behind schedule for the ERA system, largely because of flaws in how the schedule was developed. Further, to acquire a major system like ERA, NARA needs to improve its information technology (IT) management capabilities, and although it has made progress in doing so, its efforts are not yet complete.

Regarding alternative archiving technologies for electronic records, we found that archival organizations now rely on a mixture of evolving approaches that generally fall short of solving the long-term preservation problem. Appendix II provides a detailed discussion of these approaches.

In light of the continuing challenge of managing federal records, both electronic and otherwise, we are recommending that the Archivist of the United States develop a strategy for raising awareness of the importance of federal records management programs and for performing systematic inspections. In addition, to mitigate the risks associated with developing the new archival system, we are recommending that the Archivist reassess the schedule for this effort.

In commenting on a draft of this report, the Archivist stated that more must be done to address the enormous challenges in managing and preserving electronic records and agreed with the report's recommendations. He also offered clarifications concerning records management priority, inspections, and the ERA schedule that we have incorporated as appropriate.

Background

Advances in information technology and the explosion in computer interconnectivity brought about by the Internet are irreversibly changing the way we communicate and conduct business. Office automation applications and networked desktop computers are providing the capability to rapidly create and share electronic documents, use Web sites for executing business and financial transactions, and instantaneously communicate with individuals and groups. While the transformation from a paper-based to an electronic business environment has led to

improvements in the way federal agencies do business, both with each other and with the public, it has also created the new challenge of managing and preserving electronic records, which must be approached differently from their paper counterparts. Unlike paper records, electronic records are not tangible, come in many formats, and depend on the hardware and software with which they were created.

NARA's mission is to ensure "ready access to essential evidence" for the public, the President, the Congress, and the Courts. NARA's responsibilities stem from the Federal Records Act,[1] which requires each federal agency to make and preserve records that (1) document the organization, functions, policies, decisions, procedures, and essential transactions of the agency and (2) provide the information necessary to protect the legal and financial rights of the government and of persons directly affected by the agency's activities. Effective management of these records is critical for ensuring that sufficient documentation is created; that agencies can efficiently locate and retrieve records needed in the daily performance of their missions; and that records of historical significance are identified, preserved, and made available to the public. According to NARA, without effective records management, the records needed to document citizens' rights, actions for which federal officials are responsible, and the historical experience of the nation will be at risk of loss, deterioration, or destruction.

Under the act, NARA is responsible for oversight of records management and archiving. Records management—that is, the policies, procedures, guidance, tools and techniques, resources, and training needed to design and maintain reliable and trustworthy records systems—governs the life cycle of records from creation, through maintenance and use, to final disposition. Archiving is the permanent preservation of records documenting the activities of the government. NARA thus oversees agency management of temporary records used in everyday operations and ultimately takes control of permanent agency records judged to be of historic value.[2] Of the total number of federal records, less than 3 percent are designated permanent.

[1]44 U.S.C. chapters 21, 29, 31, and 33.

[2]NARA's regulations implementing the Federal Records Act are found at 36 CFR 1200-1280.

NARA Is Responsible for Oversight of Records Management

NARA is responsible for issuing records management guidance; working with agencies to implement effective controls over the creation, maintenance, and use of records in the conduct of agency business; providing oversight of agencies' records management programs; and providing storage facilities for certain temporary agency records. The Federal Records Act also authorizes NARA to conduct inspections of agency records and records management programs.

NARA works with agencies to identify and inventory records, appraise their value, and determine whether they are temporary or permanent, how long the temporary records should be kept, and under what conditions both the temporary and permanent records should be kept. This process is called scheduling. No record may be destroyed unless it has been scheduled, and for temporary records the schedule is of critical importance because it provides the authority to dispose of the record after a specified time period. Records are governed by schedules that are specific to an agency or by a general records schedule, which covers records common to several or all agencies. According to NARA, records covered by general records schedules make up about a third of all federal records. For the other two thirds, NARA and the agencies must agree upon specific records schedules. Once a schedule has been approved, the agency must issue it as a management directive, train employees in its use, apply its provisions to temporary and permanent records, and evaluate the results.

While the Federal Records Act covers documentary material regardless of physical form or media, records management and archiving were until recently largely focused on handling paper documents. With the advent of computers, both records management and archiving have had to take into account the creation of records in varieties of electronic formats. NARA's basic guidance for the management of electronic records is in the form of a regulation at 36 CFR Part 1234. This guidance is supplemented by the issuance of periodic NARA bulletins and a records management handbook, *Disposition of Federal Records*. NARA's guidance has two basic requirements. First, agencies are required to maintain an inventory of all agency information systems. The inventory should identify (1) the system's name; (2) its purpose; (3) the agency programs supported by the system; (4) data inputs, sources, and outputs; (5) the information content of databases; and (6) the system's hardware and software environment. Second, NARA requires agencies to schedule the electronic records maintained in its systems. Agencies must either schedule those records under specific schedules, completed through submission and approval of Standard Form 115 (SF 115), *Request for Records Disposition Authority*,

or pursuant to a general records schedule. NARA relies on this combination of inventory and scheduling requirements to ensure the management of agency electronic records consistent with the Federal Records Act.

NARA has also established a general records schedule for electronic records. General Records Schedule 20 (GRS 20) authorizes the disposal of certain categories of temporary electronic records. It has been revised several times over the years in response to developments in information technology, as well as legal challenges. (App. III provides a discussion of the evolution of electronic records guidance and legal challenges to GRS 20.)

As it stands now, GRS 20 applies to electronic records created both in computer centers engaged in large-scale data processing and in the office automation environment. With regard to computer centers, GRS 20 authorizes the disposal of certain types of scheduled electronic records associated with large database systems, such as inputs, outputs, and processing files. With regard to the office desktop environment, GRS 20 authorizes the deletion of the electronic version of records on word processing and electronic mail systems once a recordkeeping copy has been made. In addition, it authorizes deletion of electronically generated administrative spreadsheets and other administrative records that are included in recordkeeping systems that have been authorized for disposal by NARA. Since most agency "recordkeeping systems" are paper files, GRS 20 essentially authorizes agencies to destroy E-mail and word-processing files once they are printed. As already noted, records not covered by a general records schedule may not be destroyed unless authorized by a records schedule that has been approved by NARA.

GRS 20 does not address many common products of electronic information processing, particularly those that result from the now prevalent distributed, end-user computing environment. For example, although the guidance addresses the disposition of certain types of electronic records associated with large databases, it does not specifically address the disposition of electronic databases created by microcomputer users. In addition, while addressing word processing and E-mail records, GRS 20

does not address more recent forms of electronic records such as Web pages and portable document format (PDF) files.[3]

NARA Archives Permanent Records of Historical Interest

As the nation's archivist, NARA accepts for deposit to its archives those records of federal agencies, the Congress, the Architect of the Capitol, and the Supreme Court that are determined to have sufficient historical or other value to warrant their continued preservation by the U.S. government. NARA also accepts papers and other historical materials of the Presidents of the United States, documents from private sources that are appropriate for preservation (including electronic records, motion picture films, still pictures, and sound recordings), and records from agencies whose existence has been terminated, including Offices of Independent Counsel (see fig. 1).

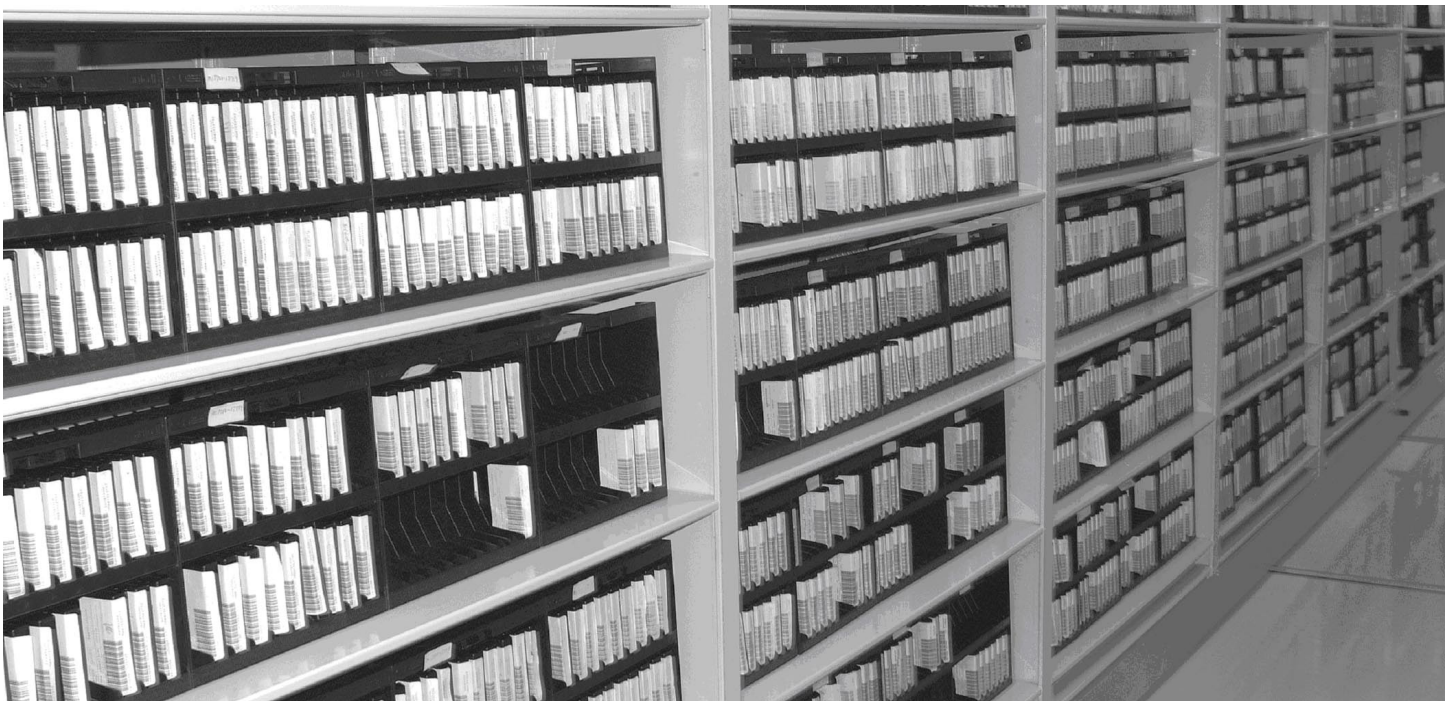
Figure 1: Removable Hard Drives and Backup Devices Used by Independent Counsel Staff

Source: NARA.

[3]PDF is a proprietary format of Adobe Systems, Inc., that preserves the fonts, formatting, graphics, and color of any source document, regardless of the application and platform used to create it.

NARA archives vast quantities of federal records in various formats. Its archival facilities (a network of regional archives) hold over 21 million cubic feet of original textual materials, while its multimedia collections include nearly 300,000 reels of motion picture film; more than 5 million maps, charts, and architectural drawings; over 200,000 sound and video recordings; about 9 million aerial photographs; nearly 14 million still pictures and posters; and over 87,000 computer data sets stored on computer tapes and cartridges (see fig. 2).

Figure 2: Master Copies of Electronic Records in NARA's Archives



Source: NARA.

In addition to its archives, NARA also manages the archival holdings of 10 presidential libraries, the Nixon presidential materials staff, and the Clinton presidential materials project. These include over 400 million paper records, over 15 million feet of film, nearly 10 million still pictures, nearly 100,000 hours of audio and video recordings, and almost half a million museum objects.

The types of electronic records that NARA currently accepts for archiving are limited to those that are independent of specified hardware or software and are in text-based formats, such as databases and certain text-based geographic information system (GIS)[4] files. NARA does not accept digital images, Web pages, word processor files, relational databases, or any records with complex structure.[5] (Although NARA does not as yet accept such files for archiving, they must still be scheduled.)

Management and Preservation of Electronic Records Pose Major Challenges

During the last four decades, archiving—the permanent preservation of information of enduring value for access by future generations—has undergone a major change. Before the advent of large bureaucracies supported by the now ubiquitous computer, archivists dealt with a scarcity of sources, with much of their efforts focused on tracking down unique manuscripts or recovering incomplete files.[6] The archived records were relatively durable—clay tablets, stone, parchment, vellum, or rag paper. Albeit scarce and often incomplete, these records come down through the centuries relatively intact and could be preserved with little or no difficulty. The growth of the government, complex organizations, and advent of the electronic age have reversed the conditions facing today's archives: rather than dealing with scarce sources, the archives are facing a flood of potentially valuable information stored on fragile materials, including pulp paper and computer tapes and disks.

While the preservation of information recorded on traditional materials such as paper or film requires significant resources, the current major archival challenge is the preservation of electronic records. Like traditional archival materials—books, papers, or film—electronic information is recorded on media that deteriorate with age. However, unlike the traditional archival materials, electronic records are stored in specific

[4]A geographic information system is a computer system for capturing, storing, checking, integrating, manipulating, analyzing, and displaying data related to positions on the Earth's surface. Typically, a GIS is used for handling maps of one kind or another. These might be represented as several different layers where each layer holds data about a particular kind of feature (e.g., roads). Each feature is linked to a position on the graphical image of a map.

[5]In January 2001, NARA directed agencies to provide a one-time "snapshot" of their public Web sites as they existed on or before January 20, 2001.

[6]National Research Council, *Preservation of Historical Records*, National Academy Press (Washington, D.C.: 1986).

formats and cannot be read without software and hardware—sometimes the specific types of hardware and software on which they were created.

The rapid evolution of information technology makes the task of managing and preserving electronic records complex and costly. Agencies are increasingly moving to an operational environment in which electronic—rather than paper—records provide comprehensive documentation of their activities and business processes. Part of the challenge of managing electronic records is that they are produced by a mix of information systems, which vary not only by type but by generation of technology: the mainframe, the personal computer, and the Internet. Each generation of technology brought in new systems and capabilities without displacing the older systems.[7] Thus, organizations have to manage and preserve electronic records associated with a wide range of systems, technologies, and formats.

The challenge of managing and preserving vast and rapidly growing volumes of electronic records produced by modern organizations is placing pressure on the archival community and on the information industry to develop a cost-effective long-term preservation strategy that would free electronic records of the straitjacket of proprietary file formats and software and hardware dependencies. This challenge is affected by several factors: decentralization of the computing environment, the complexity of electronic records, obsolescence and aging of storage media, massive volumes of electronic records, and software and hardware dependencies.

- *Decentralization of computing environment*: The challenge of managing electronic records significantly increases with the decentralization of the computing environment. In the centralized environment of a mainframe computer, it is relatively easy to identify, assess, and manage electronic records. This is not the case in the decentralized environment of agencies' office automation systems, where every user is creating electronic files that may constitute a formal record and thus should be preserved.

- *Complexity of electronic records*: Electronic records have evolved from simple text-based files to complex digital objects that may contain embedded images (still and moving), drawings, sounds, hyperlinks, or

[7]International Council on Archives, *Guide for Managing Electronic Records from an Archival Perspective* (Paris: February 1997).

spreadsheets with computational formulas. Some portions of electronic records, such as the content of dynamic Web pages, are created on the fly from databases and exist only during the viewing session. Others, such as E-mail, may contain multiple attachments, and they may be threaded (that is, related E-mail messages are linked into send–reply chains). These records cannot be converted to paper or text formats without the loss of context, functionality, and information.

- *Obsolescence and aging of storage media*: Storage media are affected by the dual problems of obsolescence and decay. They are fragile, have limited shelf life, and become obsolete in a few years. Few computers today have disk drives that can read information stored on 8- or 5¼-inch diskettes, even if the diskettes themselves remain readable.

- *Massive volumes*: Electronic records are increasingly being created in volumes that pose significant technical challenge to our ability to organize and make them accessible. For example, among the candidates for archiving are military intelligence records comprising more than 1 billion electronic messages, reports, cables, and memorandums, as well as over 50 million electronic court case files.

- *Software and hardware dependency*: Electronic records are created on computers with software ranging from word-processors to E-mail programs. As computer hardware and application software become obsolete, they may leave behind electronic records that cannot be read without the original hardware and software.

