Building a Future for Our Times: the Electronic Records Archives of the National Archives and Records Administration, U.S.

Kenneth Thibodeau, Director
Electronic Records Archives Program
National Archives and Records Administration

Preface

First, I wish to thank the President of the National Archives of Japan, Mr. Kikuchi, for inviting me to speak at this seminar. I must also express gratitude to the National Archives of Japan for the kind hospitality shown to me and to my wife. My wife, Dr. Sharon Thibodeau, is the Director of Archival Operations at the National Archives in Washington. As archivists, both of us appreciate the opportunity to attend the EASTICA Seminar, and we are sure we will profit from the experience.

Introduction

The increasing prominence of computers and digital networks in the conduct of the affairs of governments, companies, universities, and private persons compels archival institutions to look differently at themselves and their place in society. Archives are typically seen as connected to the past, because of their special role in preserving uniquely valuable historical records. But the essential mission of archives requires us to be forward looking. Our responsibility is to carry the past forward, not only to make old records available now, but also to deliver them and records created in our times to the future. Delivering records of contemporary activities to the future is a major challenge. To put it simply we don't know how. We fall short in two fundamental areas.

First, no one knows how to preserve and provide sustained access to authentic electronic records for most types of electronic records being created today. On the empirical level, there are no proven methods for preserving most types of electronic records. On the theoretical level, archival science has not articulated a set of criteria that are adequate and suitable for implementing in system design and operations in order to determine whether any digital preservation or access method preserves or presents authentic records.

Second, no one knows what information technology will be like in the future. The only thing we can safely assume is that it will be significantly different than the technology we know today. This is an important problem for archives because we have to assume that (1) in the future people will want to use the best available technologies to discover and access records from the past and (2) the best technologies will not be those that were used originally to create, process, and store the records. In many cases the best technologies will not even be those

adopted by the archives even just a few years before the researcher seeks the records.

Building for the Future

Even though we do not know the answers – and with respect to future technology cannot know the answers – archival institutions must accept the responsibility for enabling people, who will want to use the records of our times 50 and 100 years from now, to do so. In effect, we must build a future for our times by developing the ability to preserve and provide sustained access to authentic electronic records. If we do not, much crucial evidence of the late twentieth and early twenty-first centuries will be lost. But how can we reasonably and realistically prepare for an uncertain future in the context of the twofold ignorance I have described? There are several things we can do. In exploring what we can do, I will consider issues from a general perspective, and I will also tell you how the National Archives and Records Administration (NARA) of the United States is approaching the future. I will devote particular attention to our initiative to build the Electronic Records Archives system.

The first thing archives need to do is to assume an appropriate posture, one which anticipates change in the characteristics of electronic records, in the technologies used to preserve and provide access, and in the expectations and behaviors of researchers. Our plans must be permeated by dedication to meeting the unspecified and unspecifiable needs of future generations of researchers in archives. We must complement that all encompassing dedication with specific awareness of what we do not know. Dedication to meeting future requirements and awareness of current limitations are matters of attitude or posture. Will they make a significant difference? They should. Consider the alternatives. If we assume we can predict the wants and expectations of future generations of researchers based on what we know of people who use archives today, we will fall short. If we assume we can create definitive solutions for the problems of preservation and sustained access, we will – at the least – make it very difficult to take advantage better alternatives that are developed in the future. At the worst, we will create archival solutions that do not solve the problem, but compound it. So, dedication to satisfying unknown future needs and awareness of current limitations should lead us to develop plans and approaches that are open-ended and capable of evolving over time.

The second thing we can do is to recognize those things that will not or should not change. There are two distinct bases for identifying what should remain stable over time. One basis is archival science. The second basis is the legal context of each archival institution, its authority and its responsibility. But we must be very careful in determining what must not change or we will needlessly hamper our ability to deal with what will change.

Archival science provides stable principles, concepts, requirements and understanding that apply to all archival institutions and that in essence are valid for all time. It dictates, for example that we always respect those properties that differentiate records from other types of information objects that are not records, notably their provenance, original order, and archival bond. From archival science we can derive rules for modeling and managing archival objects; such as fonds, series, records, records creators, etc., in a system. But we must apply established archival knowledge wisely, not slavishly. An article that Professor Luciana Duranti and I published last year in the journal, Archival Science, demonstrates that not only basic concepts of archival science but also particular practices that were developed even hundreds of years ago are valid and applicable in the digital environment; however, the realities of the digital environment also reveal that there are some areas where archival knowledge falls short and where established archival principles and practices lead in the wrong direction, either by leading us to conclude erroneously that received knowledge is not valid in the digital environment or by failing to recognize the need to modify or supplement received knowledge.[1] One example of erroneously rejecting established principle that I have encountered is archivists who assert that original order is not applicable to electronic records because the physical placement of electronic records is not significant. This is a case of confusing a solution with a requirement. The requirement for maintaining original order of records derives from the fact that the original order expresses the relationships among records that were defined by the records creator to satisfy its needs for retrieving information in subsequent activities. When records are kept on paper and other physical media, the most effective way to make these relationships operative, is to place related records together physically in file folders or other physical containers. Putting related records together physically is a means of expressing the relationship. It should not be confused with the requirement to preserve the relationships established by the records creator. In the digital world, there are many other ways to implement relationships among records. The requirement is for archives to determine how the records creator made relationships among records and to keep those relationships intact. For example, we must continue to respect the principles of provenance and original order; however, we must not allow our knowledge of how these principles applied to traditional records to blind us to significant differences that can occur when records creation and keeping is automated. A more challenging case is where a single computer system is used by many records creators. There are, for instance, computer systems in the U.S. Government that simultaneously serve the needs of several different agencies, and even of agencies of state, local and tribal governments. What are the provenance and original order of records in such systems? There are similar systems in other nations. I ask you to think about how the principles of provenance and original order apply to such systems. There is not a single provenance because there are many independent entities creating and keeping records in such systems, but the system imposes an overall order on the records of

