Archilab

Saggistica - Strumenti/2

Gli archivi dei partiti e dei movimenti politici.
Considerazioni archivistiche e storiografiche

a cura di Siriana Suprani

(pp. 112, euro 10,33)

Analizza le problematiche relative alla gestione archivistica e all'uso storiografico delle carte di natura politica.

ARCHIVI & COMPUTER

AUTOMAZIONE E BENI CULTURALI

INDICE

Le regole tecniche per completare il recepimento della direttiva europea 1999/93/ CE non introdurranno modifiche tali da vanificare gli investimenti effettuati sulla base della precedente normativa. Certamente è stato e sarà necessario apportare alcune modifiche, conseguenti soprattutto alle ulteriori tipologie di firme elettroniche che consentiranno al mercato di offrire soluzioni di servizi di certificazione modulabili secondo l'esigenza specifica.

Ma i cittadini quando utilizzeranno la firma elettronica? Il futuro della firma elettronica è nella disponibilità di servizi che ne prevedono l'uso. È quindi uno strumento che concorrerà al processo di *e-government*, consentendo quando necessario di sottoscrivere atti verso la pubblica amministrazione. Può trovare inoltre forte impulso nella realizzazione di applicativi che consentano di rendere disponibili servizi fra privati in cui è necessaria la sottoscrizione con valore legale di una ben precisa volontà.

Kenneth THIBODEAU

Building the archives of the future. Advances in preserving electronic records at the National Archives and Records Administration[1]

Abstract: dopo aver descritto la natura dei documenti archivistici e la rilevanza di mantenere nel tempo contenuto, struttura e contesto delle entità documentarie, l'A. sottolinea la necessità di conservare in ambiente digitale i documenti informatici e la difficoltà di garantirne autenticità e accessibilità in assenza di metodi affidabili ed efficienti. D'altra parte la responsabilità dei National Archives di Washington non implica solo l'esercizio diretto della funzione conservativa, ma anche la definizione di regole e procedure per la formazione di documenti informatici presso le agenzie federali. La definizione di una strategia complessiva (Electronic Records Archives Program – ERA) ha costituito il primo punto di partenza e ha implicato, per la sua complessità, lo sviluppo di partnership con le principali istituzioni nazionali impegnate nel settore della conservazione digitale. Il modello di riferimento è la proposta di standard ISO Open Archival Information System (OAIS) Reference Model, sviluppata dal Consultative Committee on Space Data Systems e sostenuta dalla NASA. Per l'analisi dei principi e delle funzioni che sono alla base della conservazione, il progetto ha utilizzato i risultati della ricerca condotta da InterPARES, mentre per la individuazione dei metodi di trattamento conservativo si è avviata una collaborazione con il Supercomputer Centre dell'Università di S. Diego, nell'ambito di una partnership internazionale che coinvolge anche l'Università di Urbino. Le soluzioni in via di sperimentazione sono basate sullo standard XML per la produzione di "Persistent Archives". I risultati finora raggiunti sono incoraggianti, così come è significativo l'allargamento del nucleo originario delle istituzioni coinvolte nel programma di ricerca.

Information in digital form poses critical challenges for the National Archives and Records Administration (NARA). While many other institutions are facing such challenges, NARA's situation is different because of the special requirements that apply to archival institutions, NARA's unique role in the Federal Government, and the scale and diversity of the Government's programs. NARA views success in facing these challenges as entailing nothing less than building the archives of the future. In sober terms, unless we succeed in surmounting these challenges there will not be a National Archives of the United States for the digital era.

One of the key elements that distinguishes archives from other institutions which

---

[1]   An earlier version of this paper was presented at the Digital Library Forum 2000, Chicago, Illinois USA, November 19, 2000. The paper has been published on "D-Lib Magazine", February 2001, Volume 7 Number 2.