Past GAO Work Highlighted Electronic Records Challenges

In July 1999, we reported that NARA and federal agencies were facing the substantial challenge of preserving electronic records in an era of rapidly changing technology.[8] In that report we stated that in addition to handling the burgeoning volume of electronic records, NARA and the agencies would have to address several hardware and software issues to ensure that electronic records were properly created, maintained, secured, and retrievable in the future. We also noted that NARA did not have governmentwide data on the records management capabilities and programs of all federal agencies. As a result, we recommended that NARA conduct a governmentwide survey of agencies' electronic records

[8]U.S. General Accounting Office, *National Archives: Preserving Electronic Records in an Era of Rapidly Changing Technology*, GGD-99-94 (Washington, D.C.: July 19, 1999) (*http://www.gao.gov/archive/1999/gg99094.pdf*).

management programs and use the information as input to its efforts to reengineer its business processes. NARA's subsequent efforts to assess governmentwide records management practices and study the redesign of its business processes are discussed later in this report.

Agencies Are Beginning to Automate Management of Electronic Records

In response to the difficulty of manually managing electronic records, agencies are slowly turning to automated records management applications to help automate electronic records management life-cycle processes. The primary functions of these applications include categorizing and locating records and identifying records that are due for disposition, as well as storing, retrieving, and disposing of electronic records that are maintained in repositories. Also, some applications are beginning to be designed to automatically classify electronic records and assign them to an appropriate records retention and disposition category.

The Department of Defense (DOD), which is pioneering the assessment and use of records management applications, has published application standards and established a certification program.[9] The DOD standard, endorsed by NARA, includes the requirement that records management applications acquired by DOD components after 1999 be certified to meet this standard.[10] As of March 2002, DOD had certified 31 applications. NARA was testing one of the DOD-certified electronic records management applications, and it will be assessing the second version of the DOD standard to determine whether it can or should become a governmentwide standard.

Theory, Methods, and Model for Long-Term Preservation of Electronic Records Are Being Developed

NARA is not alone in facing the challenges posed by electronic records, particularly long-term preservation. There is a general consensus in the archival community that a viable strategy for the long-term preservation and archiving of electronic records has yet to be developed. Accordingly, archives scholars, national archival and library institutions, and private industry representatives are collaborating on major initiatives to develop the theoretical and methodological knowledge needed for the permanent

[9]Department of Defense, *Design Criteria Standard for Electronic Records Management Software Applications*, DOD 5015.2-STD (November 1997) (*http://www.dtic.mil/whs/directives/corres/html/50152std.htm*).

[10]DOD 5015.2-STD requires that records management applications be able to manage records regardless of their media.

preservation of records created in electronic systems. These initiatives include the following:

- The International Research on Permanent Authentic Records in Electronic Systems project is a major two-phase international research project in which archival and computer engineering scholars, national archival institutions (including NARA), and private industry representatives are collaborating to develop the theoretical and methodological knowledge required for the permanent preservation of authentic records created in electronic systems. The first phase of the project, focusing on records generated in databases and document management systems, was recently completed; the second phase (2002 to 2006) deals with the issues of authenticity, reliability, and accuracy of records produced in new digital environments.

- The Library of Congress' National Digital Information Infrastructure and Preservation Program is a national cooperative effort led by the Library to develop the strategy and technical approaches needed to archive and preserve digital information; NARA is also participating in this effort. The program is in an early stage; completion is not expected until 2004 or 2005, when the Library will provide recommendations to the Congress.

- NARA is collaborating in a joint effort on electronic record archiving with the Defense Advanced Research Projects Agency (DARPA), the U.S. Patent and Trademark Office, the National Partnership for Advanced Computational Infrastructure, and the San Diego Supercomputer Center. Led by DARPA, the collaboration aims to develop and demonstrate architectures and technologies for electronic archiving and the development of persistent object preservation, a proposed technique for electronic archiving (discussed in app. II).

These initiatives are all in their early stages; none of them has yet yielded proof-of-concept prototypes demonstrating the viability of a long-term solution to preserving and accessing electronic records.

Progress has been made, however, in the development of a standard model for electronic archiving systems. The Open Archival Information System (OAIS) model, which is currently emerging as a standard in the archival community, was initially developed by the National Aeronautics and Space Administration (NASA) for archiving the large volumes of data produced by space missions. However, the model is applicable to any archive, digital

library, or repository. As a standard framework for long-term preservation archives, the model defines the environment necessary to support a digital repository and the interactions within that environment. According to NASA, it also promotes the understanding and increased awareness of archival concepts needed for long-term digital information preservation and access, as well as for describing and comparing architectures and operations of existing and future archives.

Many institutions have already chosen to use the framework of the OAIS reference model to guide their digital preservation efforts, including the National Library of the Netherlands, NARA (in conjunction with the development of its electronic records archiving project), NASA's National Space Science Data Center, and many commercial organizations.

The OAIS model (see fig. 3) breaks the archiving system down into six distinct functional areas: ingest, archival storage, data management, administration, preservation planning, and access.

- In the ingest area, systems accept information submitted from outside the framework and prepare the contents for storage. This functional area also includes systems to generate descriptive information to allow future management within the archive.

- In the archival storage area, systems pass the information, now called archival information packages, into a storage repository, where it is maintained until the contents are requested and retrieved.

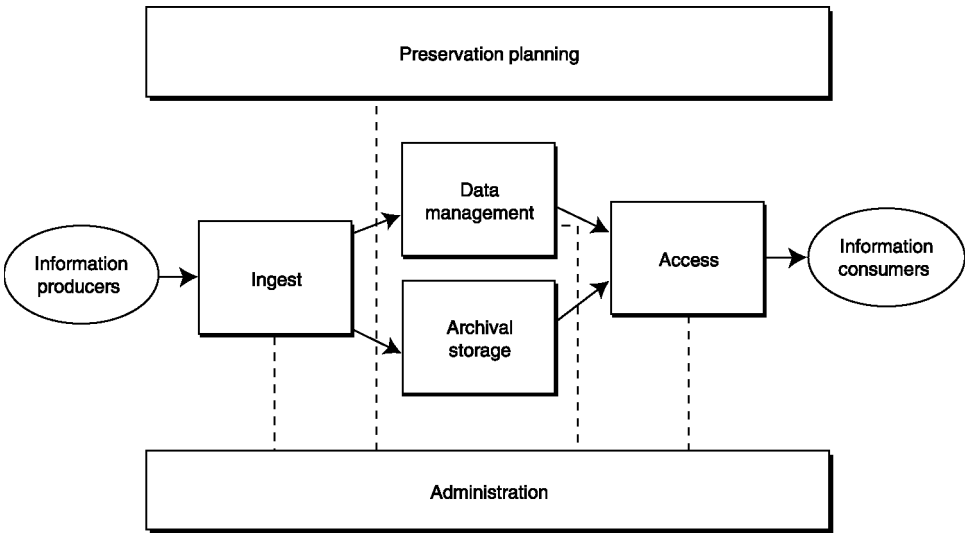
- The data management area encompasses the services and functions for populating, maintaining, and accessing both descriptive information that identifies and documents archive holdings and administrative data used to manage the archive.

- The administration area provides the services and functions for the overall operation of the archive system.

- In the preservation planning area, systems monitor the environment of the OAIS and provide recommendations to ensure that the information stored in the OAIS remains accessible, even if the original computing environment becomes obsolete.

- The access area includes systems that allow a user to determine the existence, description, location, and availability of information stored in the OAIS, allowing information products to be requested and received.

Figure 3: OAIS Model and Its Components



Source: Consultative Committee for Space Data Systems.

The OAIS framework does not presume or apply any particular preservation strategy. This approach allows organizations that adopt the framework to apply their own strategies or combinations of strategies. The framework does assume that the information managed is produced outside the OAIS, and that the information will be disseminated to users who are also outside the system. Because the model is simplified to include only functions common to all repositories, it allows institutions to focus on the approaches necessary to preserve the information.

NARA Is Responding to Challenges of Electronic Records Management

NARA is taking action to respond to long-standing problems associated with managing and preserving electronic records in archives. In 2001, NARA completed an assessment of governmentwide records management practices. This assessment concluded that although agencies are creating sufficient records and maintaining them appropriately, most electronic records remain unscheduled, and permanent records of historical value are

not being identified and provided to NARA for preservation and archiving. As a result, potentially valuable records may be at risk.

According to the study, the problems in electronic records management appear to stem from (1) inadequate governmentwide records management guidance and (2) the low priority traditionally given to federal records management functions and a lack of technology tools to manage electronic records. To address these problems, NARA now plans to (1) analyze key policy issues related to the disposition of records and improve its guidance and (2) examine and redesign, if necessary, the scheduling and appraisal process and make this process more effective through the use of technology. NARA's plans, however, do not address the low priority given to records functions. Further, these plans do not address the need to monitor performance of records management programs and practices on an ongoing basis.

NARA's Assessment of Federal Records Practices Identifies Problems

Records must be effectively managed throughout their life cycle, which includes records creation, maintenance and use, and scheduling and disposition. Agencies must create reliable records that meet the business needs and legal responsibilities of federal programs and (to the extent known) the needs of internal and external stakeholders who may make secondary use of the records. To maintain and use the records created, agencies are to create internal recordkeeping requirements for maintaining records, consistently apply these requirements, and establish systems that allow them to find records that they need. Scheduling is the means by which NARA and agencies identify federal records, determine time frames for disposition, and identify permanent records of historical value that are to be transferred to NARA for preservation and archiving. With regard particularly to electronic records, agencies are also to compile inventories of their information systems, after which the agency is required to develop a schedule for the electronic records maintained in those systems.

In 2001, NARA completed an assessment of governmentwide records management practices, as recommended in our prior work. The assessment included a recordkeeping study performed by a contractor—SRA International—and a series of records system analyses performed by NARA staff. The SRA study was based on a survey of federal employees representing over 150 federal government organizations and on 54 focus groups and interviews involving individuals from 18 agencies; the NARA staff's records system analyses focused on records management practices for key business processes in 11 federal agencies.

The resulting NARA/SRA study identified problems in agency records management.[11] Specifically, NARA's assessment of records management for key processes in 11 agencies concluded the following.

- **Records creation**: In general, the NARA study showed that the processes that were studied appeared to generate adequate records documentation.

- **Records maintenance and use**: For the most part, recordkeeping requirements were adequate, documented, and consistently applied. In addition, employees were generally able to find the records that they needed.

- **Records scheduling and disposition**: The study identified significant problems in both records scheduling and disposition. According to the study, many significant records—as well as most federal electronic records—are unscheduled. In addition to the unscheduled records, NARA identified several significant records that had been improperly scheduled. The study concluded that records scheduling was clearly a problem area.

Our review at four agencies (Commerce, Housing and Urban Development, Veterans Affairs, and State) provides confirmation of this result, eliciting a collective estimate that less than 10 percent of mission-critical systems were inventoried. The number of mission-critical systems at these four agencies was reported to be 907, according to information collected by the Office of Management and Budget in November 1999 as part of the federal government's effort to assess the Year 2000 computing challenge.[12] Thus for these four agencies alone, over 800 systems had not been inventoried and the electronic records maintained in them had not been scheduled. Scheduling the electronic records in a large number of major information systems presents an enormous challenge, particularly since it generally

[11]SRA International, Inc., *Report on Current Recordkeeping Practices within the Federal Government* (Dec. 10, 2001) (*http://www.nara.gov/records/rkreport.html*). Both the SRA study and the NARA staff analyses were reported within this document.

[12]The 24 major agencies reported 6,435 mission-critical systems. Subcommittee on Government Management, Information, and Technology, House Committee on Government Reform, *Federal Government Earns B+ on a Final Y2K Report Card*, news release (Washington, D.C.: Nov. 22, 1999).

takes NARA, in conjunction with agencies, well over 6 months to approve a new schedule.[13]

Failure to inventory systems and schedule records places these records at risk. The absence of inventories and schedules means that NARA and agencies have not examined the contents of these information systems to identify official government records, appraised the value of these records, determined appropriate disposition, and directed and trained employees in how to maintain and when and how to dispose of these records. As a result, temporary records may remain on hard drives and other media long after they are needed or could be moved to less costly forms of storage. In addition, there is increased risk that these records may be deleted prematurely while still needed for fiscal, legal, and administrative purposes.

The lack of scheduling presents particular risks to the preservation of permanent records of historic significance. NARA's study of 11 agencies found instances where valuable permanent electronic records were not being appropriately transferred to NARA's archives because these records had not been scheduled, appraised, identified as permanent, and placed under the control of the agency's records program. This lack of management control places these valuable records at increased risk of loss, destruction, and deterioration.

NARA's Records Management Guidance Has Not Kept Pace with the Challenges of Electronic Records

The NARA/SRA study identified the lack of sufficient governmentwide guidance as one cause of records management problems. As NARA has acknowledged, its policies and processes on electronic records have not yet evolved to reflect the modern recordkeeping environment: records created electronically in decentralized processes.[14] Despite repeated attempts to clarify its electronic records guidance through a succession of NARA bulletins, the current guidance remains incomplete and confusing. According to the study, for example, employees lack knowledge concerning how to identify electronic records and what to do with them once identified. The guidance does not provide disposition instructions for

[13]According to NARA, its current goals for schedule processing are 180 days for simple schedules and 365 days for complex schedules. In FY 2001 the median time for completing schedules was 237 days.

[14]National Archives and Records Administration, *An Overview of Three Projects Relating to the Changing Federal Recordkeeping Environment* (January 2001) (*http://www.nara.gov/records/rmioverview.html*).

electronic records maintained in many of the common types of formats produced by federal agencies, including PDF files, Web pages, and spreadsheets. To support their missions, many agencies must maintain such records—often in large volumes—with little guidance from NARA (see app. IV for a discussion of the records management challenges faced by selected agencies).

The NARA/SRA study concluded that while agencies appreciate the specific assistance from NARA personnel, they are frustrated because they perceive that NARA is not meeting agencies' broader needs for guidance and records management leadership. This study reported that agencies believe that NARA has a responsibility to lead the way in transitioning to an electronic records environment and to provide guidance and standards, as well as tools to enable agencies to follow the guidance. According to the study, some viewed NARA as leaving agencies to fend for themselves, sometimes levying impossible requirements that pressure agencies to come up with their own individual solutions.

Agency Records Management Programs Are Given Low Priority and Lack Technology Tools

The NARA/SRA study identified another cause of records management difficulties: the low priority generally afforded to records management programs. The study states that records management is not even "on the radar scope" of agency leaders. Further, records officers have little clout and do not appear to have much involvement in or influence on programmatic business processes or the development of information systems designed to support them. New government employees seldom receive any formal, initial records management training. One agency told NARA that records management is "number 26 on our list of top 25 priorities." The study also noted that federal downsizing may have negatively affected records management and staffing resources in agencies.

Further, records management is generally considered a "support" activity. Since support functions are typically the most dispensable in agencies, resources for and focus on these functions are often limited. This finding was echoed by a recent review of archival practices of research universities, corporate research and development programs, and federal science agencies, which noted that "agency records management programs

lack the resources to meet even the legally required standards of securing adequate documentation of their programs and activities."[15]

As indicated by the NARA/SRA study, a related issue is the technical challenge of electronic records management: effective electronic records management may require more sophisticated and expensive information technology (such as automated electronic records management systems) than was previously necessary for paper-based records management programs. Because management tends not to focus on records management, priority has not been given to acquiring or upgrading the technology required to manage records in an electronic environment. The study noted that technology tools for managing electronic records do not exist in most agencies, and further, that agency information technology environments have not been designed to facilitate the retention and retrieval of electronic records. As a result, despite the growth of electronic media, agency records systems are predominantly in paper format rather than electronic.

The study further noted that agencies planning or piloting automated electronic records management systems perform better recordkeeping than those without such tools. Typically, such agencies are already performing better recordkeeping, and they tend to invest in electronic records management systems because of the value they place on good records management. According to the study, many agencies are either planning or piloting information technology initiatives to support electronic records management, but their movement to electronic systems is constrained by the level of financial support provided for records management.

Inspections of Federal Electronic Records Programs Are Limited

A possible further cause of agency records management problems, not addressed in the NARA/SRA study, is the limited nature of NARA's current inspection program. NARA is responsible, under the Federal Records Act, for conducting inspections or surveys of agency records and records management programs and practices. Its implementing regulations require NARA to select agencies to be inspected (1) on the basis of perceived need by NARA, (2) by specific request by the agency, or (3) on the basis of a

[15]Center for History of Physics, American Institute of Physics, *AIP Study of Multi-institutional Collaborations: Final Report—Highlights and Project Recommendations*, College Park, MD (2001) *(http://www.aip.org/history/pubs/collabs/highlights.html)*.

compliance monitoring cycle developed by NARA.[16] In all instances, NARA is to determine the scope of the inspection. Such inspections provide not only the means to assess and improve individual agency records management programs but also the opportunity for NARA to determine overall progress in improving agency records management and identify problem areas that need to be addressed in its guidance.