[1] Luciana Duranti and Kenneth Thibodeau. "The Concept of Record in Interactive, Experiential and Dynamic Environments: the View of InterPARES," Archival Science 6, 1 (2006): 13-68 (Online: http://dx.doi.org/10.1007/s10502-006-9021-7)

the different creators, an order that is established and maintained by consensus of all the parties participating in the system, and that is necessary to ensure that each records creator can satisfy its particular needs for keeping and retrieving records. When thinking about such cases, if you have some difficulty in applying established principles, ask yourself whether the principle is invalid, or needs to be modified, or whether it simply needs to be applied differently because such systems are different than traditional record keeping systems. Overall, we must strike a balance between being true to the solid, immutable foundations of archival science and being open to enriching our science and improving our practice.

The authority and responsibility of each archival institution form the second basis of things that will not or should not change. The mission of an archival institution and the functions that must be carried out in order to accomplish that mission define the basic structure and functions that any system used to preserve and provide access to authentic records over time. There are certain generic functions and non-functional requirements that apply to any institution that preserves any information over time, whether the information takes the form of records or some other type. These requirements are articulated in the International Organization for Standardization's Open Archival Information System reference model.[2] However, such requirements must be refined for institutions which do preserve records in order to address the specific properties of records. Requirements for archival systems must be refined further according to the specific mission and legal status of each archival institution. Many basic facets of an archives' exercise of its mission will remain fairly stable over time. For example, there are some functions that my institution, NARA, has to assume we will always do, regardless of the increasing pace and prominence of change. We will always need to review and approve requests for disposition authority for federal records. We will always need to appraise what records should be preserved in the National Archives. We will always need to respect the principles of provenance and original order. We will always need to provide access to preserved records to anyone who wants to see them, limited only by specific legal restrictions on access.

The need to refine basic requirements to address different legal status directly applies within NARA because there are several different legal contexts in which we preserve records. NARA provides guidance and services for the management of all records of the U.S. Government. We do this under three different legal contexts. The first context is established by the Federal Records Act, which applies to all agencies of the U.S. Government. Under this law, NARA has broad authorities. No agency can destroy any record without permission from the Archivist of the United States. The Archivist also has authority to designate

[2] International Organization for Standards. Reference Model for an Open Archival Information System (OAIS). ISO 14721:2003

which records of an agency should be preserved in the National Archives and, when such records are thirty years old, he can compel agencies to transfer them to the National Archives. Under the Federal Records Act, NARA issues legally binding regulations on how agencies must manage their records, and we offer a variety of services and assistance for records management.

The second legal context is the Presidential Records Act, which applies to the records of the President of the United States and to those government officials and offices whose function is solely to advise the President. Under this law, when a President leaves office, all presidential records of his administration that remain in existence are immediately transferred to the custody and control of the Archivist. While a President is in office, no presidential records may be destroyed without the written concurrence of the Archivist.

The third legal context is one where NARA has authority to act at its discretion, but we have no power to compel anyone else to act. There are two different sub-contexts where we have discretionary authority. First, under the Constitution, NARA has no authority over the Congress or the Supreme Court, but we do preserve and provide access to the records of those bodies as a courtesy and in accordance with their direction. Second, under the National Archives Act, NARA may accept donations of historical materials from persons outside of the U.S. Government. We must manage, preserve and provide access to records according to the different laws, regulations and rules that apply in each context.

The third thing archives can do in order to reasonably prepare for an uncertain future is to make reasonable assumptions about the future. While we cannot know what new technologies will be invented in the future, or which ones will be successful, or how they will be used, there are certain things we can safely assume about information and communications technologies in general. Two general trends that are fairly certain to continue for a considerable time are the continuing growth and the dynamic evolution of technology. We can assume that computers will continue to become more common in the activities of institutions, the lives of individuals, and the interactions of groups of people. We can assume that the power, speed, capacity and usefulness of information technology will continue to grow, and we can expect prices to continue to decline for some time. We can also assume that the major information technology trend of the last decade, the growth of the Internet, will continue. Indeed, there are signs that the vision of ubiquitous computing – computers everywhere – which originated at the Xerox Palo Alto Research Center in 1988,[3] is becoming a reality. Not only are computers increasingly common on desktops, they are frequently embedded in other systems, including automobiles and airplanes, manufacturing equipment and medical devices, and even in buildings and clothing. To prosper in the future, archives must be able to function effectively in a world were computers are everywhere,

