Preservation Task Force Report

Task Force Members

Kenneth Thibodeau, U.S. National Archives and Records Administration (Chair)
Richard Blake, Public Records Office, U.K.
Paola Caruci, National Archives of Italy
Robert Chadduck, U.S. National Archives and Records Administration
Michele Cloonan, University of California, Los Angeles
Babak Hamidzadeh, University of British Columbia
P.C. Hariharan, Johns Hopkins University
Ross Harvey, Curtin University of Technology
Hans Hofman, National Archives of The Netherlands
Torbjörn Hörnfeldt, National Archives of Sweden
Richard Lysakowski, CENSA
Reagan Moore, San Diego Supercomputer Center
Christine Petillat, National Archives of France
William Rhind, CENSA
William Underwood, Georgia Tech Research Institute
Bruce Walton, National Archives of Canada

Introduction

InterPARES Domain 3, Methodologies for Preserving Authentic Electronic Records, focused on the topic of preservation, and the InterPARES Project established the Preservation Task Force to address questions in this area of investigation. The original InterPARES research plan stated, "The goal of the research in this domain is to identify and develop the procedures and resources required for the implementation of the conceptual requirements and the criteria identified in the first two domains."[1] The research plan articulated the following research questions for Domain 3:

- What methods, procedures, and rules of long-term preservation are in use or being developed?
    a) Which of these meet the conceptual requirements for authenticity identified in Domain 1?
    b) Which methods of long-term preservation need to be developed?
    c) Which of these methods are required or subject to standards, regulations, and guidelines in specific industry or institutional settings?

- What are the procedural methods of authentication for preserved electronic records?
    a) In what way can archival description be a method of authentication for electronic records?
    b) In what way can appraisal and acquisition/accession reports be constructed to allow for the authentication of electronic records?
    c) What are the procedures for certifying electronic records when they cross technical boundaries (e.g., refreshing, copying, migrating) to preserve their authenticity?

- What are the technical methods of authentication for preserved electronic records?
- What are the principles and criteria for media and storage management that are required for the preservation of authentic electronic records?
- What are the responsibilities for the long-term preservation of authentic electronic records?

Although the Preservation Task Force was charged with addressing these questions, several of the questions presume knowledge that would only be developed by other InterPARES groups. All five questions relate to authenticity and authentication, and their answers depend specifically on the articulation of these concepts and related requirements by the Authenticity Task Force. Part (b) of the second question depends in part on the work of the Appraisal Task Force. However, the Preservation Task Force could not delay starting its work until the products of the other task forces were finished, or it would not have been able to address most of these questions within the time frame of the project.

The task force therefore proceeded to address the issues of concern in the original questions, rather than literally to answer the questions as originally formulated. Pending results of the Authenticity and Appraisal Task Forces, the Preservation Task Force proceeded along lines that are essentially independent of those products. On the one hand it gathered empirical data about existing programs, plans, and technologies for preserving electronic records. On the other, it undertook a structured analysis of the process of preserving electronic records. In the empirical domain, the task force conducted a survey of programs that are preserving, are planning to preserve, or are conducting research related to the preservation of electronic records, and developed a white paper on media for storage of digital information. It also collected information on methods of authentication being used in organizations that are participating in InterPARES and preserving electronic records; however, in the absence of criteria for authentication, these data could not be analyzed.

---

[1] InterPARES Project, *Research Plan*, available at <http://www.interpares.org/researchplan.htm>.

In the analytic realm, the task force addressed the preservation of authentic electronic records by recasting the research questions. The key question addressed was:

- What activities are necessary to preserve electronic records?
  a) What are the inputs to this process?
  b) What controls govern the process?
  c) What does the process produce; that is, what are its outputs?
  d) What resources or mechanisms are necessary to carry out the process of preserving electronic records?

In addressing the above question and its sub-questions, the task force explored related issues:

- How do requirements for authenticity impact the preservation process?
  a) How is compliance with requirements for authenticity demonstrated?
  b) How can technological methods be evaluated in light of requirements for authenticity?

- How does the appraisal of electronic records impact on the preservation process?
  a) How does preservation impact appraisal?

The survey also addressed the original Domain 3 opening questions and asked three additional questions as a result of initial responses:

- What is the meaning of *preservation*?
  a) Does the meaning change when it is applied to electronic rather than paper-based records?

- Will current strategies for preserving electronic records ensure longevity and authenticity?
- How are costs for the preservation of electronic records derived? Have effective cost models been developed?