Between 1996 and 2000, NARA performed 16 inspections of agency records management programs, or about 3 per year. These reviews were systematic and comprehensive, covering all aspects of an agency's records program. However, only 2 of the 24 major executive departments or agencies were evaluated, with most of NARA's evaluations focused on component organizations or independent agencies. Moreover, these evaluations frequently bypassed the issue of electronic records.

In 2000, NARA replaced agency evaluations with a new inspection approach—targeted assistance. NARA decided that its previous approach to inspections was basically flawed: besides reaching only a few agencies, it was often perceived negatively by agencies and resulted in a list of records management problems that agencies then had to resolve on their own. Under the targeted assistance approach, NARA enters into partnerships with federal agencies to provide them with guidance, assistance, or training in any area of records management. Services offered include expedited review of critical schedules, tailored training, and help in records disposition and transfer.

However, although this approach may improve records management in the targeted agencies, it is not a substitute for systematic inspections and evaluations of federal records programs. Because the targeted assistance program is voluntary and, according to NARA, initiated by a written request from the agency, relying on it exclusively could significantly limit NARA's evaluations of federal recordkeeping. First, only agencies requesting targeted assistance—presumably those already having greater appreciation of the importance of records management—are evaluated. Second, the scope and the focus of the targeted assistance are not determined by NARA but by the requesting agency.

[16]CFR 1220.54 (a).

NARA Is Addressing Records Management Problems, but Additional Opportunities Exist

NARA has recognized that its policy and regulations for the management and disposition of electronic records must be revised to provide agencies with clear and comprehensive guidance encompassing all types and formats of electronic records. Having completed its assessment of federal records management practices, NARA now plan a two-phase project to (1) analyze key policy issues related to the disposition of records and improve governmentwide guidance, and (2) examine and redesign, if necessary, the scheduling and appraisal process and make this process more effective through the use of technology.

According to NARA, the purpose of the first phase of the project is to analyze and make decisions, as necessary, on key policy issues related to determining the disposition of records. NARA plans to evaluate current legislation, regulations, and guidance to determine if these are adequate in the current recordkeeping environment. NARA expects the outcome of the first phase, scheduled for completion by the end of fiscal year 2002, to be policy decisions that support the appropriate disposition of all government documentation in today's multimedia environment.[17] These results are also intended, as recommended in our prior work, to inform the redesign of the current scheduling and appraisal process planned for the second phase of the project, the development of electronic recordkeeping requirements, and improvements to records management guidance and assistance to agencies.

In the second phase, NARA plans to examine and redesign, if necessary, the process used by the federal government to determine the disposition of records. This is planned as a multiyear process (2003 to 2006) during which NARA intends to address the scheduling and appraisal of federal records in all formats. Currently, it takes NARA well over 6 months to approve a new schedule. According to NARA, the extensive appraisal time delays action on the disposition of records and discourages agencies from submitting schedules, potentially putting essential evidence at risk. NARA has two goals for this project: (1) making the process for determining the disposition of records, regardless of medium, more effective and efficient and dramatically decreasing the amount of time it takes to get approval for the disposition of records from the Archivist of the United States, and (2) deciding how to appropriately apply technology to support the revised

[17]NARA expects the policy review phase to be completed by the end of 2002, but according to NARA, all new or revised policies will not be in place by that date. The entire project will not be complete until 2006.

process for determining the disposition of records as part of managing records throughout their life cycle.

Although NARA's plans address the need to improve guidance and determine how to use technology to support records management, these plans do not address another issue raised in its study: the low priority generally given to records management and the related lack of management commitment and attention to these functions. Without a strategy to establish senior-level agency commitment to records management and raise awareness of its importance to the federal government, these programs are likely to continue to be regarded by agency management and employees as low-priority "support" functions.

In addition, NARA's plans do not address the issue of systematic inspections. While the results of its recent study provide a baseline of governmentwide records management practices, NARA's targeted assistance approach does not provide systematic and comprehensive information to assess progress over time. Without this type of data, NARA will be impaired in its ability to determine if it is achieving results in improving agency records management. Further, NARA may not have the means to identify agency implementation issues and areas where its guidance needs to be clarified, augmented, and strengthened. The feedback provided by inspection is especially critical now as NARA plans to redesign the scheduling and appraisal process, and improve its guidance.

NARA's Effort to Acquire Advanced Electronic Archival System Faces Risks

Archiving—the final phase of records management for permanent records—presents a significant challenge when records are electronic. In light of the growth in the volume, complexity, and diversity of electronic records, NARA has recognized that its technical strategies to support preservation, management, and sustained access to electronic records are inadequate and inefficient. To address this challenge, the agency is pursuing two strategies. Its short-term strategy is to extend the useful life of its current systems and to create some new systems for archiving electronic records and for cataloging and displaying electronic records on-line. NARA's long-term strategy, on which it is placing its primary focus, is to contract with a private sector firm to acquire (that is, obtain) an advanced electronic records archive (ERA).

However, NARA faces substantial risks in implementing its long-term strategy. NARA is not meeting its schedule for the ERA system, largely because of flaws in how the schedule was developed. As a result, the

schedule will be compressed, increasing risks. Further, although NARA recognizes that to be successful it must improve its information technology (IT) management capabilities and has made progress in doing so, these efforts are not yet complete.

NARA Is Planning to Acquire an Advanced Electronic Records Archiving System

NARA's long-term strategic initiative is to develop an advanced electronic records archive. The agency's goals for this system are to preserve and provide access to any kind of electronic record, free from dependency on any specific hardware or software, so that the agency can carry out its mission into the future.

Although the new archival system is not yet formally defined, agency documents, public presentations, and interviews with agency officials and staff indicate, in broad outline, how they envision this system. It will probably be a distributed system, allowing the storage and management of massive record collections at a variety of installations, with accessibility provided via the Internet. It may be based on persistent object preservation, an advanced form of file format conversion and encapsulation (described in app. II) that is the subject of research sponsored by NARA and other organizations. A leading candidate for performing this encapsulation and capturing the necessary information is the Extensible Markup Language (XML), which provides a means for "tagging" (annotating) information in a meaningful fashion that can be readily interpreted by disparate computer systems (XML is further discussed in app. II).

NARA has indicated that ERA will be a major system, and that it is likely that it will be developed and implemented in several phases (or "builds"), with each phase adding more functions to the system. According to NARA, its development will take several years, and it will involve a significant expenditure of resources on program management, research, and systems development activities.

NARA is planning to award the contract for the new electronic archival system in January 2004. Table 1 is a timeline showing key tasks for the program.

Table 1: Timeline for ERA Program

Key ERA tasks	Completion dates
Develop vision statement	March 1, 2002[a]
Develop concept of operations	April 1, 2002[b]
Conduct market survey	June 28, 2002
Perform analysis of alternatives	July 22, 2002
Develop cost estimates	August 19, 2002
Develop high-level conceptual and functional requirements	September 24, 2002
Develop business case/economic analysis	September 30, 2002
Develop final functional requirements	December 2, 2002
Issue Request for Information	January 13, 2003
Release Request for Proposal	August 4, 2003
Fiscal year 2004 budget for ERA In effect	October 1, 2003
Award ERA contract	January 12, 2004

[a]Completed April 18, 2002.
[b]Completed in draft on April 1, 2002.

To assist in this effort, NARA contracted with Integrated Computer Engineering (ICE), Incorporated,[18] a private company experienced in systems development and acquisition. With the assistance of this contractor, NARA has been establishing the ERA program management office. Since July 2001, the program management office has been focused on developing the capability to manage the development and acquisition of the ERA system.

NARA is also funding two independent assessments of the research into the technology that is proposed for ERA. These two independent assessments, conducted by the National Academy of Sciences, will review research that NARA is now sponsoring, as well as alternative approaches. The first assessment is a technical review of the viability of persistent object preservation, the architecture for persistent archives of electronic records that is being researched by the National Partnership for Advanced Computational Infrastructure (see app. II). This assessment—scheduled

[18]On January 15, 2002, American Systems Corporation (ASC) announced its acquisition of ICE, Inc. According to the ERA project manager, this change does not affect the status of NARA's contract with ICE, Inc.

for completion on January 31, 2003—will address the adequacy and soundness of the persistent object preservation architecture as a whole, as well as its major components, from the points of view of computer science, systems engineering, and archival sciences. NARA has stated that the assessment of the persistent object information management architecture and its technical validation should be completed before ERA is developed. In its fiscal year 2002 budget hearings, NARA referred to the articulation of the persistent object preservation architecture as the one "major dependency" in its strategy for acquiring an ERA system.

The second assessment will identify and evaluate alternative methods for digital preservation of records, examine the operational use of the Internet for digital archiving, and identify those aspects of the preservation of electronic records that cannot be adequately addressed either by state-of-the-art information technology or by technologies under development. It will also address the feasibility of commercializing new ideas from research. According to NARA, the second assessment is to be completed 6 to 9 months after the first.

ERA Schedule Faces Significant Risks

Although the ERA project is still in its initial stages, it is already falling behind schedule. As shown in table 1, the initial deliverables for design and acquisition are late: the vision statement, due March 1, was not completed until April 18, and the concept of operations,[19] due April 1, was delivered in draft form on that date and had not been finalized as of May 31. This lateness can be attributed to flaws in how the schedule was developed. In its tracking of ERA risks, NARA has acknowledged that the schedule for completion of tasks was based on incomplete work projections, and that its deadlines may not be achievable. Rather than constructing a plan based on estimates of the amount of work and resources required to complete each task, NARA constructed a "success oriented" schedule that was planned around ensuring that ERA was funded beginning in fiscal year 2004.

In addition, the ERA program management office is behind schedule on its efforts to develop the plans and guidance to strengthen its capability for managing the acquisition and deployment of ERA. In July 2001, with the help of its systems development and acquisition contractor, the office began focusing on developing these plans and procedures. We tracked

[19]A concept of operations is a document that describes characteristics of the system from the user's viewpoint.

planned and actual completion dates for 13 policy and planning documents that the program management office needs in order to develop and acquire a major system (according to NARA and its contractor). To date, however, only 7 of the 13 documents have been completed.[20] The 7 that have been delivered were late by an average of over 2 months. The initially planned delivery dates of the other 6 documents have passed; on average these are late by almost 4 months.[21]

Besides the approach taken to constructing the schedule, another contribution to schedule slippage may be NARA's slow start in hiring full-time government staff for the ERA program management office. For fiscal year 2002, NARA was authorized 16 positions for the ERA program office. However, as of April 2002, NARA had only 5 full-time staff on board.

NARA Is Strengthening IT Management Capabilities, but These Efforts Are Incomplete

Acquiring a major IT system such as the planned electronic archival system is a significant challenge for a relatively small organization like NARA, whose IT management capabilities are relatively limited. In its fiscal year 2002 budget hearings, NARA indicated that it must strengthen its IT management capabilities and infrastructure to support the ERA program, and NARA is currently taking steps to do so in three key areas: IT investment management, enterprise architecture, and information security. None of these efforts, however, is yet complete.

Sound IT Management Capabilities Contribute to Success in Acquiring IT Systems

IT investment management provides a systematic method for agencies to minimize risks while maximizing the return on investments. The Clinger-Cohen Act requires agency heads to implement a process for maximizing the value and assessing and managing the risks of an agency's IT investments. Our research of leading private and public sector organizations' IT management practices indicates that effective investment management requires the use of defined and disciplined investment management processes.

[20]The seven completed documents were the acquisition strategy, configuration management plan, risk management plan, quality assurance plan, life-cycle model, requirements management plan, and technology research plan.

[21]The six uncompleted documents were the revised program management office (PMO) organization, PMO billet roles/responsibilities, metrics plan, PMO training needs assessment, ERA PMO training plan, and program management plan.

An enterprise architecture provides a description—in useful models, diagrams, and narrative—of the mode of operation for an agency. It describes the agency in both (1) logical terms, such as interrelated business processes and business rules, information needs and flows, and work locations and users; and (2) technical terms, such as hardware, software, data, communications, and security attributes and standards. An enterprise architecture provides these perspectives both for the current environment and for the target environment, as well as a transition plan for sequencing from the current to the target environment. Managed properly, an enterprise architecture can clarify and help optimize the dependencies and relationships among an agency's business operations and the underlying IT infrastructure and applications that support these operations.

Information security is an important consideration for any organization that depends on information systems to carry out its mission. Our study of security management best practices, as summarized in our 1998 executive guide,[22] found that leading organizations manage their information security risks through an ongoing cycle of risk management. This management process involves (1) establishing a centralized management function to coordinate the continuous cycle of activities while providing guidance and oversight for the security of the organization as a whole, (2) identifying and assessing risks to determine what security measures are needed, (3) establishing and implementing policies and procedures that meet those needs, (4) promoting security awareness so that users understand the risks and the related policies and procedures in place to mitigate those risks, and (5) instituting an ongoing monitoring program of tests and evaluations to ensure that policies and procedures are appropriate and effective.

NARA Is Improving Its IT Investment Management Processes

The Clinger-Cohen Act of 1996 requires agencies to establish an IT investment process that provides the means for senior management to obtain timely information regarding the progress of investments in an information system, including a system of milestones for measuring progress in terms of cost, timeliness, quality, and the capability of the system to meet specified requirements. Weak IT investment management processes significantly increase the risk that agency funds and resources will not be efficiently expended.

[22]U.S. General Accounting Office, *Information Security Management: Learning from Leading Organizations*, GAO/AIMD-98-68 (Washington, D.C.: May 1998).

The first step toward establishing effective investment management is putting in place foundational, project-level control and selection processes. These foundational processes allow the agency to identify variances in project cost, schedule, and performance expectations; to take corrective action, if appropriate; and to make informed, project-specific selection decisions.

The second major step toward effective investment management is to continually assess proposed and ongoing projects as an integrated and competing set of investment options. This portfolio management approach enables the organization to consider the relative costs, benefits, and risks of new and previously funded investments and thereby identify the mix that best meets its mission, strategies, and goals.

NARA's IT investment management policies and processes were assessed and reported on by its inspector general (IG) in April 2000. The report identified several strengths in NARA's IT investment management processes, including having an IT investment board, a defined process for selecting projects, criteria to be applied in considering whether to undertake a particular IT investment, ratings of each investment's breadth of impact, and a determination of the net benefits and risks be identified for proposed investments. However, the IG identified weakness and made 13 recommendations for strengthening NARA's IT investment management processes. NARA concurred with all recommendations. While it has to date fully addressed only 2 of the recommendations, it plans to resolve the remaining 11 issues by September 30, 2002.

While NARA's investment management process has several strengths and NARA continues to improve process weaknesses, NARA has yet to complete its efforts to establish a mature investment management capability. Lacking a fully mature investment management process increases the risk that the electronic archival system will not be implemented on time and within budget, and that crucial resources and funds for meeting the electronic records challenges will not be invested effectively and efficiently. Specifically, if NARA management's oversight of the ERA program is not based on complete information (including comparisons of the actual cost and schedule to the estimated cost and schedule, as well as identification of project risks and benefits), the risk is increased that NARA management will not be able to determine whether the ERA program is having schedule or other problems and ensure that corrective actions are taken.

NARA Is Developing an Enterprise Architecture

The importance of enterprise architecture development, implementation, and maintenance is a basic tenet of effective IT management. Used in concert with other IT management controls, an enterprise architecture can greatly increase the chances for optimal mission performance. We have found that attempting to modernize operations and systems without an enterprise architecture leads to operational and systems duplication, lack of integration, and unnecessary expense.

Over the past several years, NARA has taken action to develop an enterprise architecture. NARA has drafted a current architecture and is working on a target architecture, but this work is incomplete.[23] However, the process to develop the electronic archival system is well under way. Without an enterprise architecture to guide its development, NARA increases the risk that the planned electronic archival system will be incompatible with existing and future operations and systems, thus wasting resources and requiring that unnecessary interfaces be built to achieve integration.

