InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Preservation Task Force Final Report

DRAFT FOR COMMENT

October 31, 2001

Task Force Members:

Kenneth Thibodeau, U.S. National Archives and Records Administration, chair,
Richard Blake, Public Records Office, UK,
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

Report of the InterPARES Preservation Task Force

DRAFT November 1, 2001

Abstract

Acknowledgements

This report was developed by work done by the InterPARES Preservation Task Force. The members of the Preservation Task Force are Kenneth Thibodeau, U.S. National Archives and Records Administration, chair, Richard Blake, Public Records Office, UK, 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, and Bruce Walton, National Archives of Canada. The Task Force acknowledges the contributions made by and the staff of the InterPARES Project, Tahra Fung, Jean-Pascal Morghese, and Peter VanGarderen, at the University of British Columbia, Kevin Glick and Richard Sloma of the American Research Team at the University at Albany, and the contributions from Lisa Beitel, Deirdre Bryden, Kathleen Burns, Robyn Hulley, Sion Romaine at University of British Columbia, Shelby Sannet of the University of California, Los Angeles, and Michael Skipper of the National Archives and Records Administration.

Research Questions

InterPARES Domain III, Methodologies for Preserving Authentic Electronic Records, focused on the topic of preservation. 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." [IP Plan] Domains I and II were concerned with authenticity and appraisal, respectively. The research plan articulated the following research questions for Domain III:

1. 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 I?
   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?

2. 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?
3. What are the technical methods of authentication for preserved electronic records?
4. What are the principles and criteria for media and storage management that are required for the preservation of authentic electronic records?
5. What are the responsibilities for the long-term preservation of authentic electronic records?

The InterPARES project established the Preservation Task Force (PTF) to address these questions. However, several of the questions presume knowledge that would only be developed by other InterPARES groups. All 5 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. Question 2.B 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.

Therefore, the PTF 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 PTF 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 PTF 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 in use in organizations which are participating in InterPARES and which are preserving electronic records; however, absent criteria for authentication, this data could not be analyzed.

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

1’. 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; i.e., what are its outputs?
   D. What resources or mechanisms are necessary to carry out the process of preserving electronic records?

In addressing question 1’, and its sub-questions, the PTF explored related issues:

2’. 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?
3’. How does the appraisal of electronic records impact on the preservation process?
   A. How does preservation impact appraisal?

In addition, the survey addressed the original Domain III question 1 and asked three additional questions as a result of initial responses:

- What is the meaning of *preservatio*n?
  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?

While the Authenticity Task Force articulated requirements, they were not received by the PTF in time to incorporate them fully in its products. Nonetheless, the PTF 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, the PTF determined that no substantial revision of the model is needed to accommodate these requirements.

PTF 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 15 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 PTF. The broad, sometimes open-ended questions further justified a qualitative design.

The target institutions were sent a consent letter which explained 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 13 of the selected 15 sites were ultimately interviewed. Some interviews were conducted in person, others by telephone.

Modeling 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 IDEF0 to describe processes or functions involved in preserving electronic records. In IDEF0, "A function model is a structured representation of the functions, activities or processes within the modeled system or subject area." [NIST 93] An IDEF0 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 processes: inputs (I) which are transformed or consumed in a process, controls (C) which 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) which are produced by acting on the inputs under conditions and constraints imposed by the controls. In IDEF0 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. [Hanrahan]

In IDEF0 diagrams, there are two basic icons: boxes are used to represent activities or processes and arrows represent ICOMs. The ICOMs are always arranged in the following manner: Inputs enter boxes from the left, Controls enter from the top, Outputs exit from the right, and Mechanisms enter from below. In IDEF0, a process may be

decomposed into its sub-processes. This is depicted by creating a new, child diagram in which the 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 IDEF0 models start at the highest level, labeled "A-0," 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. IDEF0 models "functions (actions, processes, operations), functional relationships, and the data and objects." The relationships between functions are logical, and not necessarily chronological. IDEF0 does not explicitly model temporal sequences. Moreover, in IDEF0,

> "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 important in one viewpoint may not even appear in a model presented from another viewpoint of the same subject." [NIST 93]

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, which 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. 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 roadmap which 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

{TBD}

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 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 preserving authentic electronic records permanently require the development of a unique cost model. Last, the lack of preservation policies in place is 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. This prioritization entails a risk that overall progress in this new arena will be more uneven than is necessary.

Results of Analytic Modeling

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 PTF 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.[1] 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 was 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 preservation begins with the initial act of storage and extends through reproduction of the record. To reflect the empirical situation, the PTF 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 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.

The Process of Preserving Electronic Records

At the highest level, three things 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 are the requirements for doing so. These requirements are archival

[1] It may be anticipated that with continuing progress in digital information technology we will 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.

in nature: they derive from archival science and principles and related standards and best practices for managing records. Second, preserving electronic records entails using digital information technology. The possibilities for doing so 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. They are 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 infrastructure, facilities and staff will be needed to preserve the records and to organize the process to guarantee that the records can be preserved 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 only output 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 which 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, which control 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 both a preservation strategy and one or more preservation action plans.