Although the Authenticity Task Force articulated requirements, they were not received by the Preservation Task Force in time to incorporate them fully in its products. Nonetheless, this task force has produced, in the formal model it developed of the function of preserving electronic records, a framework in which the requirements for authenticity and authentication can be applied. In essence, this model is neutral with respect to these requirements; that is, the model includes "place-holders" where the requirements could be introduced. In an initial review of the requirements for authenticity articulated by the Authenticity Task Force, the Preservation Task Force determined that no substantial revision of the model is needed to accommodate these requirements.

The Preservation Task Force review of the model of the process of selecting electronic records for preservation, produced by the Appraisal Task Force, showed that the two models are not incompatible, but some adjustments and clarifications are needed to align them so that they can readily be used together.

Research Design and Methodology

Survey Design

A survey, rather than a case study, method was selected because it was too early in the development of long-term retention strategies of most of the target respondents to study individual programs in depth. Given a target group of respondents consisting of fifteen sites known to be developing one or more of the techniques for digital preservation, the survey did not warrant a quantitative research design. Rather, the survey adopted a purposive sampling strategy—one that would show different perspectives on the problems we wanted to address—of archives, projects, and programs in the United States, Canada, and Europe. The investigators

developed a questionnaire with feedback from other members of the task force. The broad, sometimes open-ended questions further justified a qualitative design.

The target institutions were sent a consent letter explaining that, if they volunteered to participate in this study, they agreed to read the survey instrument, which was attached to the letter, and participate in an interview based on this instrument. Representatives from thirteen of the selected fifteen sites were ultimately interviewed. Some interviews were conducted in person, others by telephone.

Modelling Method

The Preservation Task Force developed a functional model of the process of preserving authentic electronic records following the Integrated Definition (IDEF) method prescribed by the InterPARES International Team. Specifically, it used IDEF(0) to describe processes or functions involved in preserving electronic records. In IDEF(0), "A function model is a structured representation of the functions, activities or processes within the modeled system or subject area."[2] An IDEF(0) model includes activities and entities. An activity is depicted as a box whose name indicates the nature of the activity. An entity either goes into or comes out of a process (activity). Three types of entities go into a process: inputs (I) that are transformed or consumed in the process, controls (C) that govern its execution, and the mechanisms (M) needed to carry it out. Only one type of entity comes out of a process: the outputs (O) that are produced by acting on the inputs under conditions and constraints imposed by the controls. In IDEF(0) diagrams, the four types of entities are always depicted as arrows in the following arrangement: Inputs enter a process box at the left side. Controls enter at the top. Outputs exit from the right, and mechanisms enter at the bottom. Given this invariant order, the entity arrows are collectively referred to as ICOMs.[3]

In IDEF(0) diagrams, there are two basic icons: boxes are used to represent activities or processes and arrows represent ICOMs. In IDEF(0), a process may be decomposed into its sub-processes. This is depicted by creating a new, child diagram in which the parent process box becomes the outer boundary of the diagram and the sub-processes are depicted as boxes within that diagram. All ICOMs connected to a box at a higher level are shown entering or exiting at the corresponding edge of the decomposition diagram. Successive decompositions can be delineated to achieve whatever level of precision or clarity is desired. Such successive decompositions constitute a decomposition hierarchy. All IDEF(0) models start at the highest level, labelled "A0," showing only one process box, which is the function being described taken as a whole, and the ICOMs that enter the function from the outside and that are output from the function. This simple notation provides a systematic and highly coherent method for describing a process to whatever degree of granularity is needed.

The boundaries of the preservation model derive from the viewpoint according to which the model is constructed. The IDEF(0) models "functions (actions, processes, operations), functional relationships, and the data and objects," according to the National Institute of Standards and Technology. The relationships between functions are logical, and not necessarily chronological. IDEF(0) does not explicitly model temporal sequences. Moreover, in IDEF(0),

> The viewpoint determines what can be "seen" within the model context, and from what perspective or "slant". Depending on the audience, different statements of viewpoint may be adopted that emphasize different aspects of the subject. Things that are

---

[2] National Institute of Standards and Technology (NIST), INTEGRATION DEFINITION FOR FUNCTION MODELING (IDEF0), Draft Federal Information Processing Standards Publication 183, 21 December 1993.

[3].Robert P. Hanrahan, "The IDEF Process Modeling Methodology" (1995), available at <http://www.stsc.hill.af.mil/CrossTalk/1995/jun/IDEF.asp>.

important in one viewpoint may not even appear in a model presented from another viewpoint of the same subject.[4]

The horizon for the viewpoint of the preservation model is determined by the scope of the InterPARES Project as whole. The project is concerned with the preservation of electronic records that have been selected for preservation after they are no longer needed for the practical purposes for which they were originally created. Therefore, the process described in the "Preserve Electronic Records" model begins with the transfer of the records from their creator, or from an agent acting for the creator, to a person whose primary responsibility is that of preserving authentic records—that is, the preserver. However, the preserver, as defined by the InterPARES project, has responsibilities that are broader than the preservation process itself. For example, the preserver is presumed to be responsible for selecting the records that are to be preserved. In the "Preserve Electronic Records" process model, the viewpoint is literally and strictly that of "the person responsible for preservation." The model's viewpoint includes only those entities and processes that someone, or some organization, carrying out the role of preserving the records would carry out. The same person or organization may have other roles or other, coincidental responsibilities, such as appraisal or reference, but coincidental responsibilities are excluded from the "Preserve Electronic Records" model. The role of preserving records includes all and only those activities necessary to ensure the transmission of authentic electronic records over time.