NARA Is Improving Information Security, but Has Not Yet Completed Key Tasks

NARA is currently strengthening its information security, having recognized that it has numerous weaknesses. Significant security weaknesses were identified by two IG assessments (conducted in fiscal years 2000 and 2001) and a NARA-initiated vulnerability assessment of its network (performed concurrently with the IG assessments). As a result of these assessments, the Archivist of the United States declared information security a material weakness in fiscal year 2000.[24] Actions taken by the Archivist to addresses these shortcomings and respond to recommendations identified in the reports include establishing an information security program, updating and developing new security policy documents, developing contingency plans and business recovery plans, and strengthening firewalls across the network to control inbound and outbound traffic. NARA said that it would implement the IG's recommendations by June 28, 2002, and by the end of fiscal year 2002 it plans to have rectified the shortcomings that led to its information security being declared a material weakness.

[23]NARA's effort to develop an enterprise architecture includes a separate effort to develop a data architecture.

[24]*Fiscal Year 2000 Federal Managers' Financial Integrity Assurance (FMFIA) Report to the President.*

However, although NARA is making progress in strengthening its information security, two additional weaknesses could affect the ERA program. First, NARA currently lacks a program for assessing agencywide information security risks. Federal guidance requires all federal agencies to establish comprehensive information security programs based on assessing and managing risks.[25] Risk assessments provide a basis for establishing appropriate policies and selecting cost-effective techniques to implement these policies. NARA intends to develop an agencywide risk assessment capability in fiscal year 2003, but it is not clear that this will allow vulnerability assessments to be completed before ERA is developed. Without a method to identify and evaluate risks, NARA cannot be assured that it has effective mechanisms for protecting its information assets: networks, systems, and information associated with ERA. Because a compromise of security in a single poorly secured system can undermine the security of multiple systems, NARA needs to complete vulnerability assessments of all systems that will interface with ERA.

Second, because NARA lacks an enterprise architecture, it may have difficulty addressing agencywide security. Federal guidance calls for agencies to make security controls for systems consistent with and an integral part of the enterprise architecture of the agency.[26] Without an enterprise architecture that addresses security issues agencywide, NARA cannot be sure that its current or future archiving systems are adequately protected.

These weaknesses may be particularly significant for ERA, because this system presents security issues that NARA has never before addressed, according to an initial assessment report on ERA prepared by NARA's systems development and acquisition contractor.[27] The proposed distributed structure of ERA introduces the security risks associated with the Internet—threats to the integrity of data and to data accessibility. According to the Federal Bureau of Investigation, Internet systems are threatened by hackers (who may be terrorists, transnational criminals, and

[25]Chapter 35 of title 44, section 1061, subchapter II—Information Security, United States Code.

[26]Office of Management and Budget, *Incorporating and Funding Security in Information Systems Investments*, Memorandum 00-07 (Washington, D.C.: Feb. 28, 2000).

[27]Integrated Computer Engineering, Inc., *Electronic Records Archives Initial Assessment Final Report*, version 1.2 (Oct. 18, 2001).

intelligence services) using information exploitation tools such as computer viruses, worms, Trojan horses, logic bombs, and eavesdropping sniffers.[28] As Internet usage increases, the Internet has become an increasingly tempting target, and the number of reported Internet-related security incidents is growing.[29] The effect on ERA of the vulnerabilities of the Internet would have to be assessed and addressed.

Conclusions

In response to the challenges associated with managing and preserving electronic records, NARA has performed an assessment of governmentwide records management—an important first step that identified several problems, including the inadequacy of guidance on electronic records, the low priority generally given to records management, and the lack of technology tools to manage electronic records. While NARA has plans to improve its guidance and address the need for technology, it has not yet formulated a strategy to deal with the stature of records management programs across government. Further, it has no strategy for acquiring the kind of comprehensive information on records management that would be provided by systematic inspections and evaluations of federal records programs. Without such a strategy, records management will likely continue to be considered a low-priority "support" activity lacking appropriate management attention, and NARA will not acquire information needed to address problems in agency records management and guidance. Inadequacies in records management put at risk records that

[28]*Virus*: a program that "infects" computer files, usually executable programs, by inserting a copy of itself into the file. These copies are usually executed when an infected file is loaded into memory, allowing the virus to infect other files. Unlike the computer worm, a virus requires human involvement (usually unwitting) to propagate. *Worm*: an independent computer program that reproduces by copying itself from one system to another across a network. Unlike computer viruses, worms do not require human involvement to propagate. *Trojan horse*: a computer program that conceals harmful code. A Trojan horse usually masquerades as a useful program that a user would wish to execute. *Logic bomb*: in programming, a form of sabotage in which a programmer inserts code that causes the program to perform a destructive action when some triggering event occurs, such as termination of the programmer's employment. *Sniffer or packet sniffer*: a program that intercepts routed data and examines each packet in search of specified information, such as passwords.

[29]For example, the number of incidents handled by Carnegie-Mellon University's Computer Emergency Response Team (CERT) Coordination Center has increased from 1,334 in 1993 to 8,836 during the first two quarters of 2000. Similarly, the Federal Bureau of Investigation reports that its caseload of computer-intrusion–related cases is more than doubling every year.

may be valuable: records providing information on essential government functions, information that is necessary to protect government and citizen interests, and information that is significant for the historical record.

NARA's effort to acquire an advanced electronic records archive is at risk. NARA is not meeting its schedule for the ERA system, largely because of flaws in how the schedule was developed. As a result, the schedule will be compressed, leaving less time for completing essential planning tasks. In addition, NARA has not yet improved IT management capabilities that would reduce the risks inherent in its effort to acquire ERA. Without these capabilities, NARA risks spending funds to acquire a system that does not meet mission needs and requirements, effectively work with existing systems, or provide adequate security over the information it contains.

Recommendations for Executive Action

To address the low priority given to records management programs across government, we recommend that the Archivist of the United States develop a documented strategy for raising agency senior management awareness of and commitment to records management principles, functions, and programs. Further, we recommend that the Archivist develop a documented strategy for conducting systematic inspections of agency records management programs to (1) periodically assess agency progress in improving records management programs and (2) evaluate the efficacy of NARA's governmentwide guidance.

To mitigate the risks associated with the acquisition of an advanced electronic archival system, we recommend that the Archivist reassess the ERA project schedule. A revised schedule should be developed, based on estimates of the amount of work and resources required to complete each task, that allows sufficient time for NARA to

- complete essential planning tasks and
- strengthen its IT management capabilities by (1) implementing an IT investment management process, (2) developing an enterprise architecture, and (3) improving information security.

Agency Comments and Our Evaluation

In written comments on a draft of this report, which are reprinted in appendix V, the Archivist of the United States generally agreed with our recommendations but provided clarifications concerning records

management priority, inspections, and the ERA schedule. NARA also provided technical comments, which we have incorporated as appropriate.

The Archivist agreed with our recommendation that NARA develop a strategy for raising agency senior management awareness of and commitment to records management principles, functions, and programs, adding that the responsibility for oversight of records management is not NARA's alone, but is shared by the Office of Management and Budget (OMB), the General Services Administration (GSA), and the heads of federal agencies. Further, he acknowledged that more needs to be done to have a major effect on agency leadership. The Archivist, however, disagreed with our conclusion that NARA does not plan to address the low priority generally given to records management.

Our conclusion was not meant to imply that NARA does not intend to address the priority of records management. We acknowledge NARA's past efforts to raise awareness of the importance of records management and its stated plans to further address this issue. Instead, our conclusion reflects the fact that NARA's written plan to reform federal records management policies and practices—which NARA refers to as its Records Management Initiatives—does not currently address this issue. We believe that to be successful, NARA must document its plans to address the low priority of records management programs across government, including specific goals, strategies, and milestones. Such a plan is critical in ensuring concurrence on planned actions among the key players that NARA mentions, including federal agencies, GSA, and OMB; that appropriate resources are assigned; and that NARA has the means to track progress against its goals.

The Archivist also agreed with our recommendation that NARA develop a strategy for conducting systematic inspections of agency records management program, but noted that continuing its past inspection program, as cited in the report, would not succeed. NARA disagreed with our conclusion that it has no plans to address the issue of records management inspections, noting that it plans to use risk management analysis while leveraging its inspection resources. The Archivist said that this approach would include an assessment of broad categories of important records across agencies, agency-specific interventions, and the use of NARA's authority to report the results of evaluations of at-risk records to OMB and the Congress.

We are not suggesting that NARA resurrect its past inspection program, which it concluded was basically flawed. However, we also do not believe that NARA's current targeted assistance approach is an appropriate substitute for systematic inspections and evaluations of federal records programs. In regard to our conclusion, it is again based on the fact that the written strategy for the Records Management Initiatives does not address the need for systematic inspections. We acknowledge NARA's statement that it plans to use a risk-based approach to addressing this issue, but we reiterate the need for a documented plan with associated goals, strategies, and milestones.

In commenting on our recommendation that NARA reassess the ERA project schedule, the Archivist stated that such a reassessment is prudent and that NARA intends to conduct such reassessments repeatedly, both periodically from an overall program management viewpoint and on a continuing basis as part of its ERA risk management activity. The Archivist noted that NARA is currently reassessing the schedule as part of its refinement of the ERA acquisition strategy, and that this reassessment will address the issues raised in our report.

Regarding the schedule for the ERA system, the Archivist noted that while some program documentation was not completed on schedule, all items on the ERA project's "critical path" have been completed on time, and NARA expects to meet all milestones on the critical path this year. We disagree. As discussed in our report, the development of key program documents—such as the ERA vision statement and the concept of operations—were affected by delays. For example, the ERA vision statement, planned for completion on March 1, 2002, was not completed until April 18, 2002, approximately 6 weeks late. Similarly, the concept of operations, due on April 1, 2002, and which NARA documentation shows as being on the critical path, was delivered in draft form on that date and had not been finalized as of May 31. Falling behind schedule in the initial stages presents risks to successful and timely completion of the ERA project and is one of the reasons we are recommending that the agency reassess its schedule.

The Archivist also disagreed with our conclusion that if the results of the two National Academy of Sciences assessments are not fully reflected in the ERA requirements, there is added risk that the technical strategy underlying the development of the system will prove not to be optimal, and that alternatives will not have been considered. The Archivist noted that NARA should receive the first National Academy of Sciences report at a time when it expects to receive the industry's response to NARA's request

for information, and that the report will provide an unbiased, expert view of the feasibility of building a system that is inherently evolutionary, addressing the core problem of digital preservation. According to the Archivist, NARA will factor both the scientific and the industry views into its articulation of a draft request for proposals. In regard to the second National Academy of Sciences report, the Archivist noted that its primary purpose is to provide input to NARA's long-range plans for addressing the continuing evolution of information technology and electronic records, and that the report will be useful in revising the ERA research plan to address new problems and opportunities identified by the experts, and in plans for successive builds of the ERA system.

We acknowledge NARA's clarification regarding the timing and use of the two NAS studies and believe this approach should assist in developing a system that will meet mission needs. Accordingly, we have revised our recommendation to reflect this.

We are sending copies of this report to the Ranking Minority Member, Subcommittee on Government Efficiency, Financial Management and Intergovernmental Relations, House Committee on Government Reform, and to the Ranking Minority Member, Subcommittee on Treasury, Postal Service and General Government, House Committee on Appropriations. We are also sending copies to the Archivist of the United States, the Secretary of Housing and Urban Development, the Secretary of State, the Secretary of Commerce, the Secretary of Veterans Affairs, and the Administrator of NASA. This report will also be available on GAO's home page at http://www.gao.gov.

If you have any questions concerning this report, please call me at (202) 512-6240 or Mirko J. Dolak, Assistant Director, at (202) 512-6362. We can also be reached by E-mail at koontzl@gao.gov and dolakm@gao.gov, respectively. Key contributors to this report were Timothy Case, Barbara Collier, Jamey Collins, David Plocher, and Megan Savage.

Linda D. Koontz
Director, Information Management Issues

Objectives, Scope, and Methodology

Our objectives were to

- determine the status of NARA's efforts to respond to governmentwide electronic records management problems and the adequacy of its future plans and
- assess NARA's efforts to acquire an archival system for electronic records.

As part of our assessment of NARA's efforts to acquire an electronic records archiving system, we were also asked to identify alternative technologies under consideration for the long-term preservation of electronic records.

To determine the status of NARA's efforts to assess and respond to governmentwide electronic records management problems and the adequacy of its future plans, we reviewed federal legislation and NARA records management guidance, available studies, and reports; surveyed NARA's appraisal archivists working with federal agencies; reviewed records management activities and obtained the views of record managers in selected federal agencies managing large volumes of electronic records—the Departments of State, Commerce, Housing and Urban Development (HUD), and Veterans Affairs (VA), as well as NASA and the Patent and Trademark Office; and reviewed legal challenges to federal electronic recordkeeping practices, including *Public Citizen v. John Carlin* and *Scott Armstrong v. Executive Office of the President*. We also reviewed NARA's documentation of its effort to redesign its approach and guidance for the management of electronic records. As part of this effort, we investigated whether agencies are scheduling their major information systems and the related databases; to do so, we asked five major agencies—Commerce, HUD, VA, State, and NASA—what portion of their major information systems were scheduled and placed under the agency records management program. We based our assessment on the inventory of Year 2000 mission-critical systems reported by 24 major agencies to the Office of Management and Budget.[30] In addition, to determine the status of the Library of Congress' National Digital Information Infrastructure and Preservation Program and its relationship to NARA's efforts to design and

[30]Subcommittee on Government Management, Information, and Technology, House Committee on Government Reform, *Federal Government Earns a B+ on Final Y2K Report Card*, news release (Washington, D.C.: Nov. 22, 1999).

acquire advanced electronic archival system, we discussed the program's objectives and schedule with Library of Congress officials.

To assess NARA's efforts to acquire an archival system for electronic records, we reviewed agency and contractors' documentation for the electronic records archive (ERA) program, including program and project phasing; on the basis of federal requirements and information industry practice, we assessed NARA's effort to develop or enhance its information technology capabilities, including information technology investment management, enterprise architecture, and information security.

To identify alternative technologies under consideration for the long-term preservation of electronic records, we reviewed archival studies and literature, and we surveyed selected digital preservation approaches used by the information industry and selected national governments. In addition, we contacted the archives of three judgmentally selected foreign countries (Australia, Canada, and the United Kingdom) that had been identified by records management professionals as using advanced electronic records management and that we had previously reviewed.[31] We also contacted the Public Record Office of Victoria, Australia; although this archive is not at the scale of a national archive, we included it because it has employed a unique technological approach to archiving electronic records.

We performed our work from June 2001 to May 2002 in accordance with generally accepted government auditing standards.

[31]U.S. General Accounting Office, *National Archives: Preserving Electronic Records in an Era of Rapidly Changing Technology*, GAO/GGD-99-94 (Washington, D.C.: July 19, 1999) (*http://www.gao.gov/archive/1999/gg99094.pdf*).

Approaches to Archiving Electronic Records Provide Partial Solutions

The challenge of managing and preserving the vast and rapidly growing volumes of electronic records produced by modern organizations is placing pressure on archives and on the information industry to develop a cost-effective long-term preservation strategy that will free electronic records from the constraints of proprietary file formats and software and hardware dependencies. Part of this strategy will involve ways to capture and use information about the records to make them accessible, as information in card catalogs does in traditional libraries. After considerable research in this area, some agreement is being reached on the metadata (data about data) required for preserving electronic records, and some practical applications are using XML (Extensible Markup Language[32]) for creating such metadata.

However, there is no current solution to the electronic records archiving challenge, and so archival organizations now rely on a mixture of evolving approaches that generally fall short of solving the long-term preservation problem. The four most common approaches—migration, emulation, encapsulation, and conversion—are in use or under consideration by the major archives. NARA is supporting the investigation of a new approach involving records conversion (known as persistent object preservation), but this has yet to mature.

Recognizing that archival solutions may be some time off, companies in the information industry are relying on off-the-shelf technology for providing access to billions of electronic records. These commercial archives, however, concentrate on electronic records of types that are relatively uniform in comparison to those that a government archive must address.

Archiving Requires Documentation of Attributes and Relationships of Records

Archives use catalogs of various types to capture information about records, information that is critical for sharing, storing, managing, and accessing records effectively—particularly in the context of millions of records. Because such information is data containing descriptive information about other data, it is referred to as metadata. Metadata are a central element of any approach to ensure that preserved records are functional. For electronic records, the metadata needed are often more extensive than information in traditional catalogs, including information that is important for preservation.