preserve information is that archives' essential responsibility is to preserve and deliver authentic records to subsequent generations of users. Records are documents accumulated in the course of practical activities. As instruments and byproducts of those activities, records constitute a primary and privileged source of evidence about the activities and the actors involved in them. While records are often conceived in terms of textual documents, such as letters and reports, they can take any form. What differentiates records from documentary materials in general is not their form, but their connection to the activities in which they are made and received. If this link is broken, corrupted, or even obscured, the information in the record may be preserved, but the record itself is lost. This fundamental difference between records and documents can be readily illustrated empirically. For example, a map of Sarajevo is a document, but a map of Sarajevo known to have been used in making a targeting decision that led to the bombing of the Chinese Embassy is an essential record of that action. The key difference between the document and the record is the specification of the context of action in which the record was involved. To preserve authentic records entails preserving the documents themselves and also their connections to the activities in which they were used. Archivists have identified three classes of attributes of records that must be preserved: their content, their structure, and their context. The first two are common to other types of documents. The context of records is expressed primarily in their relationships to other records created by the same actor. In simple terms, if we wish to mine the evidence available in the records of an activity, we need to know how those records are interrelated. The relationships among records established by the records creator give us the most immediate access to the connection between the records and the activities in which they were used and accumulated. In broad terms, the immediate context of a record is its position in a set of records as ordered by the records creator. The immediate context is the most direct path available to the significant context, the activity in which the record was used. To preserve records means to preserve them in their original order. To extend the National Archives of the United States into the digital era, then, entails being able to preserve the content, structure and context of the records. When any of these elements can only be expressed in digital form, the records must be preserved in that form. For NARA, as for other archival institutions, the difficulty of doing so is compounded by the commitment to preserve records permanently.

Unfortunately, given the current state of the art, effective methods for preserving most forms of digital information and most ways of organizing digital information are not available. The difficulty of digital preservation is further accentuated in NARA's case because of its authority and responsibility for life-cycle management of the records of all three branches of the Federal Government. NARA needs to acquire the capability to preserve and deliver any type of historically valuable record that may wind up in the National Archives or a Presidential Library. The wholesale absence of proven methods for digital preservation presses acutely on NARA. But NARA is not only responsible for preserving unique historical materials, but also for guiding all other federal agencies in creating and managing all of the records they need in performing their functions. The requirements for managing active records in support of the specific needs of ongoing activities are significantly different from those entailed by the objective of preserving and delivering authentic records to future users whose interests, objectives, methods and tools are essentially unknowable. NARA must find preservation methods for electronic records that will enable it to demonstrate the continuing authenticity of the records over unlimited time frames, but it also needs to find solutions that, at the least, do not conflict or compete with methods that serve a very different objective: that of effective and efficient management of the records in support of the current business of government. Ideally, records management methods and archival preservation methods should be complementary and mutually reinforcing. The need to balance current needs with long-term preservation and access is rendered even more complex when one recognizes that, overwhelmingly, most of the records created in the course of business are destroyed, appropriately, once the business needs have been satisfied.

The challenges NARA faces are further compounded by the scope and scale of its responsibilities. The Federal Government is a large, complex entity engaged in a bewildering variety of activities. The records NARA is responsible for preserving range from those produced in enacting laws to the personnel files of veterans, from the conduct of foreign affairs to the investigations of independent counsels; from the interdiction of narcotics to the topography of the United States. The president, the congress, the courts, and federal agencies employ a practically unlimited and continually changing variety of computer systems, digital media, and applications in conducting their business. The only reasonable assumption NARA can make is that preserving the electronic records of the Federal Government requires the ability to preserve virtually every class of digital object that has been, or may be created. Beyond that, NARA needs to contend with explosive growth in the quantities of electronic records it needs to preserve. The first accession of electronic records into the National Archives was in 1970. Since then, holdings of electronic records have grown exponentially, and the available data indicates that exponential growth will continue in the future.

How, then, will NARA build the archives of the future, the Electronic Records Archives? Adding together the technical difficulty of preserving and delivering digital information objects over an indefinitely long time, the diversity of the problems embodied in federal records, the rapid rate of growth of records in electronic form, and above all the critical importance of NARA's government-wide authority and responsibility for lifecycle management of records leads to a strategy of attacking the challenges in collaboration with other stakeholders. NARA formally established this collaborative strategy in the agency's strategic plan published in 1997, and has followed this strategy from the beginnings of its current Electronic Records Archives (ERA) Program[2]. NARA has pursued, created, and taken advantage of opportunities for

2   Electronic Records Archives Program, http://www.nara.gov/era/.

collaboration in a variety of venues. The agency is open to additional partnerships where they show promise. For example, NARA is in negotiations with several other national archives. Recently, at the request of Senator Ted Stevens, Chair of the Appropriations Committee, NARA helped the Library of Congress build a case for funding its strategic initiative in the digital library arena[3].

There are six key partnerships that form the core of the ERA Program. The foundation lies in the international effort to develop the Open Archival Information System (OAIS) Reference Model. Chartered by the Consultative Committee on Space Data Systems and spearheaded by NASA, the OAIS initiative is articulating the functionality and components of any system responsible for preserving any type of information over any length of time. The OAIS model is currently a draft ISO standard[4]. While it originated in an effort to address data requirements in the space science community, from the beginning the activity has been intentionally interdisciplinary. NARA has been active in this effort from the beginning in 1995 and has hosted fifteen of the eighteen U.S. OAIS workshops. OAIS is a reference model and not a guide to implementation. For the ERA Program, the OAIS model provides a high level framework for entities, functions, data flows and administrative activities.