The "Preserve Electronic Records" model is intentionally generic. It identifies and describes the processes necessary to preserve electronic records, articulates the inputs needed by each process, the controls under which it operates, the mechanisms necessary to accomplish the process, and the output(s) produced by each process. The model defines the relationships among these entities and processes. While the model is systematic, it does not prescribe an implementation. Rather than defining a preservation system, the "Preserve Electronic Records" model provides a comprehensive, precise, and coherent road map that institutions and persons concerned with the preservation of electronic records can use in designing, developing, and evaluating systems which address their specific requirements, objectives, and constraints.

Data

The multidisciplinary, expert knowledge of the members of the Preservation Task Force was the principal basis for elaborating the "Preserve Electronic Records" model. In addition, task force members from institutions that are actively engaged in the preservation of electronic records supplied information about their institutions' practices, and plans. Two walkthroughs were conducted to test and refine the model using empirical data. The first walkthrough was of a hypothetical system defined by combining elements from several actual systems supporting related business functions, including two legacy systems used in different organizations, and a third system that is being deployed in one of these organizations. In order to provide a rigorous test of the model, the hypothetical system was made more complex by the addition of several data types. This walkthrough was used to test version 4 of the model and refine it to produce version 5. The second walkthrough was of a single empirical case. It was applied to version 5 and used to derive version 6. This walkthrough is described in an appendix.[5]

Findings

Survey Findings

Responses to the survey of digital preservation practices, plans, and research indicate three broad themes. First, the perception of what preservation is goes beyond archival and library

---

[4] NIST, 1993.

[5] Appendix 7.

practice to the media being preserved. Traditional definitions of preservation may not apply in the digital arena, and a shift is already apparent. Second, the rush to develop the technological processes necessary to preserve authentic electronic records appears to be at the expense of directly addressing cost and policy issues at the start of projects. The problems posed by permanent preservation of authentic electronic records require the development of a unique cost model. Last, the paucity of preservation policies in place represents a distinct gap in the research design of many of the projects and possibly reflects a lack of commitment among the stakeholders in institutions. It appears that meeting the technological challenges of preserving electronic records is more of a priority within these institutions than developing policy. Such prioritization courts the risk that overall progress in this new arena will be more uneven than is necessary.

Results of Analytic Modelling

The process model "Preserve Electronic Records," developed by the Preservation Task Force, took an approach contrary to that discovered by the survey in many projects. Rather than giving priority to the technological challenges of preserving electronic records, the task force developed a model in which alternative technical solutions can be evaluated and adopted as appropriate. The model is motivated by the perception that, while preserving electronic records requires technological solutions, it is impossible to determine whether any given technology constitutes a solution on technological grounds. The criteria for evaluating technologies derive from the archival and institutional requirements that determine the goals and objectives of preservation and act as controls on the preservation process. Technology plays a role in selecting solutions only in that any solution must be feasible: the technology to implement the solution must exist and be applicable. Feasibility also includes affordability. While the "Preserve Electronic Records" model does not include or assume a cost model, it provides for the application of a cost model in developing preservation strategies and plans and in evaluating their execution.

The most fundamental finding that emerged from the structured analysis of the process of preserving electronic records was a paradigmatic shift in the concept of preservation of electronic records. This shift had both archival and technological dimensions. While the phrase "preserve an electronic record" is convenient and undoubtedly will continue to be used, in many variations it is a shorthand expression that belies reality. Empirically, it is not possible to preserve an electronic record: it is only possible to preserve the ability to reproduce the record. That is because it is not possible to store an electronic record in the documentary form in which it is capable of serving as a record. There is inevitably a substantial difference between the digital representation of the record in storage and the form in which it is presented for use.[6] It is always necessary to use some software to translate the stored digital bits into the documentary form of the record. This entails an inevitable risk that, regardless of how well the digital data were protected in storage, the record may be inappropriately altered when the stored bits are retrieved and presented for use as a record. Thus, in contrast to prevailing notions about the preservation of records in hard copy, the process of preserving an electronic record goes well beyond keeping it safely in storage. The process of digital preservation begins with the initial act of storage and extends through reproduction of the record. To reflect the empirical situation, the task force constructed the concept of a "digital component of an electronic record." An electronic record is stored as one or more digital components. Digital components have no necessary relation to the elements of documentary form recognized in diplomatics analysis of records. Rather, they are determined technologically by the way the bits are stored and by the methods (software) that must be applied