The preservation strategy encompasses a set of rules or procedures for processing the records in each of the execution activities, as well as criteria for determining whether each process defined in the strategy is executed properly and achieves its desired outcome, and specific technological methods which 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 which will be used in preserving the records. The difference between the technological infrastructure needed for preservation and the technological preservation methods used to preserve the records is that the preservation methods 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, while the 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 bit-mapped 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 bit-maps and software to render such bitmaps for viewing. The technological infrastructure needed to execute the software would include appropriate processors, storage devices, and display devices and drivers. Basically, preservation methods directly support the preservation and reproduction of electronic records from their digital components, while preservation infrastructure supports the execution of preservation methods.

A preservation action plan defines specific actions which 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 bit-mapped images would entail an action plan stipulating that, when textual records are transferred

in different formats, they should be converted to bit maps. The same strategy should also dictate an action plan specifying what to do if the software used to display the records stored in bit maps becomes obsolete. While the preservation strategy remains constant, unless there is a management decision to change it, steps in a preservation action plan are no longer action items once they are carried out. Therefore, a preservation action plan functions as an input to one or more preservation processes.

Each execution process must produce information about itself and the results achieved in its execution that is appropriate and adequate for management to determine whether a preservation strategy or action plan 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 IDEF0 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. 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, alternative, costs and other related factors. Furthermore, to develop an adequate preservation strategy and action plan(s) 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 this assessment when records are examined as part of the process of bringing them into the preservation system. The Baseline Requirements Supporting the Reproduction of Authentic 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 specifically 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 addressing 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…." There is a large gap between these processes that needs to be addressed.

Products

The Preservation Task Force produced a detailed IDEF0 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:

- IDEF0 Model, Preserve Electronic Records
- How to Preserve Authentic Electronic Records
- Case Study Applying the 'Preserve Electronic Records' model. (In progress)
- 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 to this report. The PTF also intends to produce a 'walk-through' describing how a specific body of electronic records would be preserved in accordance with the model.

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, which is currently a draft ISO standard. [CCSDS] “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.