[32]XML is a simplified subset of the Standard Generalized Markup Language (SGML) used to define portable document formats.

Metadata Provide Information Necessary to Describe Electronic Collections

The creation of accessible software- and hardware-independent electronic records requires that all materials that are placed in archives be linked to information about their structure, context, and use history. Metadata to be associated with electronic records may include information about

- the source of the record;
- how, why, and when it was created, updated, or changed;
- its intended function or purpose;
- how to open and read it;
- terms of access, and
- how it is related to other software and records used by the originating organization.

These metadata must be sufficient to support any changes made to records through various generations of hardware and software, to support the reconstruction of the decisionmaking process, to provide audit trails throughout a record's life cycle, and to capture internal documentation. Without an adequately defined metadata structure, an effective electronic archive cannot be constructed.

Numerous research projects have examined the question of defining metadata that would be sufficient to ensure digital preservation. Although archives experts note that unresolved issues remain, the work on preservation metadata is beginning to move from the research area to practice. The Public Record Office Victoria (Australia), a state archive, has published standards for the management of electronic records that includes a metadata model originally developed by the National Archives of Australia.

For incorporating metadata, the Victoria archive mandates the use of XML. XML is being actively considered by archives and researchers as a promising approach to generating metadata.

XML Enables Infrastructure-Independent Description of Electronic Records

XML is a flexible, nonproprietary set of standards for annotating ("tagging") data with semantically rich labels that permit computers to process files on the basis of their meaning.[33] Like the more familiar HTML (Hypertext Markup Language) files used on the World Wide Web, XML files can be easily transmitted via the Internet, and with appropriate software, they can be displayed by Web browsers. The difference is that HTML is used only for telling computers how to display information for a human being to view, whereas the semantically based XML tags allow computers to automatically interpret and process XML files.

XML is called extensible because it is not a fixed format. Instead, XML is actually a "metalanguage"—a language for describing other languages—which allows the design of customized markup languages for limitless different types of documents. Thus, although in the beginning stages of adoption, XML is viewed as a promising format for a wide range of applications.[34]

Several XML attributes make it attractive for archive applications. The semantic nature of XML tags makes XML suitable for recording metadata. Its extensibility would allow archives to expand their systems to accommodate evolving needs. As an open standard, it reduces the problems of proprietary software. Further, because they are basically text files, XML files can be readily interpreted by disparate computer systems. Even without the mediation of software, human beings can interpret an XML-tagged file, because XML tags are human readable (see fig. 4). This quality allows them to be preserved both on computer media and on paper (so that they would be readable both by human beings and automatically through optical character recognition).

[33]Tagging data in a standard way allows any system that recognizes the standard to readily understand and process data that conform to that standard. In tagging, a standard format is used to label each element of a data set with metadata that clarify what kind of information is being provided. Common tagging systems for electronic information—also known as markup languages—use labels set off by angled brackets to show where data elements begin and end: for example, in <label> data </label>, the second tag includes a slash to indicate that it is a closing tag.

[34]U.S. General Accounting Office, *Electronic Government: Challenges to Effective Adoption of the Extensible Markup Language*, GAO-02-327 (Washington, D.C.: Apr. 5, 2002).

Figure 4: Sample of XML Version of State Department Telegram


```
<telegram object_id="3 ">
<line count="1>AS</>
<directive>
<line count="2 ">THIS TELEGRAM MUST BE</>
<line count="3 ">CLOSELY PARAPHRASED BE-</>
<line count="4 ">FORE BEING COMMUNICATED</>
<line count="5 ">TO ANYONE.(SC)</>
</directive>
<place_of_sender>
<line count="6 ">LONDON</>
</place_of_sender>
<date_sent>
<line count="7 ">DATED
<month>DECEMBER</>
<date>6</>,
<year>1940</></>
</date_sent>
<date_received>
<line count="8 ">REC'D
<time>9:10 A.M.,</>
<date>7 TH </></>
</date_received>
<addressee>
<name>
<line count="9 ">SECRETARY OF STATE,</></>
<place>
<line count="10 ">WASHINGTON.</></>
</addressee>
<date_creation>
<line count="11 ">3984,
<date>DECEMBER 6,</>
<time>MIDNIGHT.</>
</date_creation>
<directive type="confidentiality ">
<line count="12 ">STRICTLY CONFIDENTIAL FOR THE SECRETARY AND THE UNDER </>
<line count="13 ">SECRETARY AND FOR TRUITT MARITIME COMMISSION.</>
</directive>
<date_information>
<line count="14 ">MY 3965,
<date>DECEMBER 5TH.</></>
</date_information>
<body>
<line count="15 ">THE SHIPPING SITUATION IS … <</>
</body>
</telegram>
```


Source: San Diego Supercomputer Center.

Figure 4 is an example of a text document—a World War II vintage telegram in the Franklin D. Roosevelt library—converted to XML format.[35] The XML "tags" provide the means for identifying—and retrieving—key pieces of information, such as date sent, addressee, and place of sender. If the file were viewed in an XML-compliant Web browser, the tags in the telegram would not be visible, and the telegram itself could be displayed in various ways for the convenience of the human reader. At the same time, the presence of the tags permits computer systems to perform powerful searches and exchange data.

XML is also used by the National Archives of Australia,[36] which converts files from their native formats to XML versions, while retaining a copy of the original source file. The Australian archives has also developed a metadata model, but it has not yet determined its final preservation metadata requirements.

Electronic Archives Take Combinations of Approaches to Preservation

For long-term preservation of electronic records, electronic archives must address the problems of obsolescence and aging of storage media, the dependence of electronic records on the software and hardware on which they were created, the complexity of electronic records, and the massive volumes of records created by often decentralized systems. According to one archival expert, a viable strategy for long-term preservation for electronic records would call for "a long-lived solution that does not require continual heroic effort or repeated intervention of new approaches every time formats, software, or hardware paradigms, document types, or recordkeeping practices change."[37]

Since no one solution is yet available that addresses all the problems, most archives and other institutions that preserve records use a variety of approaches, often in combination. The current approaches for dealing with the technical issues associated with long-term electronic archiving are

[35]Amarnath Gupta, *Preserving Presidential Library Websites*, San Diego Supercomputer Center, SDSC TR-2001-3 (Jan. 18, 2001).

[36]National Archives of Australia (*http://www.naa.gov.au/*).

[37]Jeff Rothenberg, *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*, Council on Library and Information Resources (January 1999) (*http://www.clir.org/pubs/reports/rothenberg/contents.html*).

- technology preservation—maintaining old technologies to allow access to old formats;
- emulation—using software running on new-technology platforms to mimic old technologies;
- migration—transferring digital materials from one hardware/software configuration to another, or from one generation of computer technology to a subsequent generation;[38]
- encapsulation—grouping together a digital object with other information necessary to provide access to that object; and
- conversion to standard formats—transforming records into objects that are relatively software and hardware independent.

The recent development of durable analog storage media (that is, media that preserve images of human-readable documents, much as microfiche does) suggests the possibility of approaches that combine those above with the use of analog rather than digital media.[39]

Technology Preservation Is a Short-Term Solution Only

Technology preservation refers to the practice of maintaining outdated equipment well after it is useful in everyday business processes. Under this approach, electronic files or records, which are saved in their native formats, continue to be accessible through the use of original hardware and software. In the short term, this is a simple and cost-effective approach, and some organizations do maintain older information systems only to be able to access their records.[40]

However, this approach is at best an interim solution to the problem of the dependence of electronic records on the software and hardware on which they were created. The solution eventually fails, because maintaining the

[38]Task Force on Archiving of Digital Information, *Preserving Digital Information* (May 1, 1996) (*http://www.rlg.org/ArchTF/*).

[39]HD-Rosetta Archival Preservation Services (*http://www.norsam.com/hdrosetta.htm*).

[40]Andrew Waugh, Ross Wilkinson, Brendan Hills, and Jon Dell'oro, *Preserving Digital Information Forever*, Commonwealth Scientific and Industrial Research Organisation (CSIRO) Mathematical and Information Sciences (undated) (*http://pigfish.vic.cmis.csiro.au/~ajw/PresDigitInfoL.pdf*).

original technology grows increasingly difficult and costly with the passage of time. Further, it does not solve the problem of aging and obsolescent storage media, which would also grow more difficult if not impossible to replace. Issues of cataloging and metadata are also not addressed by this approach. With the seemingly endless introduction of new hardware and software, the sheer number of differing formats and applications, and the cost to maintain any and all systems, technology preservation is not a feasible strategy for the long term.

Emulation Is Currently More Theoretical Than Practical for Electronic Archiving

A proposed approach to the problem of software and hardware dependence is emulation, which aims to preserve the original software environment in which records were created. Emulation software mimics the functionality of older software (generally operating systems) and hardware. Under the emulation approach, data files are stored along with copies of the creating software as well as software that emulates the hardware/operating system required to run the software.[41] This technique seeks to recreate a digital document's original functionality, look, and feel by reproducing, on current computer systems, the behavior of the older system on which the document was created. In other words, an emulation strategy means that nothing is done to the original electronic file; rather, the original environment is recreated. Since the original file remains unaltered, emulation also offers a solution to the problem of preserving the original functionality and the "look and feel" of complex digital files.

Emulation has been in practical use on computer systems for many years:

- IBM mainframes emulate previous mainframes in order to support legacy systems and allow several generations of operating system versions to be run.
- Operating system emulators allow a single computer to provide more than one operating environment (such as Macintosh and Windows).
- Emulation software allows desktop computers to run video games and legacy video gaming systems.

[41]Jeff Rothenberg, *Using Emulation to Preserve Digital Information*, Position Paper, NSF Workshop on Data Archiving & Information Preservation (Mar. 26, 1999) (*http://cecssrv1.cecs.missouri.edu/NSFWorkshop/ppaper3.html*).

However, according to one archival expert, emulation has not yet been applied to preserving archival documents in any systematic way. Although emulation could in theory be part of a solution to the problem of hardware and software independence, it is just beginning to be explored as an archival approach. Emulation is under consideration as one of various archiving approaches by the United Kingdom's Public Record Office.[42]

One problem unique to emulation is that intellectual property rights issues may be involved when either operating systems or applications are emulated.[43] Even if the software and hardware are obsolete, their copyrighted specifications are not likely to be released for the benefit of archival integrity. Further, the use of an emulated operating system or application introduces outmoded programs into a modern environment, requiring users to understand how to use them; in other words, using the old software may require expert knowledge of the outdated systems—knowledge that is likely to disappear.

Other problems with emulation include the increasing possibility that software failures will occur as the old systems continue to age and the pool of expertise concerning them shrinks. Emulation assumes that the emulated software will continue to run without maintenance. As the year 2000 date conversion problem showed, this is not a safe assumption, as it is possible that software may contain bugs that may eventually cause catastrophic loss of information.[44] Further, an emulation approach depends on several components working together (the emulation software, the original application, and the data); as the number of components increases, so does the risk of failure.

Migration of Both Media and File Formats May Preserve Records

Migration refers to the periodic transfer of digital materials from one format configuration to another, or from one generation of computer technology to a subsequent generation. In the context of archiving, migration can refer both to the media on which information resides (conversion from older to newer media or forms of media) and to the

[42]The Public Record Office is the national archive of England, Wales, and the United Kingdom *(http://www.pro.gov.uk/)*.

[43]Jeff Rothenberg, *Using Emulation to Preserve Digital Documents*, Rand-Europe, Koninklijke Bibliotheek (The Hague: July 2000).

[44]See footnote 40.

formats in which it is encoded (conversion from one file format or system to another).

The first type of migration, media migration, has been so far unavoidable: it is the standard approach to the problem of media obsolescence and aging. In media migration, records are moved from older storage media to newer media, either to avoid the obsolescence or decay of an older medium or to upgrade to a more advanced medium (often to increase storage capacities while reducing cost). However, media migration alone does not ensure that the electronic records transferred to the new media continue to be accessible, especially if their format is obsolete. As new storage technologies evolve—including extreme-longevity analog media such as the High Density Rosetta disk discussed later in this appendix—the migration process may become less frequent and more efficient.

The second type of migration, format migration, is a process of preservation by conversion: specifically, format migration is defined as rearranging the original sequence of structural and data elements of a file to conform to another configuration. Such migration occurs whenever older systems and formats are displaced by newer, often more advanced systems and formats. Many organizations have, for example, converted old database systems to newer systems, and in the process they have converted the formats of the records they contain.

The major difficulty with format migration is the risk of altering records during conversion from the source to the target format. For conversions to be successful, those performing the transition must have knowledge of the original application and data formats,[45] and the more complex the file structure, the more important this knowledge is. Whether the application is commercial or generated in house, over time this knowledge may be lost and with it the ability to perform a successful migration. For such reasons, migration has been described as cost effective only for certain types of records that remain in operational use.[46] For records in use, problems with imperfect conversion are more likely to be discovered by users, and organizational resources are more likely to be devoted to ensuring that these are resolved or mitigated.

[45]See footnote 40.

[46]See footnote 40.

Further, although format migration has occurred in many contexts in the past, it has not been extensively used in archiving. Most electronic archives are relatively new, so they are dealing with records in current formats created by systems that are still operational. Thus, they have not yet experienced the need to incorporate format migration into their processes. Rather, they treat migration as a future option for dealing with preserving the types of records that they are currently storing.

As a strategy for the long-term preservation of electronic records, relying on format migration is risky. Migration as a preservation strategy would have to be a continuous process, with conversions occurring whenever a new format needed to be introduced. With each format conversion, the possibility of loss would be increased, and the more complex the record, the more the possibility of loss. Thus, migration is at best an imperfect solution as it can potentially lead to the loss of record integrity.

Migration was selected by the United Kingdom's Public Record Office as its current archival approach. In addition to migration, the Public Records Office is also considering using emulators and viewers to access archived files in their native formats.

Encapsulation Preserves Both Records and Information about Records

Encapsulation is the combining of several elements to create a new single entity; in the context of archiving, the elements would be the records themselves, metadata identifying and describing the records, and possibly other elements (such as viewers enabling the records to be read).[47]

Unlike migration, encapsulation does not necessarily involve a change in the original file format. If the format is unchanged, encapsulation would avoid the problem of loss of integrity that migration entails. Leaving records in their native formats would leave open the possibility of processing the objects with the original software, and it would also permit subsequent transformation of the encapsulated records using methods that were not available when the records were originally placed into the archives.[48]

[47]Encapsulation, Preserving Access to Digital Information (PADI) (*http://www.nla.gov.au/padi/topics/20.html*).

[48]Ken Thibodeau, "Building the Archives of the Future: Advances in Preserving Electronic Records at the National Archives and Records Administration," *D-Lib Magazine* (February 2001) (*http://www.dlib.org/dlib/february01/thibodeau/02thibodeau.html*).

Encapsulation is currently being used by the Victoria Public Records Office in Australia.[49] The Victoria archive uses XML to encapsulate records along with standardized metadata describing each record in a Victorian Electronic Record Strategy (VERS) format.[50] The VERS format mandates the use of XML to describe and encapsulate records. However, the Victoria archive has only recently begun applying its process, and its electronics records collection is as yet small (described as "a few records"), so it is premature to judge its effectiveness for large-scale, long-term preservation.

Conversion to Standard Formats Makes Records Less Dependent on Hardware and Software

Conversion transforms records into standard text formats such as ASCII[51] or XML to increase their independence from hardware and software. This approach is currently used by the National Archives of Canada[52] and by NARA (both of which accept databases in ASCII format), as well as the National Archives of Australia,[53] which converts files from their native formats to XML, while retaining a copy of the original source file.

The Victoria archives is using a combination of conversion and encapsulation in its preservation approach, because before encapsulating selected types of documents, it is requiring their conversion (where appropriate) to Adobe Systems' Portable Document Format (PDF). PDF is a compact format that preserves all the fonts, formatting, graphics, and color of any source document, regardless of the software and hardware used to create it. Although PDF is a proprietary file format, PDF files can be shared, viewed, navigated, and printed exactly as intended by anyone with the freely distributed Adobe Acrobat Reader.

The primary shortcomings of the conversion approach are the limitations and the longevity of the selected standard.[54] For example, converting databases to ASCII format limits their usefulness: the conversion of a

[49]Public Records Office Victoria (*http://www.prov.vic.gov.au/welcome.htm*).

[50]The metadata are based on a model developed by the National Archives of Australia.