[3] M. Weiser, R. Gold, and J. S. Brown, The origins of ubiquitous computing research at PARC in the late 1908s. IBM Systems Journal. Volume 38, Number 4, 1999. http://www.research.ibm.com/journals/sj/384/weiser.html

are used for all kinds of things, and are constantly interacting with other computers as well as with people. Such interactions reflect one important way in which change in communications technology has induced major change in information technology. In the past, computer applications were designed as essentially closed systems, to the extent that they are described by the derogatory term, "stovepipe systems." Given the Internet, today's system developers assume that systems need to be open beyond simply accepting given inputs and producing given outputs. They need to be open to interact with different people, in different roles, doing different things. They also need to be open to interacting with many other systems on an as-needed basis, rather than by specific design. This technology trend will make it easier for archival institutions to adapt to a changing future. But we need to be careful in adopting technology so that we do not simply automate the way we have done things manually. That could very easily result in stovepipes which constrain our ability to adapt in the future.

The rapid growth of the Interned indicates that people increasingly want and expect to find information they are looking for on the Internet. Archives in the future need to locate themselves in the virtual space of the Internet. There are several basic features of the Internet terrain which merit consideration:

- The first feature is that from any point on the Internet, it is possible to communicate with any other point.

This feature has several implications for archives. For physical records, the only effective way for people to get access to the holdings of archives is, in most cases, to go to the archives. This is a practical restriction and, for many people, a burden which keeps them from using preserved records. But when records are in digital form, the archival holdings can go to people, wherever they are and whenever it is convenient for them. – Many archives have already recognized the benefit of this by digitizing current holdings and making them available on Web sites. – It is not only a matter of convenient access but also of outreach. Archives can reach far broader audiences on the Internet. This is reflected in the steps some archives have taken to translate their web sites into different languages. The National Archives of Japan, for example, has home pages in both Japanese and English. Thus the technology has already led some archival institutions to redefine who their researchers are. We should also be open to taking advantage of improvements in technology to develop new ways of making records accessible. At NARA for example, we are exploring how use technology to present records in ways that are more useful to history teachers, and other ways to make them more attractive to students. The converse of developing different ways to provide access is that archives must expect that people will change how they want to find and use information on the Internet. Archives should be aware of and respond to such changes.

There is also an important negative aspect of the ability to get from any point on the Internet to any other point. Hackers anywhere in the world can attack any system on the Internet. Attacks on systems take a variety of forms and they are

extremely common. This has an important implication for the design of archival systems. Archival systems must be secure, above all guaranteeing the integrity of the records preserved in them. We are designing security measures into the ERA system. To ensure the integrity of the records, we are implementing a series of controls on what can go into archival storage and in effect we will not allow any archival record to come out of storage. The only thing that can come out is a copy. If anything is done to the copy, such as removing sensitive content to create a version that can be released to the public, the copy could be placed in storage, but only by going through the existing controls, and it cannot replace the original version being preserved.

- A second significant feature of the Internet is that, in spite of the possibility of going from anywhere on the Internet to anywhere else, most points are practically invisible to most other points.

The most common way to find something on the Internet is to use a search engine. But search engines typically return information on only a fraction of the Internet. One telling example comes from the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). This is a method many institutions have adopted to expose their holdings on the World Wide Web. But a study in 2006 determined that 21% of the resources exposed through OAI-PMH are not indexed by any of three leading search engines. While Yahoo indexed 65% of the resources, Google indexed only 44%, and Microsoft Network only 7%. If information on the Web is not indexed by a search engine, it is difficult to find and it will not be found by most people.[4] Moreover, people seldom bother with information on the Internet if it doesn't show up in the first page or two of responses to their searches. A study at the University of North Carolina a few years ago found that archival descriptions rarely showed up in the first page of search engine responses even when the descriptions contained the exact words that were used in the search. The lessons for archival institutions are that it is not enough for archival institutions to put digital records on their Web sites. They must also adopt strategies that improve the probability that their records will be indexed by search engines; moreover archives must actively promote use of their holdings. Keep in mind that the value of archives to society will increase in proportion to the frequency of people finding archival records on the Internet, even when they are not looking for them.

NARA's major effort to address the challenges of the growth of information and communications technologies in general and of electronic records in particular is the development of the Electronic Records Archives system. We have articulated three fundamental requirements for the system as a whole that reflect dedication to meeting the needs of future researchers and recognizing current limitations. The system must be evolvable, scalable, and extensible.