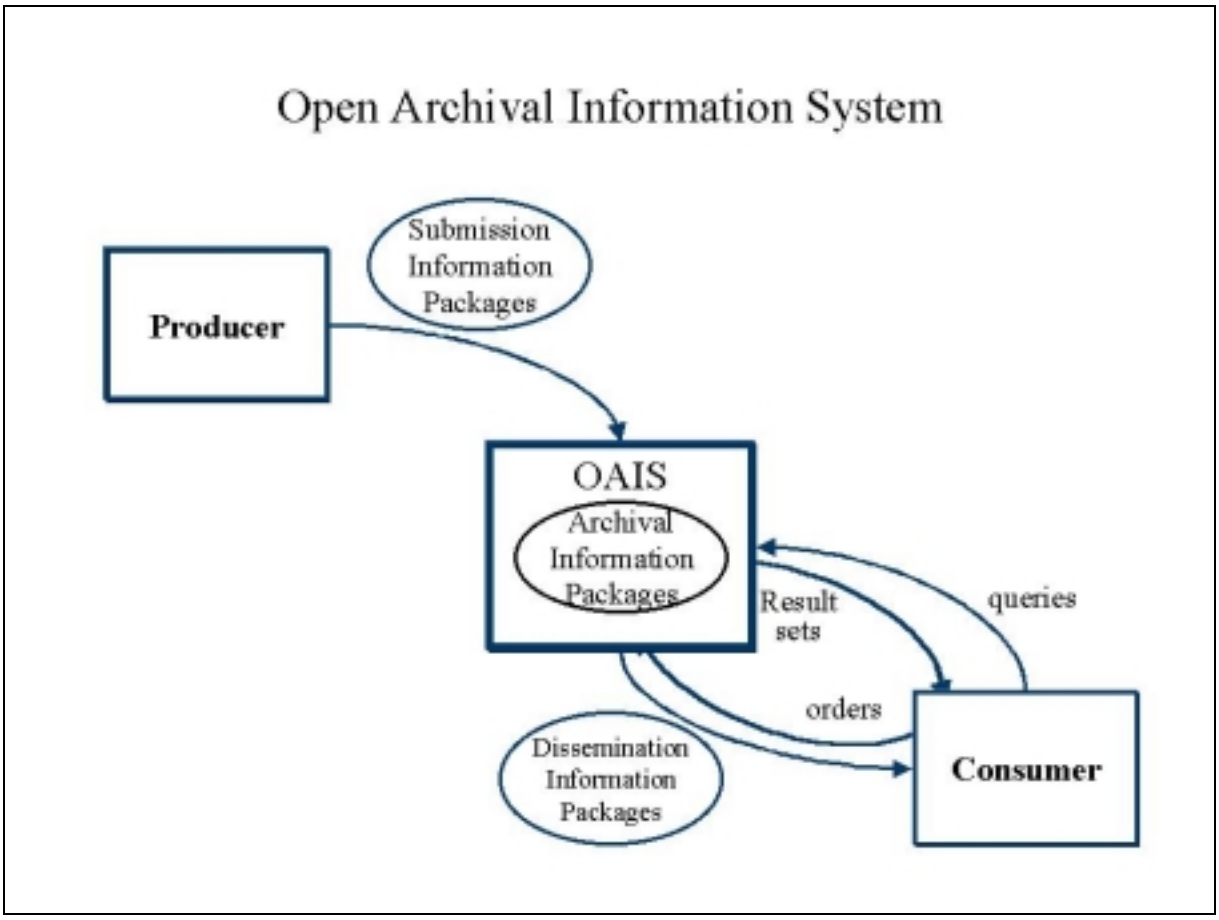


Figure 1. Open Archival Information System

Here again it is necessary to distinguish between the function described by the 'Preserve Electronic Records' model and a system which would implement that model. The preservation function might be carried out by a system which provides only the functionality described in the model. But it might equally well be implemented in a system which includes 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, but 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 has communicated its work to the committee developing the OAIS standard and is working 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 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 implementing the standards and of 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 that 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, which define the problem. It must be archival and records

management criteria which 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 which 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 both the mechanisms needed to carry out these processes and 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 which 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:

Digital Components of Electronic Records,
Preservation Control,
Archival Requirements for Preservation,
'Original' Electronic Records,
The Need to Reproduce Electronic Records, and
The Chain of Preservation,

These concepts are set out in chapter 2 of "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.

4.2 **Recommendation 4.** The InterPARES project has been so fruitful that it has not only provided valuable products in response to the research questions that it originally posed, but 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.

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

4.3 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.

4.4 Develop methods 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." {Emphasis added.}

4.5 There is a significant opportunity for the InterPARES project to contribute to the enrichment of the Open Archival Information System (OAIS) proposed as an ISO standard.

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 signaled by the concern in many disciplines of natural science with 'data lineage' or 'data parentage.'

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

References

[CCSDS] Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). Red Book. May 1999. http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf

[IP Plan] InterPARES Project. Research Plan. http://www.interpares.org/researchplan.htm

[Hanrahan] Robert P. Hanrahan, The IDEF Process Modeling Methodology. 1995. http://www.stsc.hill.af.mil/CrossTalk/1995/jun/IDEF.asp

[NIST 93] National Institute of Standards and Technology. INTEGRATION DEFINITION FOR FUNCTION MODELING (IDEF0). Draft Federal Information Processing Standards Publication 183. December 21 1993. 128 pp.

Appendices

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

---

Preservation Task Force Model Diagrams

Draft Version 5.1

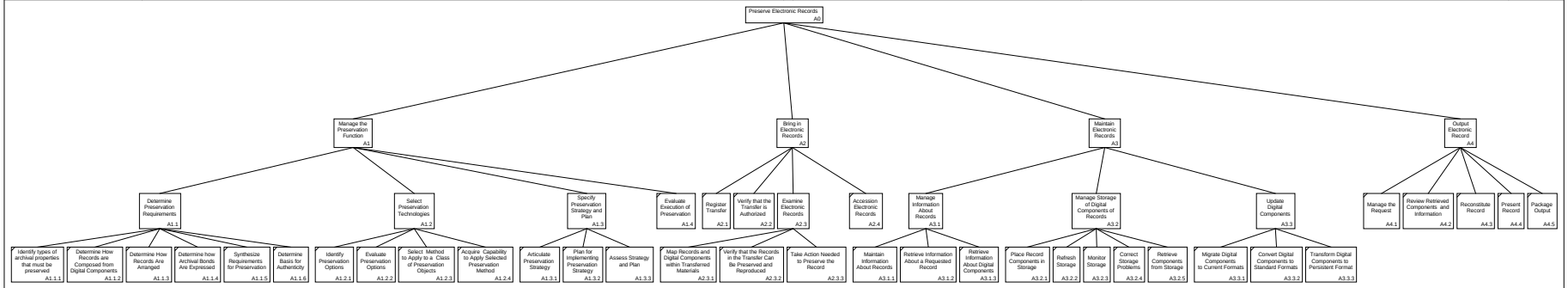
September, 2001

Task Force Members:

Kenneth Thibodeau, U.S. National Archives and Records Administration, chair,
Richard Blake, Public Records Office, UK,
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