The second major foundational collaboration for ERA is the International Research on Permanent Authentic Records in Electronic Systems (InterPARES) project[5]. As indicated by its name, this project focuses on the preservation of authentic electronic records. It is attempting to: determine the archival requirements for authenticity of different types of electronic records; to identify principles and practices to apply in selecting records for preservation to maximize the probability of successful preservation of the records; articulate the processes, inputs, outputs, controls and mechanisms for archival preservation of electronic records; to evaluate technological options of executing these processes; and develop frameworks for policies and standards for preservation. InterPARES involves representatives of ten national archives, with research being conducted by seven multidisciplinary research teams drawn from academia, government and private industry in North America, Europe, Asia and Australia[6]. NARA is a charter member of the project and supports the work of the U.S. Research Team through a grant from the National Historical Publications and Records Commission (NHPRC)[7]. In sum, NARA looks to InterPARES for rigorously developed and widely vetted archival requirements and methods for the preservation of authentic electronic records. InterPARES is building on the OAIS effort in elaborating a formal model for preservation of authentic records that is based on the OAIS reference model.

While OAIS and InterPARES are laying the foundations of the ERA program, the core of the program stems from the Distributed Object Computation Testbed (DOCT). DOCT was launched as an interagency collaboration between the Department of Defense's Advanced Research Projects Agency and the U.S. Patent and Trademark Office. NARA joined this collaboration in 1998, specifically expressing concern about long-term retention of records created, communicated and managed in advanced, high-performance computing environments. This concern was addressed in a special tasking to one of the primary research centers involved in DOCT, the San Diego Supercomputer Center (SDSC). SDSC addressed the long-term issues by inventing not only a preservation method, but an information management architecture built around the objective of preservation of arbitrarily structured sets of virtually any type of electronic record. This architecture, and the preservation method it comprises, were initially referred to under the rubric of "Collection-Based Persistent Object Preservation". SDSC articulated the persistent object approach on the basis of the OAIS model. Developed in the first year of research SDSC conducted for NARA, Persistent Object Preservation was described extensively in " D-Lib Magazine" in 2000[8]. In the second year of its work, SDSC has enriched the architecture substantially, to the point where it merits the appellation of "Knowledge-based Persistent Object Preservation"[9]. Given that NARA's ultimate objective is not that of advancing the state of the art, but technology transfer that will enable the agency to build the Electronic Records Archives, NARA has required empirical demonstration of the research results in the DOCT project. Repeated and consistent success in demonstrating the Persistent Object Preservation approach has led us to regard this approach as the most promising one ever suggested for preserving digital information in general, and electronic records in particular.

The success of SDSC's work in the DOCT project has led NARA to expand the horizon of the ERA Program. The focus remains centered on the critical problem of the absence of proven methods capable of satisfying archival requirements for preservation of authentic records, but the promise of Persistent Object Preservation enables us to look beyond the infinite horizon of permanent preservation, back to the beginnings of the records lifecycle. The retention of active records for long-term business needs is logically an incremental extension of archival preservation of permanently valuable records. There are a wide variety of business processes in government that require long-term retention of related records. Persistent Object Preservation offers substantial promise for the survival of information assets, whether they are needed for twenty-five years, seventy-five, or forever. In this perspective, the survival of electronic records

[3] Library of Congress & National Archives and Records Administration, *Challenges and Collaboration for Sustained Access to Digital Materials. Joint presentation to members of the Joint Committee on the Library, U. S. Senate*, December 17, 2000.

[4] ISO Open Archival Information System Standard. http://ssdoo.gsfc.nasa.gov/nost/isoas/.

[5] InterPARES Program. http://www.interpares.org/.

[6] Anne J. Gilliland-Swetland and Philip B. Eppard. *Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project*, in "D-Lib Magazine". July-August 2000, volume 6, number 7-8 http://www.dlib.org/dlib/july00/eppard/07eppard.html.

[7] US-InterPARES Project. http://is.gseis.ucla.edu/us-interpares/.

[8] R. Moore *et al.*, *Collection-Based Persistent Digital Archives*, in "D-Lib Magazine", March 2000, volume 6 number 3 [Part 1], http://www.dlib.org/dlib/march00/moore/03moore-pt1.html; number 4 [Part 2], http://www.dlib.org/dlib/april00/moore/04moore-pt2.html.