---

[6] With continuing progress in digital information technology we can expect to reach the point where computers can input information recorded in human-readable form. Nonetheless, this assertion will remain valid: while the digital display of a record—for example, narrative text recorded on paper—may preserve the "look and feel" of the paper version, the digital version will be inscribed on a different physical medium and the process of producing the display version from the stored version may result in alteration of the record.

to reproduce the record. Reproducing an electronic record entails (1) reconstituting it, that is, reassembling its digital components if it has more than one, or extracting any digital component stored in a physical file that contains more than one such component; and (2) presenting it in proper form.

<u>The Process of Preserving Electronic Records</u>

At the highest level, three factors control the process of preserving electronic records. First, in order to preserve records, and especially to preserve them as authentic, we need to know what the requirements are for doing so. These requirements are archival in nature: they derive from archival science and principles and from related standards and best practices for managing records. Second, preserving electronic records entails using digital information technology; the possibilities for such preservation are limited by the state of the art of information technology, which constitutes the second type of control on the preservation process. Third, the exercise of the preservation function will also be governed by requirements of the institution in which this function is carried out.

Three mechanisms are necessary to perform the preservation process: an information and communications technology infrastructure; facilities where the electronic records will be stored and processed; and persons competent to carry out the process. While the state of the art of technology determines what is possible and impossible to do, the technology infrastructure comprises the hardware, general-purpose software, and physical media used to store and process the digital components of electronic records. These mechanisms are used in all preservation activities.

There are two primary inputs to the process of preserving electronic records. The first, and most obvious, are transfers of electronic records selected for preservation. In simple terms, the records are what the process is all about. Records are preserved because they have been determined to have enduring value. That value is realized in use. So the second primary input consists of requests for the records, or for information about them. The preservation process also needs a third input, information about the records that have been selected for preservation. This information is necessary to determine what preservation methods, information technology, facilities and staff will be needed to preserve the records and to organize the process to guarantee that the records can be preserved as authentic.

The preservation process produces two primary outputs: reproduced electronic records and reproducible electronic records. A preservation system outputs a reproduced electronic record when the record is reconstituted and presented within the system itself. However, in many cases those who want to access a preserved record will want to do so on systems outside of the preservation system, such as in web browsers on their own computers. In such cases, the preservation system can output only the digital component(s) of the record along with instructions on how to reconstitute and present the record; that is, it outputs a reproducible electronic record. There are two other, derived, outputs of the preservation process: a certificate of authenticity, which attests to the authenticity of a reproduced record; and information about preservation, which attests to the integrity and reliability of the preservation process overall. A certificate of authenticity is produced when a requester demands tangible evidence that a reproduced electronic record is authentic. Information about the overall system and processes of preservation is produced either as required by higher authorities or in response to a challenge to their adequacy or appropriateness for preserving authentic records.

The process of preserving electronic records includes four principal sub-processes: a management process and three processes that carry out or execute preservation. The management process governs the other three. It establishes a comprehensive approach, which is executed in the three other processes, and it evaluates these processes to ensure that the goals and objectives of preservation are achieved. To do this, the management process interprets the external archival and institutional controls into a coherent synthesis of requirements, or preservation framework, which controls other management sub-processes, as well as the execution processes.

In each case where the appraisal process identifies a body of records as worthy of preservation, the preservation management process determines whether it is possible to preserve the records— given the technical characteristics of the records and the state of the art of information technology—and, if so, how they will be preserved. This determination feeds back into the appraisal process, enabling a two-part decision that the body of records both has enduring value and can be preserved. For each body of records thus selected for preservation, managing the preservation process requires articulating a preservation strategy.

The preservation strategy encompasses a set of rules or procedures for processing the records in each of the execution activities; criteria for determining whether each process defined in the strategy is executed properly and achieves its desired outcome; and specific technological methods that will be used to preserve the records, up to and including their reproduction. The preservation strategy acts as a control on the execution processes. A preservation strategy will entail requirements for specific information technology infrastructure needed in order to implement the strategy. Preservation management thus includes sub-processes to identify and acquire the technological infrastructure and the technological preservation methods that will be used in preserving the records. The technological preservation methods used to preserve the records are specific to classes of digital components and control the processing and maintenance of those components over time and the reproduction of records from the components; the technological infrastructure enables these methods to be executed. For example, a preservation strategy might prescribe that, in the case of textual records whose visual appearance is critical for authenticity, the records will be preserved as bitmapped images. Preservation methods used to implement such a strategy would include software to convert textual records from other formats, such as word-processing files, to bitmaps and software to render such bitmaps for viewing. The technological infrastructure needed to execute the software would include appropriate processors, storage devices, display devices, and drivers. Basically, preservation methods directly support the preservation and reproduction of electronic records from their digital components, and the preservation infrastructure supports the execution of preservation methods.