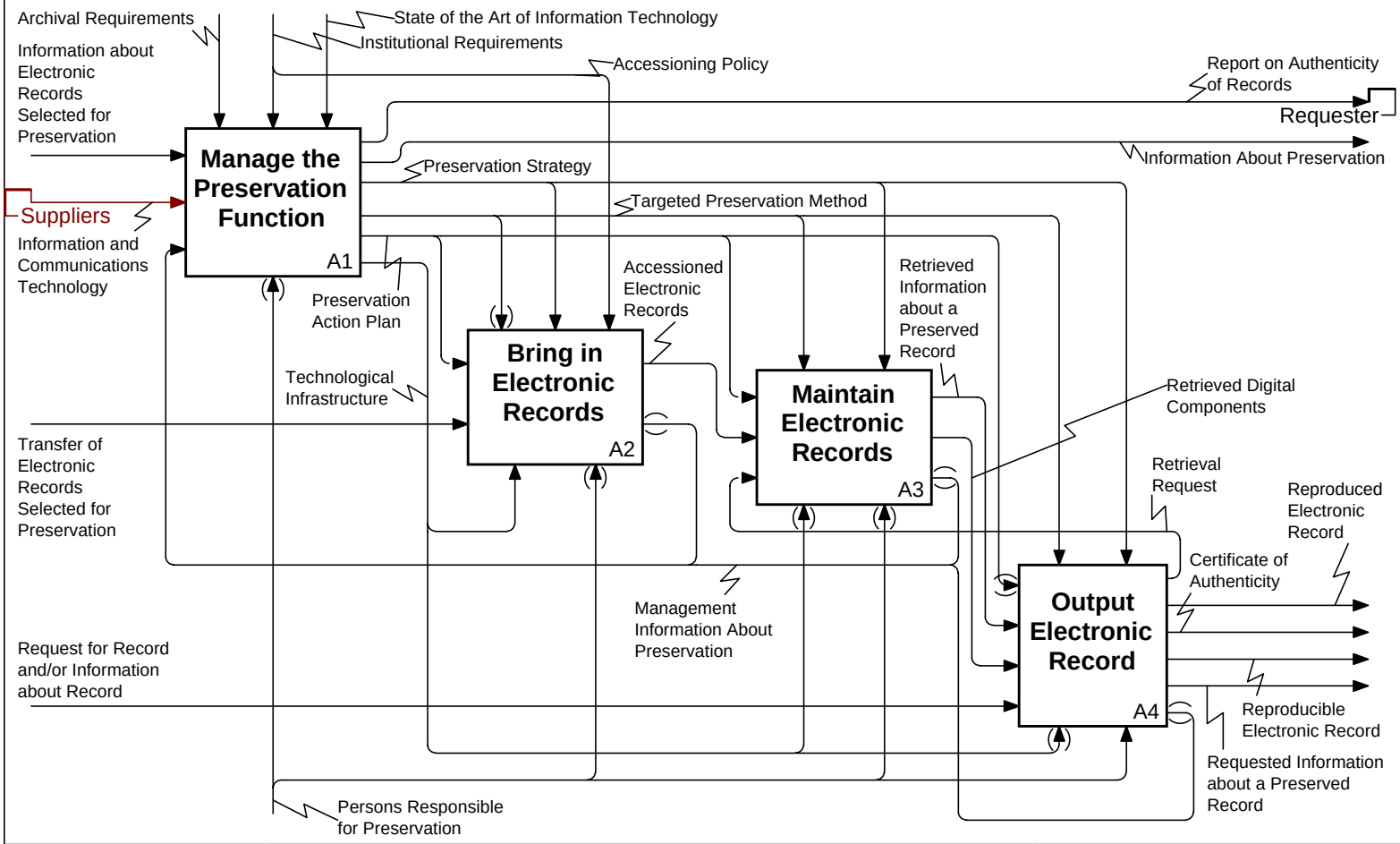
MODEL INFORMATION	
TITLE	*Preserve Electronic Records*
AUTHOR	Preservation Task Force, InterPARES Project
MODEL TYPE	IDEF(0) function model. IDEF(0) (Integration Definition for Function Modeling) is a U.S. Federal Information Processing Standard (Publication 183, as issued by the National Institute of Standards and Technology). "A function model is a structured representation of the functions, activities or processes within the modeled system or subject area." See <www.idef.com> for more information.
VERSION	5.1
VERSION DATE	September 30, 2001
VERSION HISTORY	• Febraury 15, 2001 - 2000 (InterPARES International Team Research Workshop #8) • Version 3.0, October 31, 2000 • Version 2.3 October 18, 2000 • Version 2.2 July 11, 2000 • Version 2.1 June 25, 2000 (InterPARES International Team Research Workshop #5) • Version 2. June 22, 2000 (InterPARES International Team Research Workshop #5) • Version 1. February 18, 2000. (InterPARES International Team Research Workshop #4)
PURPOSE	The purpose of this model is to articulate the functions, information and resources required to preserve permanent, authentic electronic records. The InterPARES Project will use this model to identify and develop the procedures and resources required for the implementation of the conceptual requirements and criteria identified in the project's Authenticity and Appraisal research domains.
VIEWPOINT	Person responsible for preservation
SCOPE	This model is constructed within the framework established by the Reference Model for an Open Archival Information System (OAIS), which is an ISO Draft International Standard (DIS). [See < http://ssdoo.gsfc.nasa.gov/nost/isoas/> for more information.] The 'Preserve Electronic Records' model includes 'Preserve Electronic Records' model activities and related ICOMs specifically required for the preservation and delivery of authentic electronic records. While some of these activities fall within the Ingest, Distribution and Management activities in the OAIS model, the 'Preserve Electronic Records' model excludes aspects of those activities not essential for preservation.