[9] B. Ludäscher, R. Marciano, R. Moore, *Towards Self-Validating Knowledge-Based Archives*, 11th Workshop on Research Issues in Data Engineering (RIDE). Heidelberg, Germany, IEEE Computer Society, April 2001, available at: http://www.npaci.edu/DICE/Pubs/.

and other digital information objects appears as an essential element in an information infrastructure capable of supporting electronic government, as well as electronic commerce and scientific research in a digital environment. Toward this objective, in March of 2000 NARA joined the National Science Foundation as a cosponsor of its National Partnership for Advanced Computational Infrastructure (NPACI) program. Through NPACI, NARA is supporting additional research into the development of Persistent Object Preservation both at SDSC, which is the leading-edge institution in NPACI, and at other research centers in the 46 member partnership. NARA's involvement in NPACI has also led to an expansion of that partnership, with the recent establishment of an archival research program headquartered at the University of Urbino, Italy, as a Foreign Affiliate of NPACI[10].

Another collaboration aimed at technology transfer links NARA with the U.S. Army Research Laboratory and Georgia Tech Research Institute. Called the Presidential Electronic Records Processing Operational System (PERPOS), this project is exploring, evaluating and developing advanced information technologies applicable to archival processing of electronic records. As in the Persistent Object Preservation research, the PERPOS project provides empirical demonstrations of research results. The demonstrations have focused on presidential records from the last Bush Administration, but the technologies being considered have broad applicability. Started in late 1998, the research to date has concentrated on the critical need to identify and filter out all presidential records from the totality of digital files left behind at the very end of the Administration. The total population includes a large percentage of files of operating system software, applications software, tutorials, templates and the like. In addition, the technologies under investigation offer significant potential for other processes, such as accessioning the records into the presidential library, automatically producing descriptions of the records, and finding sensitive information in the records[11].

Another type of technology transfer is targeted in the Archivist's Workbench project, funded under an NHPRC grant to the San Diego Supercomputer Center. This project aims at scaling the Persistent Object Preservation approach for applicability in smaller institutions, such as state and university archives.

Where are these collaborations leading? At bottom, they have already produced a major change in NARA's perception of what is involved in solving the challenge of electronic records. This challenge has been seen by archivists, and others, as stemming from two basic problems: the lack of durable media for storage of digital information and the rapid obsolescence of the hardware and software needed to retrieve, process and communicate the information. NARA's long experience in preserving electronic records has led us to view the media problem as manageable. The initiatives described above, however, have led us to perceive the challenge of electronic records as involving opportunities as well as problems. NARA now perceives that a solution to this challenge entails:

[10]   NPACI and SDSC Online, http://www.npaci.edu/online/v4.24/urbino.html .
[11]   PERPOS (Presidential Electronic Records Pilot Operations System), http://perpos.gtri.gatech.edu/

- Overcoming technological obsolescence in a way that preserves emonstrably authentic records;
- Building a dynamic solution that incorporates the expectation of continuing change in information technology and in the records it produces; and
- Finding ways to take advantage of continuing progress in information technology in order to maintain and improve both performance and customer service.

Thus, the Electronic Records Archives is envisaged in its totality not as a system in the usual sense, but as a comprehensive, systematic, and dynamic means of accomplishing the archival work that must be done to provide continuing access to authentic electronic records over time. Clearly, it would be shortsighted to believe that the challenge of preserving electronic records could be met simply by building a system. Any system, conceived as a final solution, even if it solved all of the known and knowable problems of obsolescence and fragile media, would itself inevitably become obsolete in what, from an archival perspective, would be a relatively short time. Furthermore, probably the only valid prediction about the future of information technology is that it will continue to change. Therefore, the solution to the challenge of digital preservation must incorporate the capability to accommodate and incorporate changing technology and unforeseeable products of that technology. Finally, a solution to the challenge of electronic records should take advantage of improvements in information technology as well as address retrospective problems such as format obsolescence. It would be unrealistic to expect that future users would be satisfied with having their access to electronic records limited to what had been available under antiquated technology. Researchers today would hardly be satisfied if access to old records required entering queries on punch cards, in FORTRAN or COBOL, with output limited to printouts in upper case. Similarly, we must anticipate that in the future there will be improved options available for ingest, preservation, and archives management as well as access.

The Electronic Records Archives, at present, is the objective of a set of interrelated collaborations which are largely research and development activities. Although a number of demonstrations and prototypes have been developed, ERA is, for all practical purposes, under development. It will undoubtedly be several years before the vision of the archives of the future is fully realized in an operational mode. Nevertheless, enough progress has been made to lay out both a developmental strategy and a conceptual model.