[51]The ASCII character set of 128 characters includes the familiar letters, numbers, and punctuation of the roman alphabet, along with certain other characters such as spaces, tabs, and carriage returns.

[52]National Archives of Canada (*http://www.archives.ca/*).

[53]National Archives of Australia (*http://www.naa.gov.au/*).

[54]See footnote 40.

relational database to flat ASCII database tables will eliminate the embedded information about the relationships among data elements.[55] Conversion to XML, on the other hand, may involve fewer such limitations, but it depends on the XML standard remaining in use and accessible.

NARA is investigating an advanced form of conversion combined with encapsulation known as persistent object preservation (POP). Under this approach, records are converted by XML tagging and then encapsulated with metadata. According to NARA, the persistent object transformation approach would make electronic records self-describing in a way that is independent of specific hardware and software. The architecture for POP is being developed through the National Partnership for Advanced Computational Infrastructure. The partnership is a collaboration of 46 institutions nationwide (including NARA) and 6 foreign affiliates, with the San Diego Supercomputer Center serving as the technical resource.

According to NARA, persistent object preservation would accommodate preservation of persistent but evolving collections by providing the ability to dynamically reconstruct data collections on new technology. The result would be a system that could upgrade individual technical components and migrate media while safeguarding the archived records. POP would thus not only enable the use of future, advanced technologies, it would also reduce threats to integrity and authenticity, because POP would not require changes in the preserved data. However, POP may not be sufficiently mature to be translated into system design.

Migration to Durable Analog Media May Offer Hybrid Approach

An archive that stores records digitally must use media migration as a preventive measure to avoid decay and obsolescence. However, the use of analog storage offers a possible alternative that may diminish the need for media migration. Whereas all current media now record digital information as 0's and 1's, analog storage of documents is suggested by a new product, called a High Density Rosetta, developed by Norsam Technologies (see fig. 5).

[55]A relational database allows the definition of data structures and storage and retrieval operations. In such a database the data and relations between them are organized in *tables*. A table is a collection of *records* and each record in a table contains the same *fields*. Certain fields may be designated as *key*s, which means that searches for specific values of that field will use indexing for increased speed. Interdependencies among these tables are expressed by data values.

Figure 5: The Long Now Foundation Rosetta Disk Language Archive

Source: Rolfe Horn, courtesy of the Long Now Foundation.

The nickel-plated disk, which has a life expectancy that is orders of magnitude longer than current electronic media,[56] allows the analog storage of information and images that are readable via an electron or optical microscope. Such a medium could avoid the obsolescence created by software-reliant media. The plates are physically inscribed by an ion

[56]The manufacturer claims a life expectancy of at least 1,000 years and a temperature threshold of 500° C.

beam, through a process known as ion milling.[57] This medium can store on each side of its 2-inch plate over 196,000 pages (with electron microscope retrieval) or 5,000 to 18,000 pages (with optical microscope retrieval). Using a text-based coding system such as XML would permit both coded (software readable) and image (human readable) information to be stored on this long-lived medium. The migration issue would then arise if new software were to be adopted, but the image information would persist.

The High Density Rosetta is being used by the Long Now Foundation to create an extreme-longevity archive of selected languages.[58] According to the foundation, 50 to 90 percent of the world's languages are predicted to disappear in the next century, many with little or no significant documentation. As part of the effort to secure this critical legacy of linguistic diversity, the foundation initiated the Rosetta Project,[59] an effort to develop a contemporary version of the historic Rosetta Stone. The project's goal is the development of a permanent archive of 1,000 languages. For storage of this archive, the project is using the High Density Rosetta to micro-etch text of archived languages at a scale readable by a 1,000-power optical microscope.

Information Technology Industry Relies on Off-the-Shelf Technologies to Provide Access to Electronic Collections

While government and academic institutions are searching for a permanent solution to electronic records archiving problems, the private sector, also concerned about and affected by the potential loss of electronic records, relies on existing information architectures and off-the-shelf technologies to make accessible massive volumes of electronic records dating back over two decades. These archiving achievements do not meet the rigorous requirements for permanence and authenticity that are demanded by a government archive, nor are their owners required to process, store, and access the full range of complex file formats encountered by governments. However, they do illustrate the capability to provide storage and access to large quantities of data. Two of the most notable private sector efforts are the Internet Archives and the Google archive of Usenet messages.

[57]Ion milling is an etching process in which high-energy gallium ions produced by a focused ion beam machine knock atoms from the surface and micro-engrave into any given medium.

[58]The Long Now Foundation (*http://www.longnow.org*).

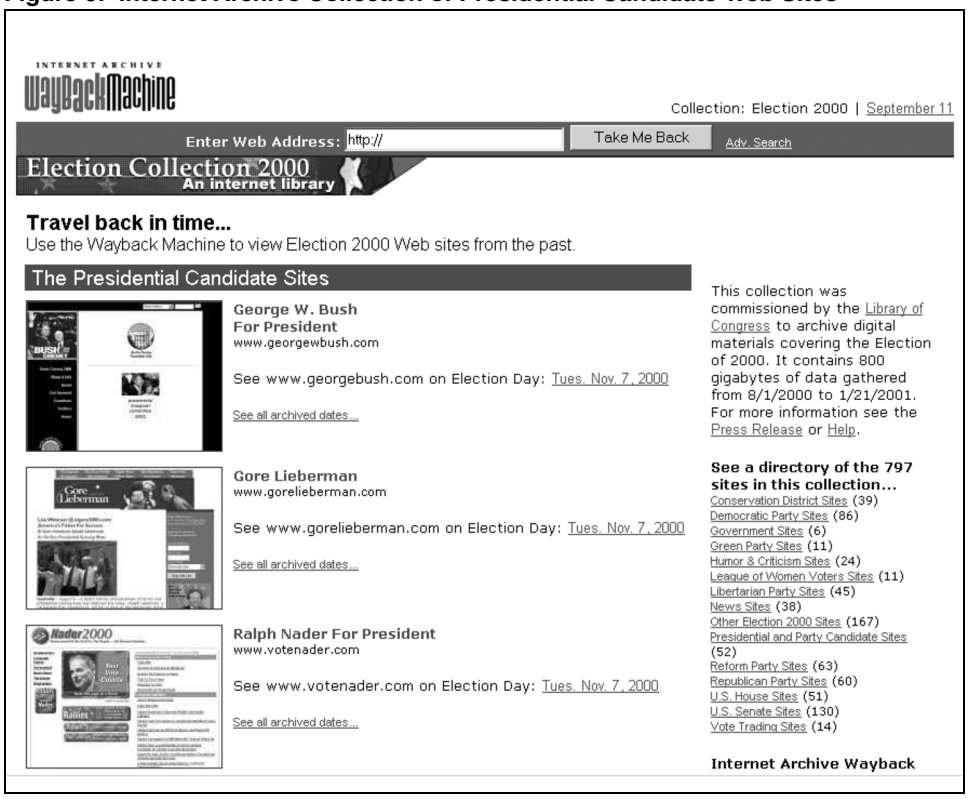
[59]The Rosetta Project (*http://www.rosettaproject.org:8080/live*).

Internet Archives

The Internet Archives has created a digital library of Internet sites and other born-digital cultural artifacts. It is attempting to archive the entire publicly available Web, offering free access to researchers, historians, scholars, and the general public. Anyone with access to the Internet can, through the Internet Archives Web site,[60] navigate the Web at any moment in time from 1996 to the present. This collection of Web pages contains over 100 terabytes, or 10 billion Web pages, and it is currently growing at a rate of 12 terabytes per month. The stored and accessible 100 terabytes is larger than the amount of data contained in the world's largest libraries, including the Library of Congress, making it the largest known database in existence. Without the efforts of the Internet Archives, these 10 billion Web pages might have been lost. As it is, they provide a record of the origins and evolution of the Internet, as well as a reflection of societal interests and opinions at different moments in time. This is particularly true in the case of Web sites such as those of presidential candidates (see fig. 6) and of monumental events such as the September 11 attacks, both of which have prominence on the Internet Archives Web site as "Special Wayback Collections."

[60]Internet Archives (*http://www.archive.org/*).

Figure 6: Internet Archive Collection of Presidential Candidate Web Sites



Source: Internet Archives.

According to the Internet Archives, it has achieved inexpensive storage on a major scale: it uses off-the-shelf technology at a cost of about $4,000 per terabyte. As a preservation strategy, the Internet Archives currently uses media migration to avoid media obsolescence and take advantage of technological advances to reduce costs. As a safety measure, backup copies of a part of the collection are also created.

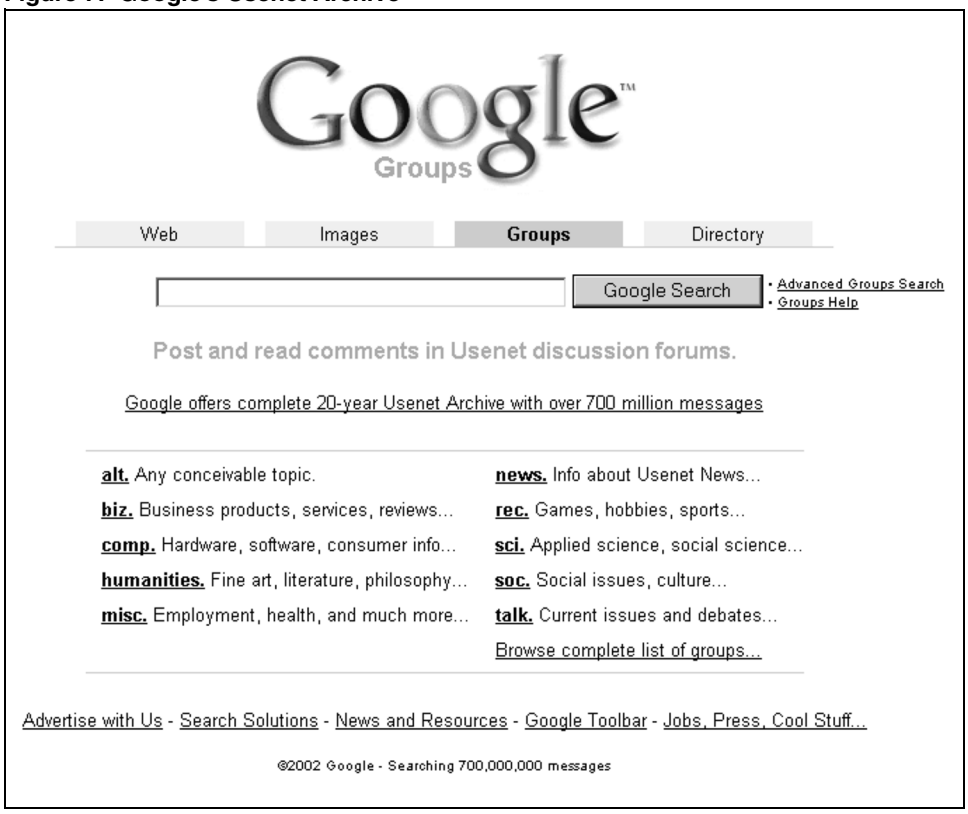
Google

Google claims to have the largest index of Web sites available on the World Wide Web and the industry's most advanced search technology. Google's Web site also contains an archive of Usenet messages that cover the past 20 years (see fig. 7).[61] Usenet is a collection of text messages that are posted on Internet electronic bulletin boards. These bulletin boards—which

[61]Google Groups (*http://www.google.com/grphp?hl=en*).

existed before E-mail, Web browsers, and the Web itself—provide avenues for communication in an open forum, allowing others to read and reply. Some notable "posts" included in Google's Usenet Archives are the first post mentioning Microsoft (1981), the first post mentioning a compact disc (1982), and the posts sent just after the September 11 attacks.

Figure 7: Google's Usenet Archive



Source: Google.

Google currently provides access to more than 700 million messages dating back to 1981, and this number is rapidly increasing. Google's collection is by far the most complete collection of Usenet articles ever assembled. Before Google's acquisition of the archive, posts without activity were usually deleted from the live discussion forums after a few days or weeks, and therefore they were not viewable or searchable by users. Some feel

that Google's Usenet archive is an irreplaceable and invaluable reference, representing "the human side of the Internet" through first-hand accounts of historical events.

NARA's Electronic Records Guidance Has Evolved

A review of the development of electronic records guidance issued by the National Archives and Records Administration (NARA) over the last several decades demonstrates the extent to which the rapid evolution of information technology has posed significant challenges for NARA in its role of providing guidance to federal agencies concerning the management of electronic records under the Federal Records Act.[62]

NARA provides guidance for electronic records management and disposition largely through two sets of guidance:

- the electronic records management regulation, which provides general responsibilities for agency management of electronic records;[63] and
- the general record schedules, which provide disposal authorization for specific categories of temporary records common to most agencies.[64]

The history of these two sets of guidance reflects the evolution of NARA's electronic records guidance.

Electronic records management was given a formal role in 1968 when NARA, then the National Archives and Records Service (NARS) of the General Services Administration (GSA), established a unit to develop policies for selecting and preserving electronic records. This Data Archives Staff undertook to develop three sets of guidance: (1) inventory guidance—forms for inventorying magnetic tape files; (2) environmental guidance—recommendations for proper handling and storage of magnetic tape; and (3) GRS 20—a general records schedule for computerized records.

Of that guidance, GRS 20 emerged as NARA's first significant electronic records guidance. It was intended to cover electronic records created by mainframe applications in the then-dominant agency data processing operations. The major purpose was to address the efficient disposition of those electronic records, including destruction of unneeded temporary records and transfer to NARS (NARA) of permanent records.

[62]44 U.S.C. chapters 21, 29, 31, and 33.

[63]36 CFR Part 1234. This rule is supplemented by NARA's *Records Management Handbook* and periodic guidance on specific issues, e.g., NARA Bulletin No. 2000-02 (Dec. 27, 1999).

[64]GRS 20 (August 1995).

The 1972 GRS 20, entitled *Data Automation Program Records*, stated, "This schedule covers machine readable records, related documentation required for their servicing, and files related to the automatic data processing (ADP) procurement, operations, and management functions." GRS 20 divided these records into categories that "correspond roughly to the typical organizational and functional structure found in most ADP installations and their parent organizations."[65]

According to recent NARA summaries, the 1972 GRS 20 was meant "to provide disposal authority for specific categories of temporary records associated with mainframe applications. Excluded from its coverage, and all subsequent revisions, were the types of records generated by large data systems that might have archival value."[66] The clear meaning of the 1972 GRS 20, however, was that it was not meant merely to identify and provide for efficient disposal of "ancillary materials common to most data processing operations."[67] Quite the contrary, the guidance identified a range of records that should be scheduled through filing of a Standard Form 115. These ranged from various temporary records to potentially permanent records, such as master data files.

GRS 20 was revised in 1977.[68] While the 1977 revision restructured the 1972 electronic records categories, it retained the earlier purpose of providing disposition instructions for virtually all records associated with data processing operations—temporary and permanent, program and administrative.[69]

In 1983, GSA issued Bulletin FPMR B-127, *Archives and Records*, which provided guidance on records created or maintained "using personal computers and electronic information storage or transmission equipment

[65]GRS 20, *Data Automation Program Records*, FPMR 101-11.4 (Apr. 28, 1972).

[66]GRS 20 (August 1995).

[67]History of General Records Schedule 20, Electronic Records (*www.nara.gov/records/grs20/20hist.html*).

[68]GRS 20, *Machine-Readable Records*, FPMR 101-11.4 (Feb. 16, 1977).

[69]Administrative records are those created in the performance of common facilitative functions that support an agency's mission activities, but do not directly document the performance of mission functions. Administrative records are temporary. Program records are those created in the performance of the unique functions that stem from an agency's mission. Program records may be temporary or permanent; they must be scheduled.

(electronic filing and electronic mail)."[70] According to the bulletin, "The proliferation of personal computers in many Federal agencies and the implementation of sophisticated electronic filing and/or mail systems has created a need for adaptation of traditional records management techniques for the control and disposal of records and information." The bulletin then reiterated that the disposition of all records regardless of physical form is controlled by the Federal Records Act and instructed agencies to ensure "that appropriate internal controls are instituted to prevent the loss or alienation of official records created or acquired in electronic form."