USED AT:	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 10/9/2001 REV: 9/30/2001	■ WORKING DRAFT RECOMMENDED PUBLICATION	READER	DATE	CONTEXT: TOP A-0



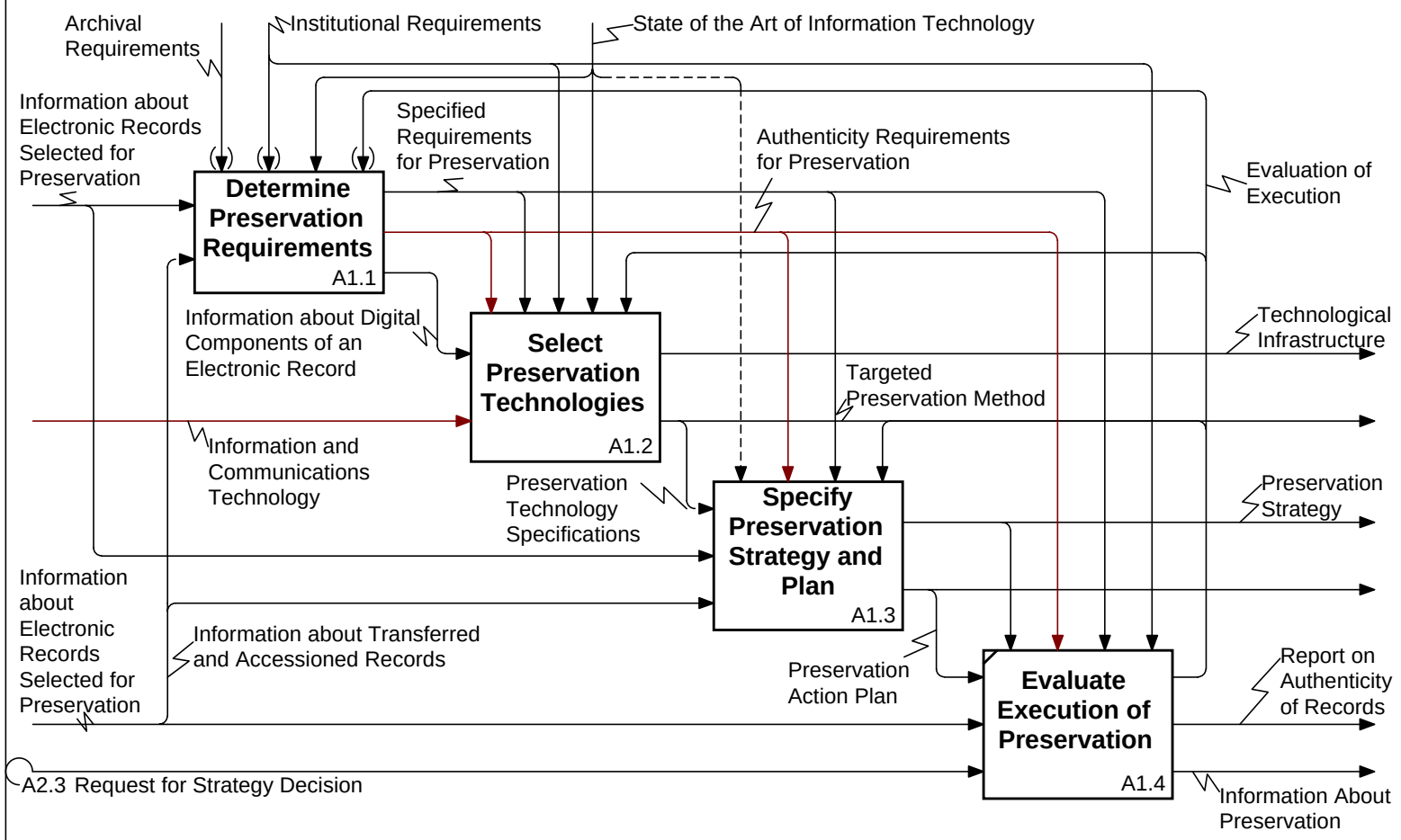
NODE:	TITLE:	NUMBER:
A0	Preserve Electronic Records	

USED AT:	AUTHOR: Preservation Task Force	DATE: 2/17/2000		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 9/30/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A-0