A preservation strategy also defines specific actions that should be taken with respect to the body of records, either at specified times, such as when the records are first brought into the preservation system, or under specific conditions, such as when the media on which the digital components of the records are stored need to be replaced. For example, a preservation strategy of preserving textual records as bitmapped images would entail converting textual records transferred in different formats to bitmaps. The same strategy should also specify what to do if the software used to display the records stored in bitmaps becomes obsolete. The preservation strategy remains constant, unless there is a management decision to change it.

Each execution process must produce information about itself and about the results achieved in its execution that is appropriate and adequate to enable management to determine whether a preservation strategy is successful and, if not, what corrective action is needed.

The reports of the Authenticity and Appraisal Task Forces give rise to different types of findings, specifically highlighting the need to review the work of the three task forces with the objective of synthesizing results where appropriate and of identifying where additional analysis is needed to align the products.

One such undertaking would be to develop a third IDEF(0) model that links those of the Appraisal and Preservation Task Forces. The Appraisal Task Force's model may be described as constructed from the viewpoint of the preserver exercising the role of selecting electronic records for preservation, while the Preservation Task Force's model may be described as constructed from the viewpoint of the preserver exercising the role of preserving electronic records. The third model would take the viewpoint of the preserver given its responsibility for coordinating both processes. This effort would not be merely academic. A substantial result that should be expected from constructing the third model would be articulation and clarification of the feedback loop between selection and preservation. Currently, when appraisal identifies a body of electronic records as having enduring value, information is needed about the feasibility of preserving the records. In the case where the preservation system has the capability and capacity to preserve

the records, confirmation of this fact may be all that is needed to reach a selection decision. However, most cases will require more extensive communication between the preservation and selection processes. For the records to be preserved successfully, the two processes must reach complete agreement on terms and conditions for transfer of the records from the active system to the preservation system or, alternatively, from the state of active or open records to the state of closed, inactive, preserved records. These terms and conditions prescribe the initial steps in the preservation strategy. Where the preservation system does not have the capability or capacity of preserving the records, there should be additional communication between the two processes concerning requirements, alternatives, costs, and other related factors. Furthermore, to develop an adequate preservation strategy for a body of records, preservation management will need information about the appraiser's benchmark assessment of authenticity as soon as it is available.

It does not appear that the "Preserve Electronic Records" model needs to be modified in any substantial way to accommodate the benchmark and baseline requirements produced by the Authenticity Task Force. The Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records do not apply to the records themselves nor to the preserver. Rather, they are criteria that the appraiser should use to assess authenticity when selecting records for preservation. The result of applying the benchmark requirements is information articulating a presumption of authenticity. The "Preserve Electronic Records" model provides for receipt and preservation of this information as part of the chain of preservation. This model also provides an opportunity for updating the assessment when records are examined as part of the process of bringing them into the preservation system. The Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records do apply to the records and to the preserver, but they are largely contextual in character. The "Preserve Electronic Records" model can satisfy these requirements as it stands, although it would probably be beneficial to make this more explicit.

Nonetheless, there are points that should be explored simultaneously from both authenticity and preservation perspectives. For example, the first baseline requirement requires that "the content of the record remains unchanged after reproduction." Given that this requirement applies to "transfer, maintenance, and reproduction," clearly the operative meaning of "unchanged" is with respect to the state of the content as delivered by the records creator. However, a variety of factors, such as the fragility of digital storage media, may result in some partial loss or corruption of content. The requirement should be enhanced by specifying that any such loss or corruption should be documented and, perhaps, by indicating when such problems would be critical.

The benchmark requirements include provision for documenting whether the creator established the documentary forms of records. The baseline requirements require documentation of "the impact of the reproduction process on their form." If the creator did not articulate documentary form, there is no obvious basis for determining the impact of reproduction. There is a large gap between these processes that needs to be addressed.