[4] Frank McCown, et al

The Requirement for Evolution: Requiring the system to be capable of evolving is a direct response to the expectation of continuing change in information and communications technologies. We are convinced that every piece of hardware and software that is used in the system, both from the start and repeatedly over time, will become obsolete and unsupportable. Therefore, it must be possible to replace any piece of hardware or software with no loss of required functionality or capacity and with no significant impact on the records preserved in the system. Continuity, enhancement and expansion of functionality and capacity are achieved through a Service Oriented Architecture. In this architecture, each major component of the system sees other components as specified services, with interfaces that define how to request service from another component and how that component will deliver service. Hardware and software products are encapsulated within services, so that, if a product is replaced, it is necessary to establish an interface between the new product and the service in which it resides, but in many cases there will be no need to change the interfaces among service components. Independence of the records from the system is achieved by defining and associating metadata with the records such that, if the system were to disappear, the archives would be intact. In other words, the metadata is sufficient so that we could export the metadata and related electronic records from the system and the metadata would provide all the information needed by some other system to properly preserve, manage, and provide access to the records.

Figure 1 illustrates two services, the Business Applications Service and the Storage Manager Service. Business applications, such as records scheduling or transfer, can interact with storage in three ways: they can send data to storage, they can request data from storage, and they can receive data from storage. The Business Application Service defines a standard way for each of these interactions to take place, regardless of the different features of individual business applications. Thanks to the Business Application Service, the Storage Manager does not need to know anything about any of the individual business applications. It only needs to know what to do in each of the three standard interactions. Conversely, the Storage Manager Service converts storage interactions for different types of storage hardware and media, eliminating the need for any business application to know anything about the storage hardware or media being used. Thus, the two services allow business applications and storage technology to evolve independently of each other.

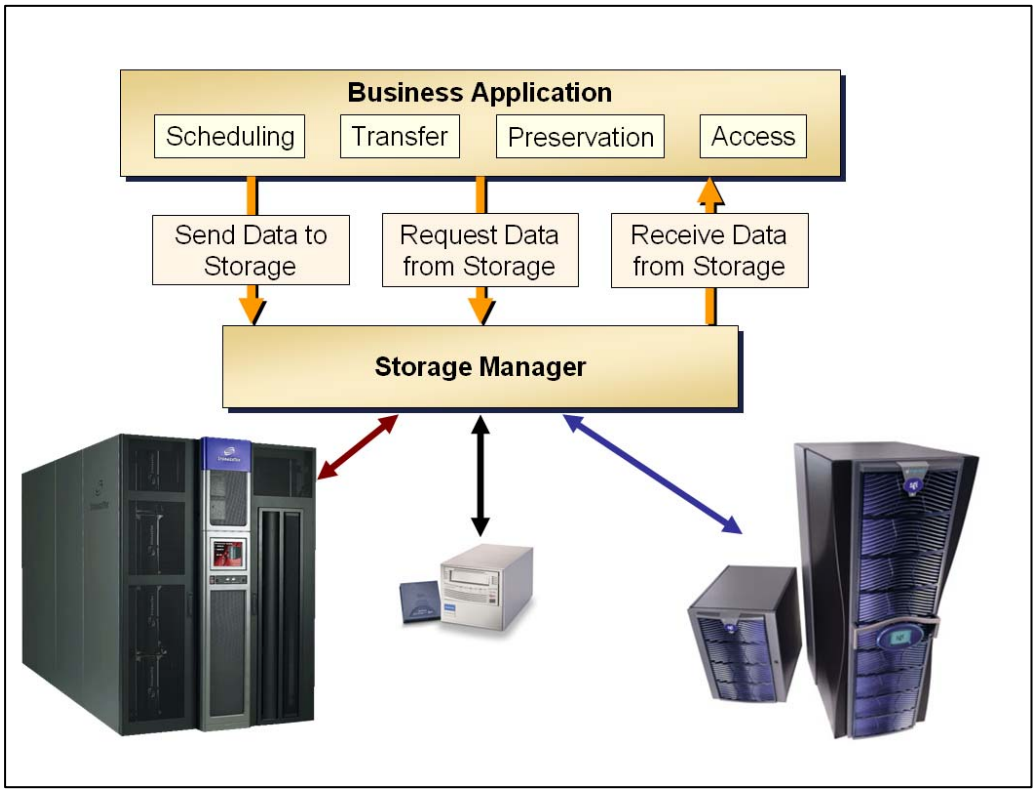
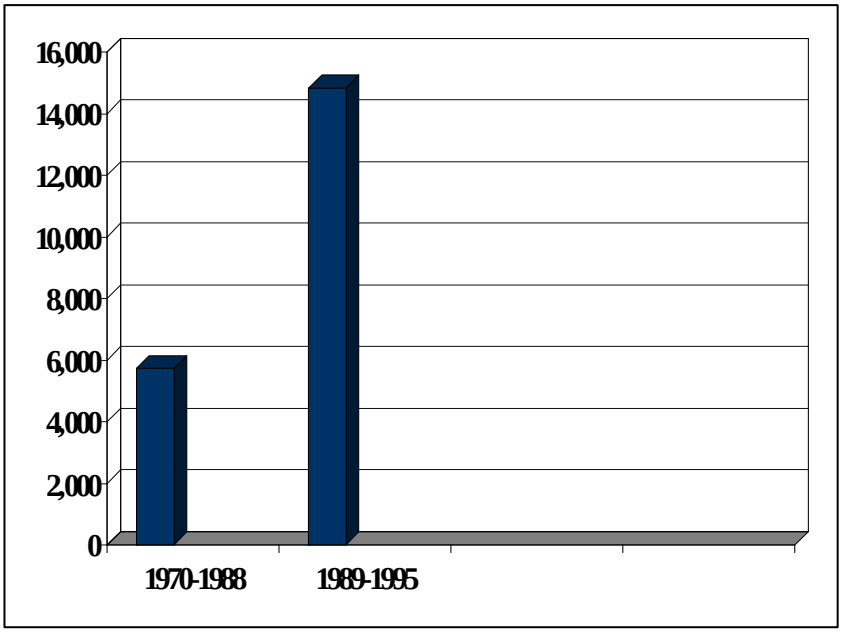