The developmental strategy is depicted in *figure 1*. The strategy extends the collaborative approach already in place. The developmental strategy has four major components. First, it aims at building archival solutions on the basis of technologies being developed to support electronic government, electronic commerce, and research. Base technologies, represented by the grid at the bottom of *figure 1*, are ones which are being developed to support a wide spectrum of applications, largely independently of concern for long-term preservation and access. They include the eXtensible Markup

Language (XML) family of standards and various "mediation" and "grid" technologies that enable different computing platforms and storage resources to interact coherently. Second, on the base of such general technologies, the strategy envisages developing an information management architecture capable of preserving and delivering digital information across generations of information technologies and that is applicable across as broad a range of requirements as possible. Ideally, solutions at this level, which are represented by the bottom level of the pyramid in *figure 1*, should be applicable in digital libraries, scientific data centers, and even in cases of records retained to meets the needs of current business, as well as in archives. This architecture will embody the persistent object preservation approach discussed above. It will need to be refined to address the special needs of archives, which are differentiated from other institutions by their mission of preserving authentic records. The third element of the developmental strategy — the "Framework" layer in *figure 1* — consists of solutions that specifically address these needs. It is expected that these solutions will either fine tune more general solutions or supplement them. The intention is that solutions even at this level will have broad applicability to a variety of archives.
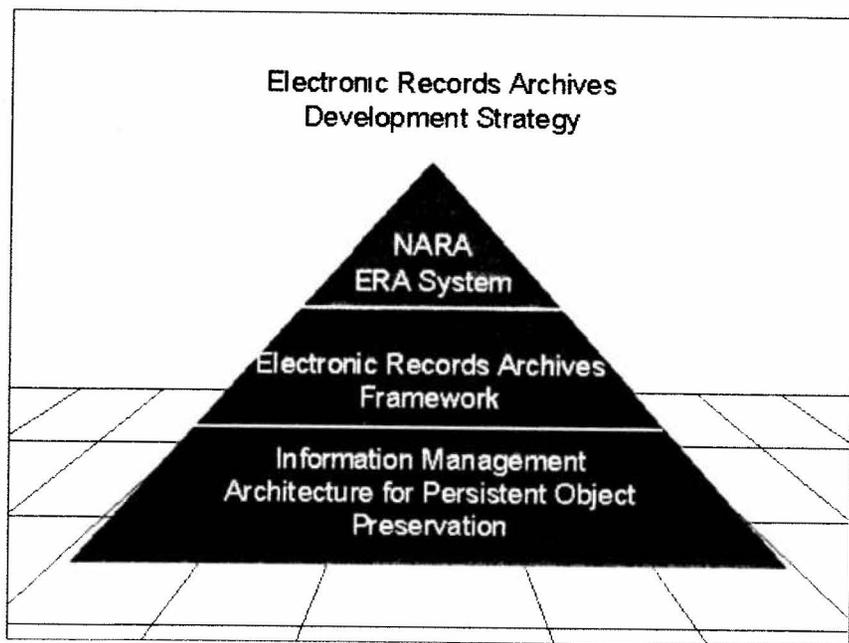
The NHPRC grant, the University of Urbino Affiliate of NPACI, and the InterPARES project, all mentioned above, are contributing to this end. Finally, NARA's specific needs will be the focus of the last, and smallest of the developmental efforts. NARA's needs derive from its responsibility for the National Archives of the United States, for presidential libraries, and for records management in the Federal Government, and also from requirements to conform to special legislation and regulations, such as the Freedom of Information Act and rules governing security classified records. The unifying theme in this developmental strategy, as indicated in the figure, is to develop as much of the solution as possible on the broadest available base. It recognizes that the market for archival technology is inadequate to drive and support the technical developments required for ERA, and that the broader the base of support for the technologies used in ERA, the more likely that those technologies will be robust. Furthermore, given NARA's responsibility for lifecycle management of records throughout the Federal Government, it is highly desirable to use technologies that can be applied to the management of records retained for ongoing business needs, as well as for archival preservation and access.
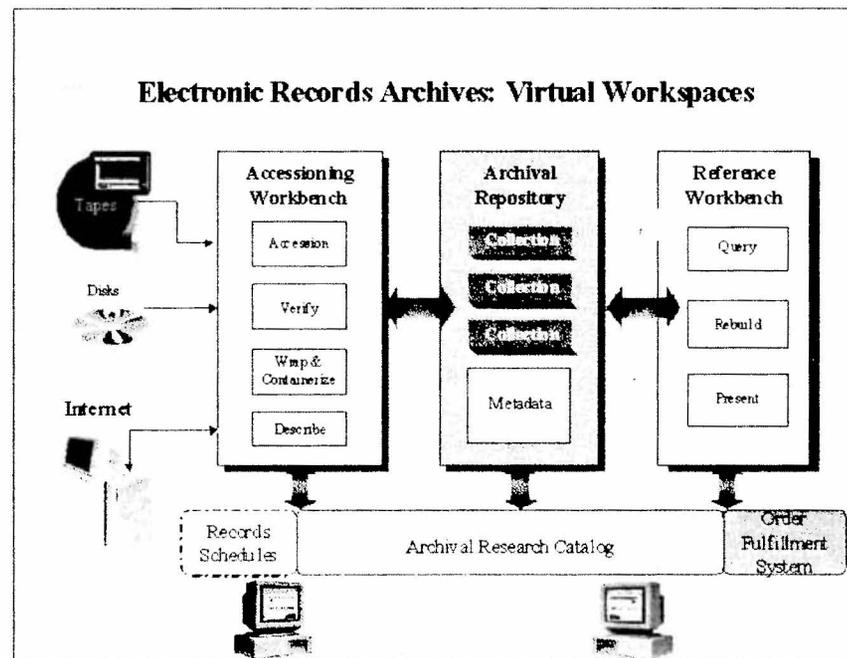