Products

The Preservation Task Force produced a detailed IDEF(0) model of the process of preserving electronic records, a report explaining basic concepts of the model and providing simplified views of the model, a case study illustrating application of the model, a report on the results of its survey of current digital preservation practices, and a report on digital storage media:

- IDEF(0) Model, "Preserve Electronic Records"
- How to Preserve Authentic Electronic Records
- Walkthrough Applying the "Preserve Electronic Records" Model.
- M. Cloonan and S. Sanett, "Preservation Strategies for Electronic Records," Round 1 (2000–2001). June 2001. <http://is.gseis.ucla.edu/us-interpares/CloonanSanettPreservationReport.pdf>

- P.C. Hariharan, "Media." <http://is.gseis.ucla.edu/us-interpares/Mediareport.pdf>
- W. Underwood, "Preserving Authentic and Reliable Electronic Records in JARS." June 2000. <http://is.gseis.ucla.edu/us-interpares/jars.pdf>

The first three of these are included as appendices.[7]

Relationship to Existing Standards

The Open Archival Information System Reference Model

The basis for the content of the preservation process model is the Open Archival Information System (OAIS) Reference Model, a new ISO standard that was being developed while the Preservation Task Force articulated the "Preserve Electronic Records" model.[8] "An OAIS is an archive, consisting of an organization of people and systems, that has accepted the responsibility to preserve information and make it available for a Designated Community." The "Preserve Electronic Records" model is built on the basic assumptions of the OAIS that the records are produced outside of the archival system, that they are to be available to a user community which is also outside of the archival system, and that the archival system is thus a mediator which takes information from producers and delivers it to users over long periods of time. Thus the OAIS model has a much broader scope than the "Preserve Electronic Records" model. The reference model is intended to apply to any type of information, not just records. For example, the information preserved in an OAIS might be scientific data, or it might be information about physical objects in a museum. At a high level, it may be said that the "Preserve Electronic Records" model is a specification of an OAIS for the specific classes of information objects comprising electronic records and archival aggregates of such records.

---

[7] Appendix 5, Appendix 6, and Appendix 7.

[8] Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System (OASIS), Red Book (May 1999,) <http://www.ccds.org/documents/pdf/CCDS-650.0-R-1.pdf>.
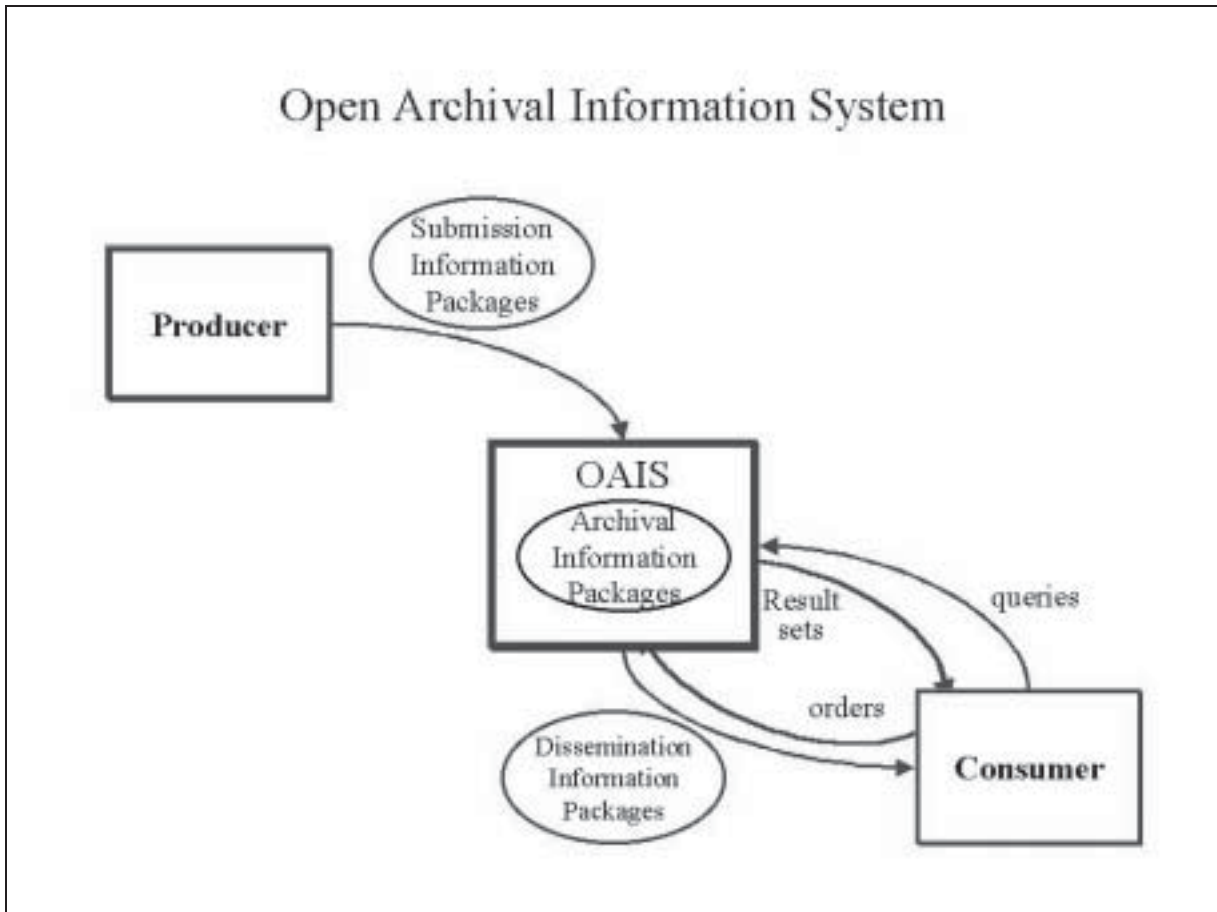
Figure 1. Open Archival Information System