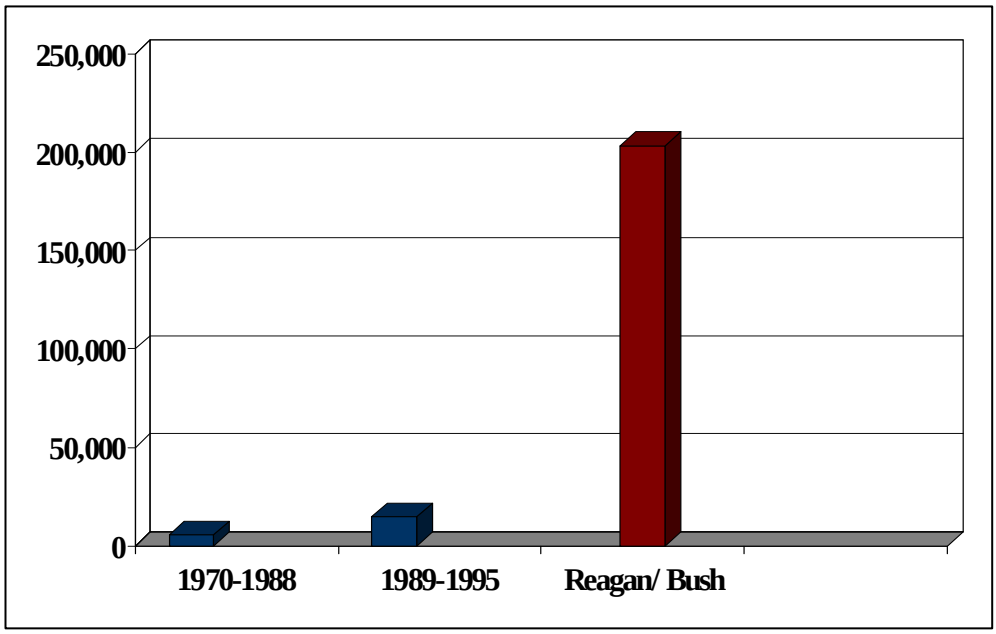
Figure 1. Illustration of Service Oriented Structure

The Requirement for Scalability: The system must be scalable in two directions. It must be possible to increase the size of the system in order to accommodate the expected growth in the quantities of electronic records we preserve. Real data about the volumes of electronic records being created in the U.S. Government is very limited. However, we have to anticipate enormous growth in our digital holdings. Our experience amply supports this anticipation. While paper and other traditional records still constitute the overwhelming majority of the records preserved in our National Archives, over the past decade our holdings of electronic records have increased one hundred times faster than traditional records. It was, in fact, data on growth that led NARA to establish the Electronic Records Archives Program. The relevant data can be displayed in three graphs. Graph 1 shows the transfers of digital files to the National Archives from the first transfer in 1970 to 1995. The data are divided into two groups, with the division marking the creation of the Center for Electronic Records in 1989. NARA created the Center in recognition of the need to focus its efforts in this area. As the graph shows, the effort was successful. There were substantial increases in transfers after the Center was created. In the 19 years before the Center, NARA received 5,720 files of electronic records. In its first 6 years, the Center brought in 14,835 files, increasing the average annual rate of transfers 9 times.