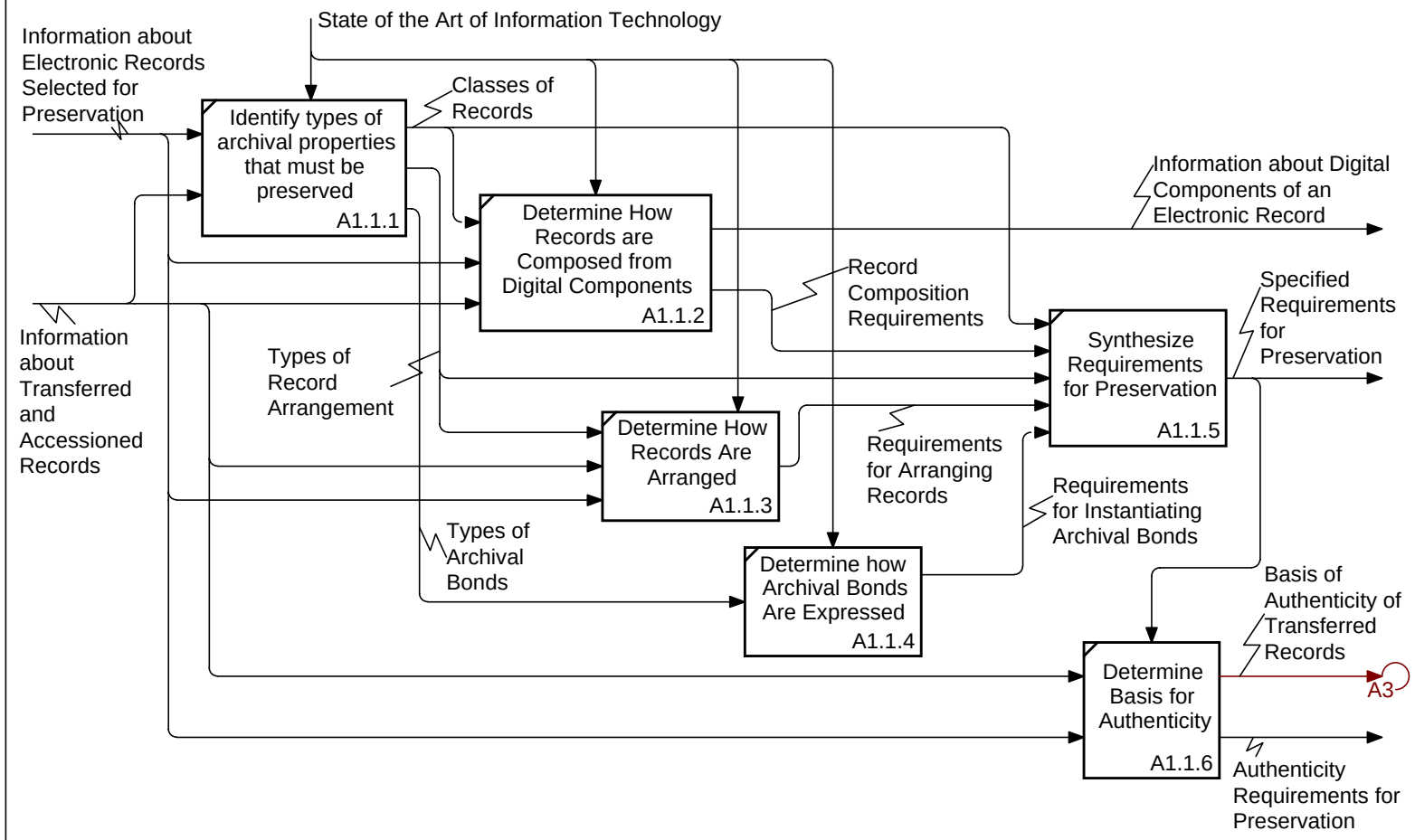
NODE:	TITLE:	NUMBER:
A0	Preserve Electronic Records	v 5.1

USED AT:	AUTHOR: Preservation Task Force	DATE: 2/18/2000		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 9/30/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A0



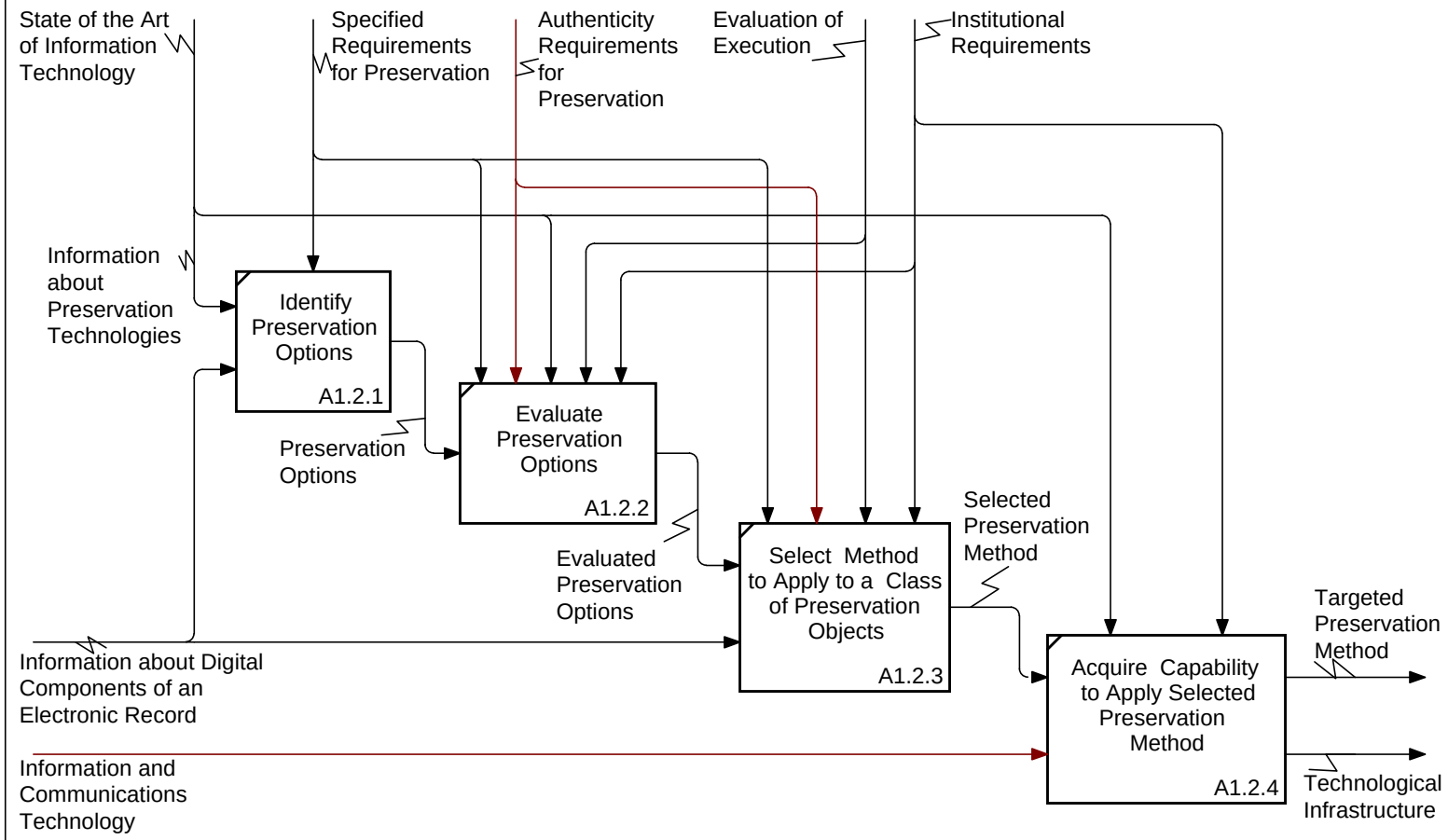
NODE:	TITLE:	NUMBER:
A1	Manage the Preservation Function	v 5.1