Two pieces of similar guidance followed in 1985. First, NARA issued Bulletin 85-2 to provide general guidance "on how to manage records created, stored, or transmitted using personal computers or other electronic office equipment including word processors."[71] This bulletin again rooted electronic records management in the fundamental requirements of the Federal Records Act: "The creation, maintenance, and disposition of all official records regardless of physical form is controlled by the provisions of [the Federal Records Act and implementing regulations]."

Two weeks after issuing Bulletin 85-2, NARA issued an ADP Records Management regulation.[72] This rule was the first version of the regulation still found at 36 CFR 1234. The rule consolidated guidance consistent with the goals of the 1968 Data Archives Staff, requiring each agency (in very summary terms) to

- establish a program for the management of ADP records, including classifying, preserving, and scheduling machine-readable records; and
- ensure proper care, handling, and storage of magnetic computer tapes and disk packs.

The next major step in the evolution of NARA's electronic records guidance occurred in the 1988 revision of two general records schedules: GRS 20, now entitled *Electronic Records*, and GRS 23, *Records Common to Most*

[70]GSA Bulletin FPMR B-127 (June 17, 1983).

[71]NARA Bulletin No. 85-2 (June 18, 1985).

[72]36 CFR 1234, 50 FR 26939 (June 28, 1985).

Offices within Agencies.[73] The revisions significantly modified the scope of both general records schedules and, for the first time, provided disposal authority for personal computer records in GRS 23.

With regard to GRS 20, the 1988 revision altered its scope, stating, "This schedule applies to disposable electronic records routinely stored on magnetic media by Federal agencies in central data processing facilities." As opposed to the broad purpose of the 1972 and 1977 versions, which had been to provide disposition guidance for all electronic records associated with data processing operations, the 1988 GRS 20 discussed only disposable records. All references to scheduling records were removed. This change was not limited, however, to GRS 20. It reflected a NARA decision that all general records schedules should pertain only to disposable records. The intent was to rely on other guidance to provide instructions about scheduling and disposition of permanent records, such as the regulation at 36 CFR 1234 and the *Appraisal Guidelines for Permanent Records*, now published as an appendix in NARA's *Disposition of Federal Records* handbook.

The second major change in 1988 was the GRS 23 treatment of records generated on personal computers. Like the 1988 GRS 20, the 1988 GRS 23 was explicitly limited to disposable records: "The records covered by this schedule relate to routine internal administrative and housekeeping activities." GRS 23 provided disposal authority for temporary administrative records generated by end-user applications on stand-alone or networked computers. This included word processing files, spreadsheets, and administrative databases. In addition to authorizing the destruction of administrative or housekeeping records when no longer needed, the 1988 GRS 23 authorized the deletion of electronic versions of records created after they were printed to hard copy, unless the records were maintained only in electronic form. If the electronic record was maintained only in electronic form, it could be deleted only after the expiration of the retention period authorized for the hard copy by the GRS or a NARA-approved SF 115. As NARA subsequently stated, its acceptance of paper recordkeeping for electronic records was based on the assessment that even with the growing use of computers, "agencies continued to maintain records produced with office automation applications in organized paper files, especially since end-user applications were not

[73]GRS 20 (June 1988); GRS 23, *Records Common to Most Offices within Agencies* (June 1988).

designed to classify, index, and maintain documents for their authorized retention period …" Thus, the revised GRS authorized deletion of word processing and E-mail records after they had been copied to paper or microform.[74]

The 1988 revisions to GRS 20 and 23 were followed by the 1990 revision to NARA's electronic records management regulation.[75] This revision continued the purposes of the 1985 bulletins, but provided more detailed mandates for "procedures to manage electronic records, to provide for the selection and maintenance of electronic storage media, and to follow the legal requirements for the disposition of such records." Agency requirements under this still valid and largely unchanged regulation include the following:

- develop and implement an agencywide electronic records management program;
- establish procedures for addressing records management requirements before approving new electronic records systems or enhancements to existing systems; and
- specify the location, manner, and media in which electronic records will be maintained to meet operational and archival requirements, and maintain inventories of electronic records systems.

While NARA endeavored to create a comprehensive electronic records management scheme through the combination of affirmative guidance, such as the 1990 regulation, and the revised general records schedules, the GRS 20 principle that paper printouts could substitute for electronic records became the focus of controversy through a lawsuit challenging the 1989 destruction of White House E-mail tapes. The case, *Armstrong v. Executive Office of the President*, spanned several years and involved multiple issues and court rulings. In a 1993 ruling in that case, the U.S. Court of Appeals ruled that paper printouts of E-mail messages were not adequate substitutes for electronic versions stored on computer tapes because they "may omit fundamental pieces of information which are an integral part of the original electronic records, such as the identity of the

[74]GRS 20 (August 1995).

[75]*Electronic Records Management*, 55 FR 19216 (May 8, 1990).

sender and/or recipient and the time of receipt."[76] Thus, the court rejected the government's argument that "electronic records are merely 'extra copies' of the paper versions," and concluded that "since there are often fundamental and meaningful differences in content between the paper and electronic versions of these documents, the electronic versions do not lose their status as records and must be managed and preserved in accordance with the FRA."

Largely in response to the court's findings, NARA revised GRS 20 in 1995.[77] First, as an organizational matter, it moved the electronic records instructions from GRS 23 into GRS 20 in order to have a single general schedule for all disposable electronic records. This resulted in combining instructions for the broad format categories of word processing files, electronic mail records, and electronic spreadsheets with those for specific functional categories of administrative records, such as backup files, finding aids, and systems operations records. Second, as a substantive matter, NARA now instructed agencies to "identify records created using office automation and to maintain them in a recordkeeping system that preserves their content, structure, and context for their required period." According to the GRS,

> "Only after the records have been properly preserved in a recordkeeping system will agencies be authorized by GRS 20 to delete the versions on the electronic mail and word processing systems. As indicated, most agencies have no viable alternative at the present time but to use their current paper files as their recordkeeping system. As the technology progresses, however, agencies will be able to consider converting to electronic recordkeeping systems for their records."

Thus, NARA stated in the 1995 GRS, "Program records that have been transferred to the recordkeeping system will not be affected by GRS 20." However, because NARA accepted the use of paper files as appropriate recordkeeping systems for electronic records, this logic permitted the disposal of electronic versions of records that required retention or permanent preservation. Accordingly, while GRS 20 did not authorize the destruction of program records, it did permit the destruction of electronic copies of those records.

[76] *Armstrong v. Executive Office of the President*, 1 F. 3d 1274 (Aug. 13, 1993).

[77] GRS 20 (August 1995).

In 1997, a Federal District court, in *Public Citizen v. John Carlin*, overturned the 1995 GRS 20, finding that it did not go far enough to direct agencies to protect electronic records.[78] The court ruled that NARA should not have treated electronic records as disposable simply because they could be copied into another form:

> "[The] differences between electronic and paper records illustrate the fact that the administrative, legal, research, and historical value of electronic records is not always fully captured—indeed, is usually not captured—by paper or microfiche copies. Electronic records therefore do not become valueless duplicates or lose their character as 'program records' once they have been printed on paper; rather, they retain features unique to their medium."

The court also found that NARA failed to perform its statutory duty to evaluate the value of records for disposal: "By categorically determining that electronic records possess no administrative, legal, research or historical value beyond paper print-outs of the same document or record, the Archivist has absolved both himself and the federal agencies he is supposed to oversee of their statutory duties to evaluate specific electronic records as to their value."

In response to the district court ruling, NARA established an Electronic Records Work Group to review the 1995 GRS 20 and make recommendations for revisions. It also issued a number of pieces of guidance to reflect the District Court's ruling.[79]

On August 6, 1999, the U.S. Court of Appeals for the D.C. Circuit upheld NARA's GRS 20, reversing the District Court decision that had overturned the 1995 GRS 20.[80] The Court of Appeals rejected the lower court's reasoning that NARA had authorized destruction of all types of word processing and E-mail records without regard to content: "GRS 20 does not authorize disposal of electronic records per se; rather, such records may be discarded only after they have been copied into an agency recordkeeping system."

[78] *Public Citizen v. John Carlin*, 2 F. Supp. 2d 1 (D.D.C. 1997).

[79] See, e.g., NARA, *Disposition of Electronic Records*, Bulletin 98-02 (Mar. 10, 1998); U.S. General Accounting Office, *National Archives: Preserving Electronic Records in an Era of Rapidly Changing Technology*, GAO/GGD-99-94 (Washington, D.C.: July 1999).

[80] *Public Citizen v. John Carlin*, 184 F.3d 900 (D.C. Cir. 1999).

The court acknowledged that an electronic recordkeeping system would be superior to a paper recordkeeping system, but it also agreed with NARA that agencies should be free "to maintain their recordkeeping systems in the form most appropriate to the business of the agency." Thus the court said,

> "We agree with Public Citizen that electronic recordkeeping has advantages over paper recordkeeping, but our duty as a reviewing court is to ask only whether the Archivist's policy choice is arbitrary or capricious; manifestly it is not. All agencies by now, we presume, use personal computers to generate electronic mail and word processing documents, but not all have taken the next step of establishing electronic recordkeeping systems in which to preserve those records. It may well be time for them do so, but that is a question for the Congress or the Executive, not the Judiciary, to decide."

Finally, the court found that the 1995 GRS 20 met the *Armstrong* test of requiring that electronic records be stored in a manner that captures all relevant transmission data.

As a result of the Court of Appeals ruling, NARA instructed agencies to again use the 1995 GRS 20 to dispose of temporary electronic records after recordkeeping copies were filed in electronic, paper, or microform recordkeeping systems.[81] NARA did say, however,

> "We believe there may be better alternatives to GRS 20 for disposition authority for electronic copies of program records and expect to develop those alternatives as part of a comprehensive review of the policies and procedures for scheduling and appraisal of records in all formats. The Court decision provides the Government time to include electronic copies in this overall review. Our review may result in significant changes in the way that agencies schedule their records in the future. When we have completed this review, we will promulgate new guidance."

On October 10, 2001, NARA published a notice seeking public comment on a petition for rulemaking filed by the Public Citizen Litigation Group (a plaintiff in both *Public Citizen v. John Carlin* and *Armstrong v. Executive Office the President*) requesting NARA to revise its electronic records management regulations.[82] In this notice, NARA stated that it was currently "evaluating alternatives to GRS 20 for disposition authority as part of a comprehensive review of the policies and procedures for scheduling and

[81]NARA Bulletin 2002-2 (Dec. 27, 1999).

[82]66 FR 51739 (Oct. 10, 2001).

appraisal of records in all formats." As of May 2002, this review was ongoing.

Agencies Are Managing Large Volumes of Important Electronic Records

Agencies are facing the complex challenge of managing electronic records and in some cases maintaining these records on a long-term basis. For example, because of their particular missions, NASA, the Patent and Trademark Office, Veterans Affairs (VA), and the State Department must each electronically manage millions of electronic records, either long-term or permanently. In some instances, the volumes of electronic records that these agencies manage are far larger than the volumes of permanent electronic records that NARA currently archives. The experiences of these agencies highlight electronic records management and the gaps in existing guidance.

National Aeronautics and Space Administration

NASA is committed to the long-term preservation of massive volumes of electronic space science data and images of our solar system. The observational data sets from NASA missions record the continually changing aspects of our Earth and represent an asset that must be retained in a findable, accessible, and usable state. The agency proposed to permanently maintain these data within the agency in order to support future science usage. Presently, NASA's National Space Science Data Center archives over 20 terabytes of digital space science data from past and present NASA missions, of which 3 terabytes are currently electronically accessible. In addition, the Hubble Space Telescope has created a data archive of over 7 terabytes of images of our solar system, and continues to archive an additional 3 to 5 gigabytes every day. Archiving and ensuring data integrity of all these electronic records require periodic data renewal cycles, involving migration from old to new media, resource-intensive data reorganization and reformatting, or even recreation of related software.

Because these records are of permanent value and NARA has no means to archive them in any useful way, NASA retains custody of them. They accordingly fall into an undefined category: they are permanent records that NARA cannot archive. The current arrangement by which they are maintained is not covered by NARA guidance. Nor is NASA's archiving approach covered by this guidance, which does not cover migration and archival formats (other than flat ASCII files on tape), management of digital images, or maintenance of electronic records in databases for extended periods of time.

U.S. Patent and Trademark Office

The Patent and Trademark Office manages and indefinitely preserves millions of digitized patents and trademarks. Patent examiners must have access to a complete collection of the history of U.S. patents in order to research prior art before approving new patents. Recently, the office replaced the examiners' collection of paper patents with EAST (Examiners Automated Search Tool) and WEST (Web Examiner Search Tool), which are complete electronic patent collections containing the full text of over 2.5 million U.S. patents and full images of over 6.5 million U.S. patents and over 14.5 million foreign patents. In addition, the Patent and Trademark Office has digitized the text and images of over 2.7 million trademark applications and registration. The Patent and Trademark Office has been using XML[83] to develop and implement systems to support the filing, examination, publication, and archival storage of intellectual property documents in electronic format.

The Patent and Trademark Office's digitization program has highlighted an issue that is not adequately addressed by NARA guidance: that is, when a record exists in many versions (electronic, paper, microform, etc.), which should be considered primary? Many of the patent files that have been digitized were originally paper files, and it has been argued that destroying the original paper versions after digitization has led to or risked loss of important information.[84] Just as converting an electronic original to paper may lead to information loss, so may the reverse. NARA guidance does not address this issue, leaving agencies at risk of losing information.

[83]Extensible Markup Language (XML) is discussed further in appendix II.

[84]The potential problem of information lost during the conversion from paper to electronic patents was identified in a recent Congressional hearing: when searching electronic patent databases for prior art, patent searchers miss relevant patents. As noted in testimony by an association representing patent researchers, this is due to a unique problem related to how an invention is described: "in many, if not most, cases the invention is never fully described 'in the words.' The patent law requires only that the specification, including the drawings, together be understandable and enabling to one of ordinary skill in the art to make and use the invention. 'The words,' in many if not most cases, merely 'flesh out' what is shown in the drawings and do not replicate 'in words' what is in the drawings, but are ancillary thereto. Thus, in a patent database electronic search one is often presented the additional problem of 'searching' for 'words' which were never there to begin with." —Testimony of James F. Cottone, President, National Intellectual Property Researchers Association, Oversight Hearing on the U.S. PTO of the Subcommittee on Courts and Intellectual Property of the House Judiciary Committee (Thursday, Mar. 9, 2000) (*http://www.house.gov/judiciary/cottone.htm*).

Department of Veterans Affairs

VA must manage and preserve, for 75 years, millions of electronic medical and benefit records. An integral part of VA's enrollment process for each veteran applying for health benefits is the use of several Veterans Health Information Systems and Technology Architecture (VISTA) databases to enter and verify veteran eligibility information. This information must be maintained in the system and accessible for the life of the veteran in order to document entitlement to health care benefits, which VA has determined to be a maximum period of 75 years. One enrollment database alone contains information for 9 million veterans.

VA patient enrollment records present another instance of the confusion regarding scheduling requirements for electronic records and for records in multiple versions. Although VA is working toward a completely electronic process, enrollment records are initiated on paper because of current legal requirements for ink signatures. In general, however, VA does not schedule electronic records when it has scheduled the paper version. It is NARA policy, however, that electronic records must also be scheduled. According to VA, another key challenge that it faces is ensuring the validity and authenticity of electronic records, and it would like to see adequate guidance and standards about electronic signatures from NARA so that all government agencies are using the same approach.

Department of State

State electronically preserves over 25 million diplomatic cables and more than 400,000 digital images of correspondence of the Secretary of State. The State Archiving System (SAS) is a repository for over 25 million cables, from 1973 to the present, documenting the conduct of U.S. foreign policy. The cables are managed electronically for 25 years before they are due to be transferred to NARA. However, if the cable records in SAS had been transferred to NARA for archiving, they would no longer have been accessible to users.

NARA has responded to the State Department's archiving and access needs by developing a new system (Access to Archival Databases), which is expected to be available in the summer of 2002. This system will allow NARA to provide on-line access to archived State Department cables. When the system is available, the cable records will be transferred to NARA for archiving.

In addition, the Secretariat Tracking and Retrieval System (STARS) tracks approximately 440,000 digital images of foreign policy memoranda and

correspondence of the Secretary of State from 1986 to the present. Both STARS and SAS must not only preserve the records, but also maintain reliable and rapid access to the image data. As technologies change, preserving and providing access to the records present complex electronic records management challenges.