Graph 1. Transfers of Digital Files to the National Archives, 1970 - 1995

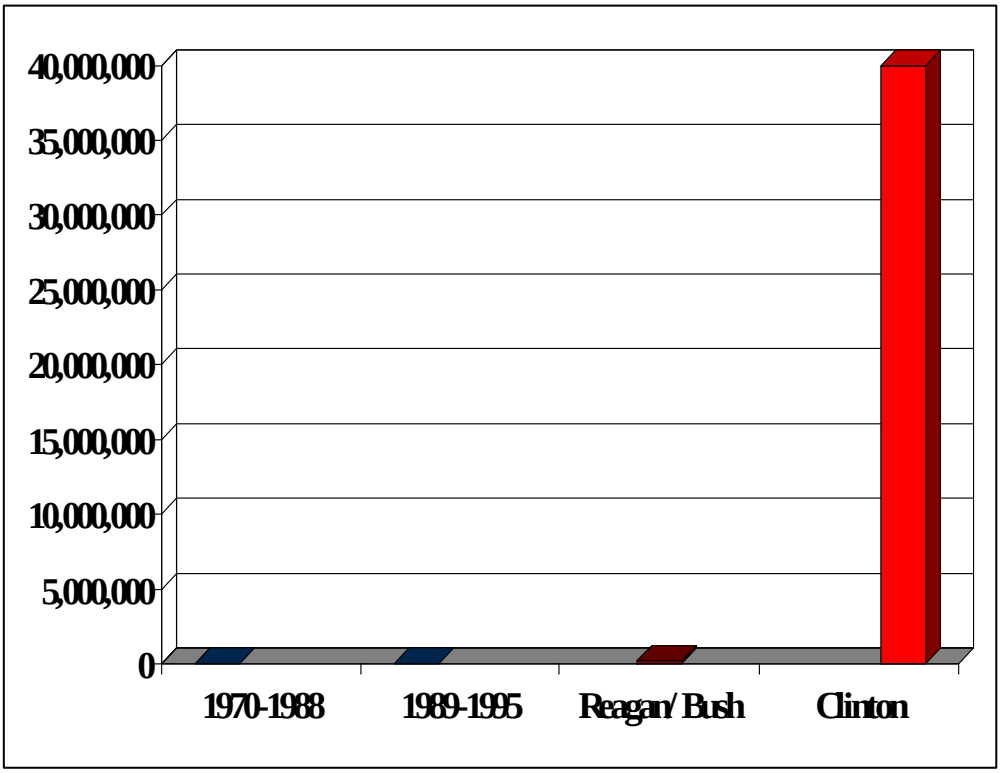
Graph 2 adds data showing the first transfer of electronic records from the White House, which happened in 1993, at the end of the first Bush administration and included records from both Presidents Reagan and Bush. Obviously, the presidential electronic records dwarfed everything that came before.



Graph 2. Transfers of Digital Files from the Reagan & Bush Administrations

Graph 3 adds the transfers of electronic records from the Clinton Administration. At this scale, all the previous transfers recede into the x-axis of the graph. When we estimated these transfers, in 1998, we determined that the systems NARA had in place could not be expanded to handle the volume of Clinton records. A new start was clearly required, one that would produce a system capable of growing without practical limits. It was this recognition that caused NARA to establish the Electronic Records Archives Program and that made scalability a basic requirement. We expect transfers of electronic records at the end of the current Administration to be much greater than those from the Clinton White House.

It must also be possible to implement an Electronic Records Archives system on a small scale. There are two reasons for this. First, NARA is responsible for some relatively small collections of electronic records which must be managed apart from other records. So we must be able to take the ERA architecture and implement it in a relatively small stand-alone system. The second reason for small scale is external. NARA does not have enough resources to do everything we need to do, not to speak of everything we want to do. Nevertheless, we recognize that we are a large archival institution and that we have much larger resources than sister institutions, such as the archives of the fifty states, and also archives in local and tribal governments and in universities. We hope that our investment in information technology produces solutions which can be useful in smaller archives.



Graph 3. Transfers of Digital Files from the Clinton Administration

The Requirement for Extension: The requirement that the system be extensible illustrates recognition of the limits of current knowledge. Given continuing changes in information technology, we know that there will be new types of electronic records created in the future, but we do not know what their specific formats or behaviors will be. Therefore, we need a system that can be extended repeatedly over time to enable us to bring in, manage, preserve, and provide access to new types of electronic records, regardless of their technical characteristics.

In addition to satisfying these underlying requirements for evolvability, scalability and extensibility, the structure and functions of the ERA system will support the processes that NARA carries out in order to accomplish its records management mission. The system will be a large and complex system which NARA will use in all of the legal contexts in which we operate; that is, for federal, presidential, congressional, Supreme Court and donated private records. The Electronic Records Archives may be described as a set of systems nested within one another, somewhat like a Chinese carved ivory ball. From a functional perspective, the design includes the following major elements:

- The Outer System, supporting lifecycle management functions for records of all types;
- The Inner System, processing and storing electronic records;
- Interior Frameworks for searching for records and information about records and for preserving electronic records; and
- Archival "mini-systems" providing systematic, case-by-case lifecycle management of individual bodies of records.

At a very general level, the two outer layers and the archival mini-systems support functions that we will always need to carry out, while the search and preservation frameworks will enable NARA to extend the system to deal with new types of electronic records and also to address uncertainty about the specific technologies that will be available for preservation and access. Because of its size and complexity, NARA is developing the ERA system in an incremental process. The initial system is scheduled to be put into operation in May 2008. After that, we will gradually add functionality and increase the size of the system through 2011. This incremental approach obviously will provide an empirical demonstration that the system is evolvable and scalable. It will also demonstrate extensibility because over the course of the development, we will gradually add the ability to deal with more and more types of electronic records.