USED AT:	AUTHOR: Preservation Task Force	DATE: 4/8/2001		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 9/30/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A1



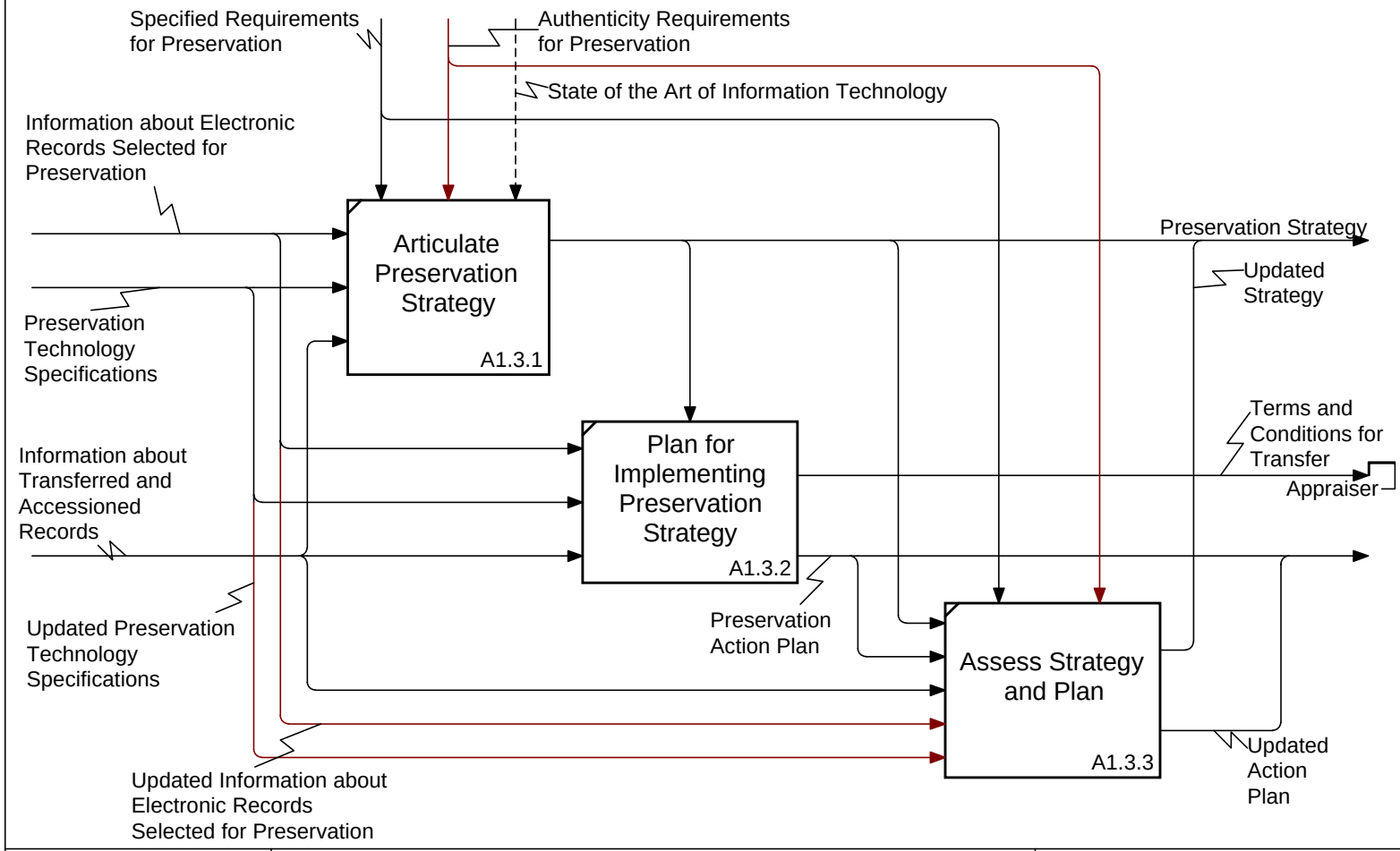
NODE:	TITLE:	NUMBER:
A1.1	Determine Preservation Requirements	v. 5.1 draft