The State Department's records management office has sole responsibility for maintaining SAS, and it has had to proceed with the long-term management and preservation of the system records—periodically updating and migrating all the images to reflect new technologies—without guidance from NARA. NARA guidance does not address updating or migration of file formats.

Comments from the National Archives and Records Administration

National Archives at College Park
8601 Adelphi Road College Park, Maryland 20740-6001

May 30, 2002

Joel C. Willemssen
Managing Director
Information Technology Team
General Accounting Office
441 G Street NW
Washington, DC 20548

Dear Mr. Willemssen:

Thank you for the opportunity to review and comment on the draft report on challenges in managing and preserving electronic records. The report recognizes the enormous challenges the Federal Government faces in managing and preserving electronic records and many of the actions the National Archives and Records Administration has taken to meet those challenges. Nevertheless, we agree that more must be done, and we support the report's recommendations. We would like to clarify several points in the report, however, and have suggested some technical corrections in an attachment.

Records Management

The report recommends that we develop a strategy for raising agency senior management awareness of and commitment to records management principles, functions, and programs. We certainly agree with this recommendation, and are active on a number of fronts to raise senior management awareness of and commitment to records management in Federal agencies. Such activities include:

- The Deputy Archivist of the United States and I along with senior NARA program officials have held a series of meetings with agency heads on the importance of records management and specific agency records issues.
- The Deputy Archivist and I speak at agency conferences to emphasize the importance of records management. For example, in April I addressed senior leadership at the Treasury Department's records management conference.
- NARA has developed tools (e.g., PowerPoint presentations) that agencies can use to do their own management briefings. These have been popular with agency records management officers.
- NARA developed specific guidance for senior agency management, *Documenting Your Public Service*, which was distributed to all senior officials at the start of the Administration.
- NARA works with the Office of Management and Budget (OMB) to include a records management emphasis or implications in new guidance to agencies such as the OMB Circular A-130 revision, annual OMB Circular A-11 revisions, and the Government Paperwork Elimination Act.

- NARA is the managing partner for the Electronic Records Management E-Government Initiative, which involves a coalition of Federal agencies working together to develop policies and tools to improve electronic records management.

Despite all of these activities, however, we agree that more needs to be done to have a major effect on agency leadership. Effective records management must be a partnership, a concept reflected in the U.S. Code. As laid out in 44 U.S.C. chapters 29 and 35, the responsibility for oversight of records management is shared by NARA, the Office of Management and Budget (OMB), and the General Services Administration. Of equal importance, the head of each Federal agency is charged with the responsibility to make and preserve records (44 U.S.C. 3101) and "establish and maintain an **active, continuing"** records management program (44 U.S.C. 3102, emphasis added).

Federal agency management will not take an interest in records management unless it can help them meet their business needs. The recent *Report on Current Recordkeeping Practices within the Federal Government*, which we commissioned, found that when agencies have a strong business need for good recordkeeping, such as for legal or operational needs, their recordkeeping practices are better. As part of the strategy we are developing for our Records Management Initiatives, we plan to create incentives for agencies to work with us in a "virtuous cycle" where our records management program adds value to the agencies' business processes and as a result records are kept long enough to protect rights, ensure accountability, and document the national experience. We disagree with the GAO report's conclusion that NARA does not plan to address the low priority generally given to records management. Our whole approach is predicated on the assumption that records and records management are integral aspects of agencies' business architectures. In addition, our plans recognize that we need to show more leadership with Federal agencies and the Congress on records management issues.

The GAO report also recommends that we develop a strategy for conducting systematic inspections of agency records management programs. While we agree with the thrust of the recommendation, continuing our past inspection program as cited in the report will not succeed. When NARA undertook the Records Management Initiatives to rethink completely how we do records management in the Federal Government, we put our evaluation program on hold, pending changes to the program, because it was clear we needed to do things differently. For example:

- The evaluation program could at best conduct 3 agency evaluations a year, meaning it would take at least 60 years to cover the major agencies of the Federal Government.
- Each evaluation was extremely labor intensive involving staff from multiple units (headquarters and field) up to a year.
- Because the evaluations were of records management programs, responsibility for responding to them fell to records management staff, not the program staff who actually managed the records. Where records management is not closely identified with the business process, it will not be effective.
- Many of the recommendations were broad, could take years to implement, and could be extremely resource intensive. Frequently agencies lost interest in the issues,

especially if there was a change in records officer before the action plan was completed.

- Program effectiveness was very uneven. A few agencies (e.g., IRS) completed their actions plans in a timely fashion. Yet even though we have not started a new evaluation in several years, there are a number of agencies that have not completed their action plans.

In addition, the *Report on Current Recordkeeping Practices within the Federal Government* concluded that while NARA should work with individual agencies, "given the availability of resources . . . NARA may wish to carefully consider which agencies should be selected for assessment. The situational factors at some agencies may limit the likelihood that specific, *or any*, intervention options can improve RM."[1] While heeding this caution, we plan to make evaluations, surveys, and inspections part of the strategy we are developing to assess how well records are managed in agencies as a result of our Records Management Initiatives. We disagree with the GAO report's conclusion that NARA has no plans to address the issue of records management inspections. Using risk management analysis while leveraging our inspection resources, our approach will include looking systematically at broad categories of important records across agencies as well as undertaking agency-specific interventions. We also plan to make more use, as necessary, of our authority to report the results of evaluations to OMB and the Congress, especially on issues related to at-risk records.

Electronic Records Archives

The GAO report recommends that we reassess the Electronic Records Archives (ERA) project schedule. We believe that such reassessment is prudent and intend to conduct such reassessments repeatedly, both periodically from an overall program management viewpoint and on a continuing basis as part of our ERA risk management activity. We are currently reassessing the schedule as part of our refinement of the ERA acquisition strategy. This reassessment will address the issues that the report raised, and we will report the results of our reassessment to both GAO and our Congressional committees.

We would, however, like to clarify two points of special importance related to the ERA project schedule. First, the report states that "NARA is not meeting its schedule for the ERA system...." Although some program documentation deliverables have not been completed on schedule, all items on the "critical path" have been completed on time, and we expect to meet all milestones on the critical path this year.

Second, the report suggests aligning the project schedule for deliverables from the study NARA is sponsoring by the Computer Science and Technology Board of the National Academy of Sciences (NAS) with the system acquisition schedule. The NAS study is divided into two parts, with a separate report to be issued at the end of each. The division of this study into two parts reflects the fact that the preservation of electronic records is an open-ended, evolving challenge for which there can be no one-time solution. NARA

[1] SRA International, Inc., *Report on Current Recordkeeping Practices within the Federal Government*, December 10, 2001, p. 32. Emphasis in original.

has both near-term and long-term needs to preserve electronic records. The critical near-term need is to stem and prevent the loss of valuable electronic records of the Federal Government by developing the capability to preserve and provide access to them. The long-term need must incorporate the expectation of continuing and often unpredictable change into NARA's long-range planning.

The first part of the NAS study will assess the technical recommendations NARA has received from research we cosponsored, with the National Science Foundation, in the National Partnership for Advanced Computational Infrastructure (NPACI). It will focus on the information management architecture proposed by NPACI for persistent archives of digital information. The most basic requirement for any digital archives is for a solution that is sustainable in the face of continuing and ultimately unpredictable change in information technology. Otherwise, the solution itself will come to embody, in a relatively short time, the very problems it purports to solve.

Thus, as the GAO report correctly notes, the infrastructure independence of the basic architecture is a "major dependency" for the acquisition of the ERA system. The NAS report on this topic "is expected to address the adequacy and soundness of the architecture as a whole and its major components." But the GAO report asserts, "NARA's planning has left little opportunity for the assessment results to be reflected in the ERA design without disrupting the acquisition process and increasing the risk to the ERA schedule." We disagree with this conclusion, which assumes NARA will make design decisions about the ERA system prior to receipt of the NAS report. In fact, NARA will not even begin to address design until well after the NAS report is received. The delivery of the first report, projected for January 2003, is timed to fit into the schedule for development of the ERA system.

NARA should receive the first NAS report in the same time frame that we receive industry's responses to our planned request for information. Those two information sources will be complementary. The NAS report will provide an unbiased, expert view of the feasibility of building a system that is inherently evolutionary, addressing the core problem of digital preservation. The industry responses will indicate how close the market is to supporting the development of a system that is independent of infrastructure. NARA will factor both the scientific and the industry views into its articulation of a draft request for proposals.

The GAO report also asserts, "If these results [of the two NAS reports] are not fully reflected in the requirements, there is added risk that the technical strategy underlying the development of the system will prove not to be optimal, and that alternatives will not have been considered." We disagree with this conclusion. NARA is articulating requirements to reflect its mission needs and the interests and needs of its stakeholders. The requirements will state what the system must do, not how it should accomplish these goals. Rather than dictate a solution, NARA will ask industry to propose the optimal methods of satisfying our requirements. Any other approach would create unnecessary and inappropriate barriers to acquiring the best possible solution that the market can provide. In this context, it should be noted that the NPACI architecture for persistent

archives is a notional architecture. It does not specify any particular hardware, software or network architecture. Furthermore, after contract award for design of the system, the ERA Program will enter a requirements definition and refinement stage ending with a System Requirements Review, collaborating with the contractor to finalize what is to be built. This will be another opportunity to fold in additional research-related information.

With respect to the second part of the NAS study, the GAO report states, "By this date [October 1, 2003], the Request for Proposals for the electronic archival system will be released, leaving little or no opportunity for the results of the second assessment to influence the first build of the system." The primary purpose of the second NAS report, however, is to provide input to NARA's long range plans for addressing the continuing evolution of information technology and electronic records. As stated in the NAS contract statement of work, the second part of the study "will provide a more comprehensive discussion of the digital archiving and preservation issues and options confronting the National Archives and Records Administration." The second NAS report will be useful in revising the ERA research plan to address new problems and opportunities identified by the experts, and in plans for successive builds of the ERA system. Even in the initial build, we intend to provide the second NAS report to contractors to develop designs for the initial build of the ERA system. Given that design work will start only after award of the contract, the contractor will be able to take the NAS assessment into account in developing its design, and NARA will be able to use it in evaluating the design.

Thank you for considering our comments. As your report recognizes, we face enormous challenges in managing and preserving electronic records, and we welcome the perspective GAO brings to these issues. Work we already have underway will be instrumental in meeting the report's recommendations, and we will be pleased to report to you and the Congress regularly about our progress.

If you have any questions, please contact Lori Lisowski, Director of Policy and Communications, at 301-837-1850.

Sincerely,

JOHN W. CARLIN
Archivist of the United States

Enclosure

Glossary

administrative records	Records created by several or all federal agencies in performing common facilitative functions that support the agency's mission activities, but do not directly document the performance of mission functions. Administrative records relate to activities such as budget and finance, human resources, equipment and supplies, facilities, public and congressional relations, and contracting. Administrative records are temporary and are covered by general record schedules.
business process	A collection of related, structured activities—a chain of events—that produce a specific service or product for a particular customer or customers.
data architecture	The framework for organizing and defining the interrelationships of data in support of an organization's missions, functions, goals, objectives, and strategies. Data architectures provide the basis for the incremental, ordered design and development of systems or subject databases based on successively more detailed levels of data modeling.
electronic record	In the context of the federal government, any information that is recorded by or in a format that only a computer can process and satisfies the definition of a federal record in 44 U.S.C. 3301.
electronic recordkeeping system	An electronic system in which records are collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition.
enterprise architecture	An institutional systems blueprint that defines in both business and technology terms an organization's current and target operating environments and provides a road map for moving between the two.
Extensible Markup Language (XML)	A flexible, nonproprietary set of standards for tagging information so that it can be transmitted using Internet protocols and readily interpreted by disparate computer systems.
federal records	In the context of federal recordkeeping, 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 U.S. government under federal law or in connection with the transaction of public business, and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, operations, or

	other activities of the government or because of the informational value of the data in them.
metadata	Data containing descriptive information about other data.
office automation records	Electronic records created by means of office automation software, such as word processors, spreadsheets, other desktop applications, or electronic mail.
office automation	The techniques and means used for the automation of office activities, in particular, the processing and communication of text, images, and voice.
permanent records	Records that NARA appraises as having sufficient value to warrant continued preservation by the federal government as part of the National Archives of the United States.
Portable Document Format (PDF)	A proprietary de facto standard for electronic document distribution worldwide. Created by Adobe Systems, the portable document file format preserves all the fonts, formatting, graphics, and color of any source document, regardless of the application and platform used to create it.
program records	Records created by each federal agency in performing the unique functions that stem from the distinctive mission of the agency. The agency's mission is defined in enabling legislation and further delineated in formal regulations. Program records may be temporary or permanent; they must be scheduled.
record	See *federal records*.
recordkeeping system	A manual or automated system in which records are collected, organized, and categorized to facilitate their preservation, retrieval, use, and disposition.
recordkeeping	The act or process of creating and maintaining records.
records management	The planning, controlling, directing, organizing, training, promoting, and other managerial activities involved in records creation, maintenance and use, and disposition in order to achieve adequate and proper documentation of the policies and transactions of the federal government.

records management application	The term used by the Department of Defense's *Design Criteria Standard for Electronic Records Management Software Applications* (DOD 5015.2-STD) for software that manages records. The primary management functions of such software are categorizing and locating records and identifying records that are due for disposition.
records schedule	A document providing mandatory instructions for what to do with records no longer needed for current business, with provision of authority for the final disposition of recurring and nonrecurring records.
technical reference model	A taxonomy that provides a consistent set of service areas, interface categories, and relationships to address interoperability and open systems; part of an enterprise architecture.
temporary records	Records appraised as having temporary or limited value and approved for destruction either immediately or after a specific period of time.
Usenet	An Internet-based worldwide distributed discussion system. Usenet consists of a set of "newsgroups" with names that are classified hierarchically by subject. "Articles" or "messages" are "posted" to these newsgroups by people on computers with the appropriate software; these articles are then broadcast to other interconnected computer systems via a wide variety of networks.
XML	See *Extensible Markup Language*.
XML document	A text document marked up with hierarchically arranged descriptive tags and attributes conforming to the XML standard. An XML document can also begin with declarations that refer to other files providing further instructions for interpreting and displaying data elements.

GAO's Mission

The General Accounting Office, the investigative arm of Congress, exists to support Congress in meeting its constitutional responsibilities and to help improve the performance and accountability of the federal government for the American people. GAO examines the use of public funds; evaluates federal programs and policies; and provides analyses, recommendations, and other assistance to help Congress make informed oversight, policy, and funding decisions. GAO's commitment to good government is reflected in its core values of accountability, integrity, and reliability.

Obtaining Copies of GAO Reports and Testimony

The fastest and easiest way to obtain copies of GAO documents at no cost is through the Internet. GAO's Web site (www.gao.gov) contains abstracts and full-text files of current reports and testimony and an expanding archive of older products. The Web site features a search engine to help you locate documents using key words and phrases. You can print these documents in their entirety, including charts and other graphics.

Each day, GAO issues a list of newly released reports, testimony, and correspondence. GAO posts this list, known as "Today's Reports," on its Web site daily. The list contains links to the full-text document files. To have GAO e-mail this list to you every afternoon, go to www.gao.gov and select "Subscribe to daily E-mail alert for newly released products" under the GAO Reports heading.

Order by Mail or Phone

The first copy of each printed report is free. Additional copies are $2 each. A check or money order should be made out to the Superintendent of Documents. GAO also accepts VISA and Mastercard. Orders for 100 or more copies mailed to a single address are discounted 25 percent. Orders should be sent to:

U.S. General Accounting Office
441 G Street NW, Room LM
Washington, D.C. 20548

To order by Phone: Voice: (202) 512-6000
TDD: (202) 512-2537
Fax: (202) 512-6061

To Report Fraud, Waste, and Abuse in Federal Programs

Contact:

Web site: www.gao.gov/fraudnet/fraudnet.htm
E-mail: fraudnet@gao.gov
Automated answering system: (800) 424-5454 or (202) 512-7470

Public Affairs

Jeff Nelligan, managing director, NelliganJ@gao.gov (202) 512-4800
U.S. General Accounting Office, 441 G Street NW, Room 7149
Washington, D.C. 20548

PRINTED ON RECYCLED PAPER

United States
General Accounting Office
Washington, D.C. 20548-0001

Official Business
Penalty for Private Use $300

Address Service Requested

Presorted Standard
Postage & Fees Paid
GAO
Permit No. GI00