The system we are currently deploying is limited to federal records, but we are already working on a system expansion to encompass presidential records. The ability to handle presidential records must be in place when President Bush leaves office in January 2009

The Lifecycle Management System

Given that NARA deals with electronic records in the context of policies, criteria and procedures that NARA uses to carry out its mission to manage the lifecycles of the records of the U.S. Government, in identifying the requirements for the ERA system, we determined that the system had to implement those policies, criteria, and procedures. As a matter of principle, we need to exercise lifecycle management functions in a coherent and consistent manner. We also recognized that the development of the ERA system gave NARA the opportunity to improve how it manages the lifecycle of all types of records. Therefore, it was decided that ERA should support lifecycle management functions for all records. The ERA system will support not only electronic records, but also traditional textual records on paper, analogue audio-visual records, cartographic and architectural records and others. The outer system supports these records management processes policies, criteria and procedures that NARA uses to carry out its mission to manage the lifecycles of the records of the U.S. Government. The outer ERA system provides automated workflow, execution of transactions, completion and processing of forms, enforcement of policies and business rules, and collaboration for the following records management functions: records disposition, transfer, preservation, and access.

To illustrate how ERA will support management of all types of records, NARA manages the disposition of federal records through records schedules that identify types of records agencies create in the course of business, determine how long each type should be kept to satisfy the needs of the creating agency, and indicate whether they should be transferred to the National Archives after that time. In general, records schedules apply to categories of records that agencies regularly and repeatedly produce. Thus records schedules are typically in effect for many decades. The ERA system will enable agency records managers to create records schedules and submit them to NARA for approval. It will allow NARA and agency officials to track the schedules through the review process and support the appraisal of records. It will group documents produced during the production and review of records schedule and manage them as operational records of NARA.

In defining requirements for the ERA system, NARA undertook a thorough review of how it carries out archival processes. The inescapable and difficult challenge of digital preservation also led NARA to change how it exercises the appraisal and preservation functions, and to couple the two processes. With respect to preservation, for physical records, preservation has been a function that began only after records are transferred to NARA's legal custody. In most cases, preservation actions are limited to placing the records in archival containers in appropriate storage spaces. Additional preservation actions relatively rare and they are taken on the basis of an assessment that only looks at whether such action is needed. Theses assessments have only a coincidental relation to the exercise of other archival functions. But, given continuing change in information technology, we have to assume that virtually all the electronic records we preserve will require

preservation actions and that preservation planning must start at the earliest opportunity.

Traditionally, appraisal was done at NARA as an intellectual exercise intended to decide if records had sufficient enduring value to be preserved after the records creator no longer needed them for its activities. If records had such value, the logical conclusion was, in most cases, that they should be preserved in the National Archives. With electronic records, however, an assessment that records should be preserved is meaningless unless the archives can reasonably assert that it will be able to preserve such records and provide access to them over the long term. Thus, we have decided to start preservation planning during appraisal. Initially, we determined that the appraisal process for electronic records must include an assessment of whether and how NARA will be able to preserve and provide sustained access to the records.

Moreover, even though we cannot know what technologies might be used to access electronic records 20 years or more into the future, we recognize that digital preservation actions can have important consequences for access, so that when we plan for preservation of electronic records we also plan for sustained access. We are, in effect, implementing the concept of the chain of preservation that was articulated by the first InterPARES project.[5] That is, we are ensuring that we can preserve electronic records authentic and provide access to them over the long term by integrating appraisal, preservation, and access in a comprehensive process that extends from the records creators' record keeping system through their transfer to NARA to ongoing preservation and access in the distant future. This process is detailed in a Lifecycle Management Plan for each series or set of records appraised as permanent. The Lifecycle Management Plan provides a way to insert controls into an archival system to ensure that archival principles are respected.

On reflection, we decided that it would be beneficial to extend this process to all types of records, and we are forming Lifecycle Management Teams comprised of representatives of all the archival units involved in any way with any body of records. Each team will be responsible for ensuring that the Lifecycle Management Plan is properly executed. Anticipating that, after appraisal is completed, there will be significant changes either in the records, in the way records creators keep and manage their records, or in the means NARA uses to transfer, preserve and provide access, the Lifecycle Management Teams are also expected to revise the plans appropriately.

The ERA system will do more than support the workflows for appraisal, preservation and access processes. Records Schedules, Lifecycle Management Plans, and other documents produced in records management activities are being designed so that the system can automatically extract from them data it will use in

[5] InterPARES. The Long-Term Preservation of Authentic Electronic Records: Findings of the InterPARES Project. September 2001. Available at www.interpares.org/book/index.cfm

managing the records. For example, NARA operates Federal Records Centers, where agencies may deposit inactive records that remain under agencies' legal control even though they are in NARA's physical custody. These records may be either temporary or permanent. The ERA system will use information contained in records schedules to initiate and carry out disposition actions on electronic records in Record Center custody. For temporary records, when the schedule data indicates the records are due to be destroyed, after notifying and obtaining concurrence from the controlling agency, it will destroy the records as provided in the applicable records schedule. Similarly, for permanent records, it will initiate and manage the process for transferring the records to NARA's legal custody.