It is necessary to distinguish between the function described by the "Preserve Electronic Records" model and a system that would implement the model. The preservation function might be carried out by a system that provides only the functionality described in the model. But it might equally well be implemented in a system that features additional functionality, including the appraisal of records, the management of current and temporary records, and reference and dissemination functions.

This reveals another aspect in which the "Preserve Electronic Records" model is narrower than the OAIS: the preservation model does not include all activities related to making records available—only those that are inextricable from the preservation function. The preservation function extends to the production of copies of records, because that is necessary to guarantee their authenticity, but it does not include order agreements as described in the OAIS model or any "value-added" dissemination or access services. Similarly, the preservation model does not include processes, which inform potential users what records are being preserved or what conditions govern access to the records.

While the "Preserve Electronic Records" model is narrower than the OAIS model, the InterPARES model has substantial more depth on the topic of preservation in general and, obviously, the preservation of authentic electronic records in particular. The Preservation Task Force communicated its work to the committee responsible for the OAIS standard and worked with that committee to enhance the standard in light of our findings.

Information Technology Standards

The "Preserve Electronic Records" model does not explicitly adopt or implicitly entail specific information technology standards, such as those concerning various digital data formats, storage media, information interchange, etc. Instead, the model provides a context, specifically in the preservation management process, for identifying, evaluating and adopting such standards as appropriate. One of the principal controls on the entire preservation process, the state of the art of information technology, includes current standards. Other things being equal, the preserver should develop preservation strategies that adopt those standards that best support archival objectives; however, other things often will not be equal. In archives of corporations and universities, for example, the information technology infrastructure may be largely determined by corporate information technology architecture, leaving the preserver no option but to develop strategies that can be implemented on that infrastructure. Other major factors that will constrain the preserver's adoption of standards include the costs of doing so and the availability of products that implement the standards as well as support for those products. For example, other things being equal, to achieve the archival goal of permanent preservation, the preserver would tend to select storage media that are subject to standards and are durable. However, in an environment of continuous change in information technology, the preserver needs to anticipate that the longer any given type of digital medium is kept, the more expensive and difficult it will be to maintain.

Conclusion

Much attention to the preservation of electronic records has focused on the twin problems of the relatively short life expectancy of digital media and the rapid obsolescence of hardware and software. The InterPARES project started with recognition of these problems and cast the preservation issue in terms of evaluating practical methods for solving them. The research plan called for the Preservation Task Force "to identify and develop the procedures and resources required for the implementation of the conceptual requirements [for preserving authentic electronic records] and criteria [for appraising electronic records] identified in the first two domains." This formulation of the problem of preserving electronic records clearly situates it not in technology, but in the interface between the goal of preserving electronic records and the technology on which they depend. Technology itself is not a problem. If we did not need to preserve records beyond the life expectancies of hardware, software, and digital media, we would not have any preservation problem. Similarly, technology cannot determine the solution. It is archival and records management requirements, that define the problem. It must be archival and records management criteria that determine the appropriateness and adequacy of any technical "solution." The question "What is the best technological method for preserving electronic records?" is as meaningless as the question "What is the best medicine for making people healthy?" Neither can be answered without specifying the conditions they are meant to address. The InterPARES project defined these conditions as the archival requirements for authenticity and the archival criteria for selecting records to be preserved.

As previously stated, because the InterPARES task forces on authenticity, appraisal and preservation worked in parallel, the Preservation Task Force could not formulate solutions based on specific conceptual requirements and criteria. Nonetheless, through communications and cross-fertilization among the task forces in the entire course of the research, the Preservation Task Force has been able to produce a model of the process of preserving electronic records that does in fact identify the *procedures* and *resources* needed to implement the requirements and criteria. The procedures are the processes defined in the "Preserve Electronic Records" model, and the resources include the mechanisms needed to carry out these processes as well as the information about both the processes and the records that needs to flow across processes. This model does not describe a computer system, and it does not itself reach conclusions about what technological systems, tools, or methods are best suited for preserving electronic records. Rather, it provides an extensive, detailed and highly coherent framework for identifying and analyzing the specific challenges faced in implementing appraisal decisions that select specific bodies of electronic records to be preserved. This framework guides the evaluation of

technological options and the articulation of specific preservation strategies addressing both the archival and technological characteristics of the records to ensure the continuing availability of authentic copies of the records across time and generations of technology.