Figure 1

Figure 2

What will a solution built in this manner look like? As indicated above, the concept of the Electronic Records Archives is based on the Open Archival Information System reference model. The OAIS model defines a general framework for any system designed to preserve information assets over time. It assumes that the information is produced outside of the system and is intended for subsequent delivery to users or customers who are also outside of the system.

Internally, an OAIS has three basic functions: ingest, which brings information packages into the system, storage which maintains them over time, and dissemination, which supports queries and delivery of information to users. In the ERA concept, we envision executing these functions in three virtual workspaces. As shown in *figure 2*, the first virtual workspace, the Accessioning Workbench, is where sets of records will be brought into the archives. The second workspace is the Archival Repository where sets of records are kept over time. The third is the Reference Workbench where researchers' queries are processed and responsive sets of records are reassembled and presented. The processes occurring in these virtual workspaces are those described by Moore et al[12].

There are four critical properties of the virtual workspace concept. First, each workspace will be designed to have built-in capability for the business processes that need to occur there regularly. For example, the Accessioning Workbench will have functionality to verify that any sender who sends records to ERA for preservation in the National Archives is authorized to send such records (part of the "Accession" function shown in *figure 2*). Similarly, the Accessioning Workbench must be able to prepare sets of records for storage in the Repository (shown as "Wrap and Containerize"), and the Reference Workbench must be able to recreate the structure of any set of records retrieved from the Repository and place the records in their proper order in that structure (depicted as "Rebuild"). Second, each workspace needs to be designed to facilitate the application of special purpose tools when needed. A simple example (shown on the left in *figure 2*) is the need to accept input on various media. The National Archives receives electronic records on a variety of media. Given that there may be a significant lapse between the time when the agency wrote the files and the transfer of those media, NARA often needs to read obsolete media. The hardware and software needed to read any given medium constitute a tool set. Other types of tools might be much more complex. For example, natural language processing capabilities might be used to identify sensitive information that should not be disclosed when a record is released. The basic concepts behind tool sets are that they provide capabilities not needed most of the time, and that it should be easy to apply them when needed,

discard them when no longer needed, or replace them when better tools become available. The third key property of the virtual workspaces is that they are loosely connected through middleware such as software mediators or application programming interfaces. As described in the ERA program[13], this makes the overall system relatively independent of the particular information technology used in it at any time. If a hardware or software component in one workspace is replaced, the functionality of the system as a whole is maintained by modifying the middleware that enables interoperation between workspaces. It would not be necessary to change anything within another workspace. The fourth key property is that the virtual workspaces are defined in terms of functionality: the work performed in each space. This does not necessarily entail differences in the technologies used to implement the required functionality. In fact, it is assumed that all three virtual workspaces will share a common set of enabling technologies. For example, all three will need storage and data management capabilities such as those embodied in the Storage Resource Broker and Extensible Metadata Catalog developed by the San Diego Supercomputer Center.