1.1.1 The Electronic Records System

The Electronic Records System will provide tools and capacity for transfer, processing, storage, preservation, and access for electronic records. For transfer, the system will include software tools that agencies can download to extract electronic records from their systems and package them for transfer to NARA. When such packages are received, the system will automatically place them in temporary storage. ERA will associate every transfer with a records schedule item or other disposition agreement. It will use information in the Lifecycle Management Plan to determine if the records are what they are supposed to be. The initial validation of received records will be wholly automatic. The system will check if the records are in the digital formats we expect. It will examine their internal structure and contents to see if they correspond to what we know about the records. If the records satisfy NARA's expectations, the system will move them from temporary storage to archival storage. If any discrepancies or problems are discovered in validation, the system will notify an archivist who will work with the agency to resolve any issues. Once the records are in archival storage, the system will enable archivists to inspect the records manually, when necessary.

The Electronic Records System will execute preservation and access processes using the Preservation and Search Frameworks.

1.2 Interior Frameworks

The frameworks for preservation and access are the system components that enable an evolving and extensible approach to deal with new types of electronic records that are invented over time and to bring new technologies to bear both to improve service to researchers and other users and to implement superior solutions for long-standing problems. Among the limitations we recognize in the current state of affairs is the dearth of proven digital preservation techniques. We are also well aware that there are shortcomings with existing search technologies. Even apart from that, we know that over time we will need to use a variety of techniques for finding records. For example, NARA currently preserves very little in the form of digital photography or motion pictures, but we know we will receive substantial volumes of records in these formats in the near future. When

we do, we will need to add to the ERA system software tools for analysis, characterization, recognition and mining of pictures and videos. Similarly, we will need to incorporate digital preservation software products as they become available. Moreover, given that electronic records in different formats need different tools, the system will have to be able to include multiple preservation and access software tools at any given time, and it should have a way to select the best tool for a given job. The Search and Preservation Frameworks will make these things possible.

The Search Framework is a business service provided by the system. This service manages the execution of searches against assets in the ERA system. An "asset" is either an electronic record being preserved in the system or a business object, such as a Records Schedule or Lifecycle Management Plan. The Search Framework makes it possible to incorporate several commercial search software products in the system at the same time. The Search Framework chooses the most appropriate search engine to use for a given set of search criteria. In the initial system, however, we will only implement one search engine, the FAST product from the Norwegian company of that name.

The ERA Preservation Framework is a service that manages the execution of digital preservation processes in the system. The preservation section of the Lifecycle Management Plan for each body of records being preserved in ERA will specify conditions when preservation processes should be performed on those records. Possible conditions include a determination that the format of the records has become obsolete or the introduction of a superior preservation tool into the system. Such conditions will act as triggers in the system. When a trigger condition is met, the system will automatically execute preservation processes specified in the Lifecycle Management Plan. As the first step in a preservation process, ERA will compare the particular requirements for preserving the records covered by a plan with the specifications of the capabilities of available software tools in order to select the most appropriate tool. The Preservation Framework is not included in the initial system we are currently deploying. We plan to develop a prototype of the framework next year and to implement it in the system in 2009.

The Preservation and Search Frameworks may be described as forming the intersection between what will not, or should not, change, namely the records, with what has changed or will change, namely the technology used to encode and process the records.

Archival Mini-systems

Each Lifecycle Management Plan articulates a set of specific requirements for bringing for a series, sub-series, or other body of records into the ERA system; verifying that the right records – and all of the right records – are in fact transferred; preserving those records over the long run, maintaining the necessary context of provenance, archival bond and original order; ensuring that any restrictions on access are enforced; and providing authorized access to the records.

In effect, it defines what an archival system for those records must do. The architecture of the ERA system is rich enough in functionality and flexible enough in execution that we can implement within it an unlimited number of virtual mini-systems, each one devoted to satisfying the particular requirements for preservation and sustained access to a specific body of records. This will allow NARA not only to implement controls and processes that address the different rules governing federal records, presidential records, congressional records, and private records, but also, within each of these classes of preserved materials, to fine tune controls and processes for particular aggregates of records.

At the same time, encompassing all of these mini-systems within the overarching ERA system will allow us to improve service to researchers by providing them with a single Web portal where they can find information about all records NARA preserves, where they can request access to or copies of any of these records, and where they can get direct access to electronic records.

Overall, then, the structure of the ERA system might be better compared not to a set of nested balls, but to a gothic rose window, where there are multiple different systematic constructs interrelated in rich, intricate and harmonious ways. The Lifecycle Management System forms the outer perimeter of the structure. Other components interact with people and other systems through the Lifecycle Management System. Lifecycle Management Plans reside in this outer structure, but they define archival mini-systems which extend out beyond the periphery of ERA on the supply side to extract records from the systems in which they are created and stored and transfer them to NARA and on the demand side to inform people about the records and to enable them to obtain access, copies or versions. Within the confines of ERA, the archival mini-systems make us of the Electronic Records System and the Preservation and Search Frameworks to implement archival principles and legal requirements, to ensure that records disposition actions conform to plans, and preserve records in authentic form, and to optimize responses to user requests. The Preservation and Search Frameworks and the Service Oriented Architecture of the system as a whole will enable NARA to incorporate the expectation of change in the way we approach our mission. They will allow the system to grow in size, to extend to new types of records, to evolve both by overcoming technological obsolescence and by taking advantage in improvements in technology over time, and to respond to the changing expectations and behaviors of users, including NARA staff, officials of other government agencies, and the public.