Thus the "Preserve Electronic Records" model can be a guide to implementation, but it does not prescribe an implementation. There is greater value in this model than there would be in one that described how to design a particular preservation system. It would be simplistic, and erroneous, to assume that a single technical solution would be optimal in all circumstances. The "Preserve Electronic Records" model can be used to develop solutions that address varying circumstances, including not only diversity in the characteristics of the records to be preserved, but also variety in the external requirements imposed on the preserver, and in the goals and objectives to be achieved in preserving the records.

Recommendation 1. The primary recommendation that comes out of this work, then, is for analysts and institutions to use the "Preserve Electronic Records" model as a framework for developing solutions to the challenges of preserving electronic records.

Recommendation 2. Use of the "Preserve Electronic Records" model should be based on understanding of the particular characteristics of electronic records and what those characteristics entail for preserving these records, as summarized in the foundation concepts:

    a)   Digital Components of Electronic Records,
    b)   Preservation Control,
    c)   Archival Requirements for Preservation,
    d)   "Original" Electronic Records,
    e)   The Need to Reproduce Electronic Records, and
    f)   The Chain of Preservation.

These concepts are set out in Appendix 6, "How to Preserve Authentic Electronic Records." Key to all of these concepts is the recognition that the chain of preservation for electronic records must extend over their entire life and that the process of preserving electronic records extends to and includes reproducing them.

Recommendation 3. Solutions to the preservation of specific bodies of electronic records should be inherently dynamic. The solutions need to be dynamic for two different reasons. First, most archives and other preservers will accumulate electronic records over time. Over time, the specific properties of the records being brought into the archives will change. The preservation system must be capable of being expanded, adapted, or modified to accommodate new and different types of electronic records, and new ways of organizing, accessing, and presenting such records. Second, the goal of preserving electronic records is not to keep them, in archives or elsewhere, but to make them available to persons who have a need for, or an interest in, them. While the preserver has a fundamental responsibility for providing access to authentic records, their availability will be impacted by the continuing evolution of information technology. Preservers should assume that future users would want to use the best available technology for access to the records. The design of preservation systems should take into consideration the need to be able to interface with evolving technologies for information discovery, retrieval, communication and presentation.

Recommendation 4. The InterPARES Project has been so fruitful that it has gone far beyond providing valuable products in response to the research questions that it originally posed. It has also raised the threshold of research by articulating issues that are entailed by the original questions, but not explicit in them, by identifying new questions and by opening up lines of research that should provide grounds for valuable results for years to come. For example, the project has moved beyond its foundation in the science of diplomatics to recognize that, in the digital environment, many of the concepts and methods that traditionally were applied to individual documents need to be applied to sets of records. This insight needs to be explored more fully. The work of the Preservation Task Force has focused on defining a comprehensive framework for preserving authentic electronic records. This work should not stop when the current

project ends. The archival profession, our collaborators, and our stakeholders, have an interest and responsibility to see that further progress is made.

Much more work is needed to analyze the data and information requirements for executing the processes defined in the preservation model.

The model should also be applied to additional test cases both to validate and enrich it. The model should also be extended to address the application of specific technologies for overcoming technological obsolescence.

Methods must be developed for analysis and categorization of the documentary forms of electronic records and criteria for determining which elements or aspects of documentary form must be preserved to ensure the integrity of the record.

While the Authenticity Task Force found that it was not possible to construct a typology of electronic records from which requirements for authenticity could be derived, the concept of authenticity elucidated by that task force entails preserving documentary form. Benchmark Requirement 5, for assessing the authenticity of records, requires evidence that "the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator." Similarly, Baseline Requirement 2, for reproducing authentic copies of electronic records, entails documenting "the impact of the reproduction process on their form, content, accessibility and use."

There is a significant opportunity for the InterPARES project to contribute to the enrichment of the Open Archival Information System.

While the scope of the OAIS model extends far beyond the domain of records, that model could be informed by archival understanding of authenticity. Regardless of the nature of the information objects being preserved, those responsible for preserving them should be able to attest to and explain the authenticity of the products they deliver to their customers. Such a need is signalled by the concern in many disciplines of natural science with "data lineage" or "data parentage."

The accomplishments of the InterPARES Project should be applied to related areas of concern, such as the process of archival description.