The 'glue' that will hold together all of the ERA virtual workspaces is the Persistent Object Preservation architecture being developed by the National Partnership for Advanced Computational Infrastructure. The 'objects' that can be preserved under this approach can be any digital information that needs to be preserved. For archives, this ranges from individual records, to files of records, entire series of files, and ultimately to an entire archival fonds; that is, the totality of records created by a person or organization. Key to the ability of the persistent object approach is that it handles in a consistent fashion any arbitrarily complex object at any level of an arbitrary structure. The essential process is to transform the object to a persistent form. This entails identifying and characterizing all significant properties of the objects that are to be preserved. These properties are expressed in formal models. For example, individual records are modeled according to XML Document Type Definitions (DTDS). The appearance of the records can be captured through eXtensible Style Sheets. Files and other aggregations of records may be modeled as DTDS or as XML schemas[14]. Complex collections of records can be modeled using XML Topic Maps, which add semantic meanings to the syntactic information captured in DTSS and schemas[15]. The records are transformed by tagging or encapsulating them in metadata defined in the applicable models, eliminating other technical characteristics that are proprietary, dependent on specific hardware or software, or otherwise subject to obsolescence. But the persistent archives approach also provides the additional possibility of simply wrapping objects in their native formats in metadata that identifies them and characterizes them. This leaves open the possibility of processing the objects with the original software, as long as it remains available and operative. It would also enable

[12] R. Moore *et al.*, *Collection-Based Persistent Digital Archives*, cit.

[13] Electronic Records Archives Program. http://www.nara.gov/era.
[14] Extensible Markup Language. http://www.w3.org/XML.
[15] ISO Topic Maps Standard. http://www.dlib.org/org/dlib/february01/thibodeau/ibodeau.html.

The Electronic Records Archives
Tiered Structure

Special Tool Sets

↑ Responsiveness ↑ Adaptability ↑

Archival Functions

↑ Persistence ↑ Comprehensiveness ↑ Assurance ↑

Enabling Technologies

Infrastructure Independence
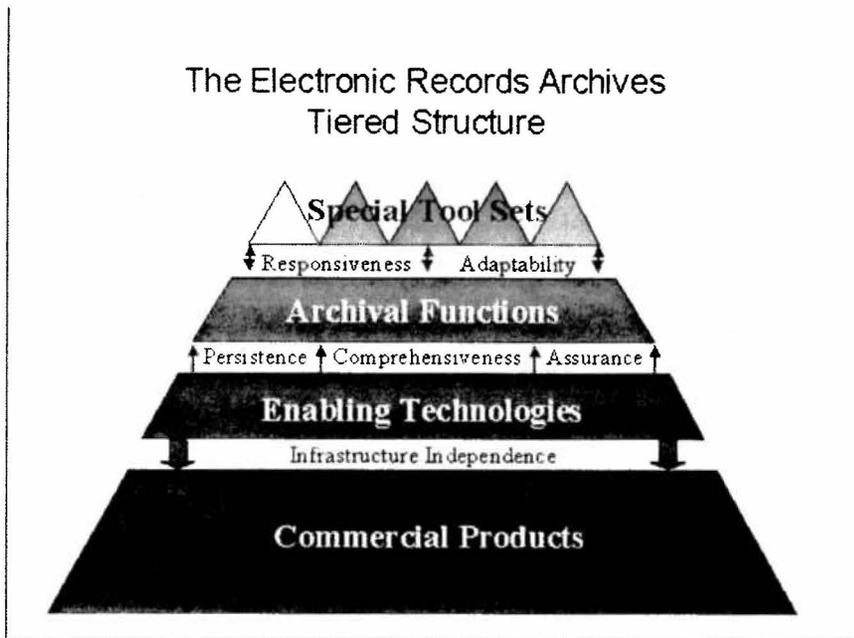
Commercial Products

Figure 3

subsequent transformation of the objects using newer methods not available when the objects are originally ingested into the Archival Repository.

Following the persistent object approach, collections are not maintained in the Archival Repository as structured sets. Rather, the members of a set are retained, along with models and metadata that define the structures of the set and the position of each member in that structure. When access to the information is desired, the set model is used to build the appropriate structure using current technology, and the members are placed appropriately in the materialized structure. Thus, potentially, a collection of records could sit in an Archival Repository for many decades without the archives having to take any specific action unless it needs to provide users with access to the records. In order to counteract obsolescence that will inevitably occur, repeatedly, across decades, all the archives needs to do is to update the software mediators it uses to translate the models and metadata stored in the repository into forms that current technologies can interpret. It does not need to update the models or the metadata, or the collections of records they describe.

Figure 3 provides a structural view of an ERA system, with the functional compo-

nents described above, built according to the developmental strategy previously outlined. The bottom tier of the structure is composed of commercial products. NARA intends to use hardware and software available in the market place as much as possible. Among others, products will be needed to satisfy basic requirements for high speed, high bandwidth communications; scalable, high-assurance, distributed processing; high-volume distributed, redundant storage; and security that is effective in a distributed environment.

However, ERA will need to be independent from the technology used. There are two reasons for this: first to ensure the persistence of the records being preserved, and second to take advantage of progress in commercial products. Infrastructure independence will be achieved through products labeled, at the second tier of the diagram, as "Enabling Technologies." Enabling Technologies include transformation methods for persistence of objects and collections of objects, storage management that makes the persistence of collections independent of the storage systems used, metadata management capable of handling all preservation and delivery requirements, and mediation methods to enable retrieved objects to be delivered to target technology.

The next higher tier in *figure 3* is labeled, "Archival Functions." This is the level of the virtual workspaces. The Enabling Technologies will ensure that the archival functions are performed in a coherent manner across all functions and all types of records and collections of records, providing a high level of assurance of both the persistence of the archives and the soundness and thoroughness of the processing. At the top of the structure are the archival "Tool Sets" that are used to address special processing needs. The Tool Sets enable the Archival Functions level to be optimized for regular work, while the system as a whole can be responsive to special needs and adaptable to new requirements and opportunities.

At this time ERA is essentially a vision. It is a vision for building a comprehensive, trustworthy means for addressing what is not only a moving target, but one which is rapidly growing both quantitatively and in complexity, and along paths that are not wholly predictable. In spite of the enormity of the challenge, there are substantial reasons for optimism that the vision will be realized:

- NARA has been able to form productive partnerships with other agencies, other governments, private business and academia to address the challenge.
- These partnerships are bringing together expanding numbers of world-class experts in a wide range of disciplines — including archival science, computer science, electronics engineering, chemistry, information science, and library science — to address the challenge.
- This work is leveraging much larger investments being made to develop the next generation information infrastructure for electronic commerce, electronic government, and research itself.
- The work is being widely publicized and subject to review, including in venues that employ rigorous peer review.

Finally, there is an empirical basis for optimism in that, in the research and development work that NARA is sponsoring, NARA demands empirical confirmation of research results. While much more work remains to be done, the research conducted to date has been validated using a variety of collections that approximate the diversity of historical materials that NARA is responsible for preserving. NARA provided some of these collections from its own holdings. Others were provided by our partners. They include the Department of Defense's Combat Area Casualties Current File from the Vietnam War, and its Gulf War web site; two million legacy patent application case files from the U.S. Patent and Trademark Office; the 1997 Vote Archive Demo of Roll Call Votes from the House of Representatives and the Senate Legislative Activity database from the U.S. Senate; a collection of one million e-mail messages from the Internet; TIGER/Line files from the Bureau of the Census; Digital Line Graph data from the U.S. Geological Survey; the contents of PC hard drives from the White House under former President Bush; the World Wide Web site of the Franklin D. Roosevelt Presidential Library; and the Art Museum Image Consortium collection of digital images from the California Digital Library.

Seamus ROSS [1]

Cyberculture, cultural asset management, and ethnohistory. Preserving the process and understanding the past

Abstract: *il saggio affronta il problema complesso della conservazione del materiale presente sul web, con particolare riferimento alla difficoltà di documentare in modo stabile oggetti di natura dinamica. La fluidità di tutte le componenti dell'ambiente cyber e delle pratiche degli utenti rendono indispensabile individuare una rigorosa metodologia di selezione che permetta di conservare nel lungo periodo informazioni e documenti autentici e affidabili, ma anche i necessari elementi di contesto.*

Analizzando alcuni progetti in corso di sviluppo sull'archiviazione del web, l'Autore conclude con l'affermazione che solo un approccio basato su responsabilità e conoscenze distribuite, che includa il coinvolgimento degli istituti di conservazione (biblioteche, archivi, musei) e dei soggetti produttori, potrà affrontare efficacemente i problemi della conservazione.

Introduction

As a child I spent three months each summer in a rustic seaside village without a television. I filled many of those TV-less evenings by reading articles from a battered, but much loved, copy of the twenty-nine volume 11th edition of the *Encyclopaedia Britannica* that my parents had bought for a couple of dollars at a yard sale. There seemed to be so much knowledge that had been gleaned from experimentation, the study of natural phenomena, the analysis of historical sources, and the investigation of remains of the past. Each article synthesised the work of many scholars and reflected many different methods of study. They all seemed to depend on or be linked in some way to many other articles. In the many decades that have passed since the 11th edition was produced our knowledge has exploded and indeed many of the ideas expressed in that edition – which could be described as a statement of human knowledge as it was then known – have been refined, revised, or even rejected. But the 11th edition had all the hallmarks that we expect from an information resource:

[1] See http://www.hatii.arts.gla.ac.uk.