USED AT:	AUTHOR: Preservation Task Force	DATE: 2/12/2001		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 9/30/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A1



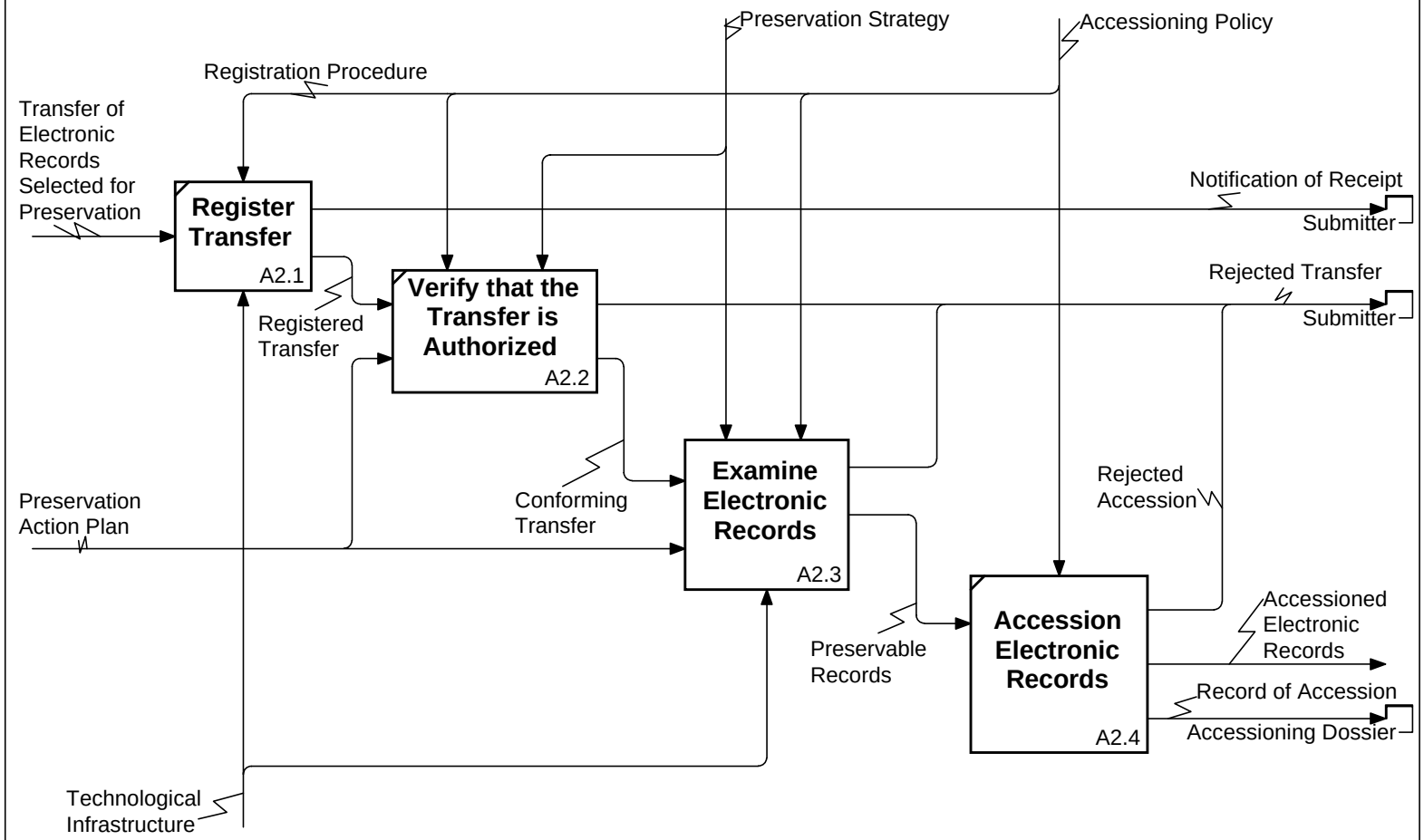
NODE:	TITLE:	NUMBER:
A1.2	Select Preservation Technologies	v. 5.1

USED AT:	AUTHOR: Preservation Task Force	DATE: 5/11/2001		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 10/6/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A1



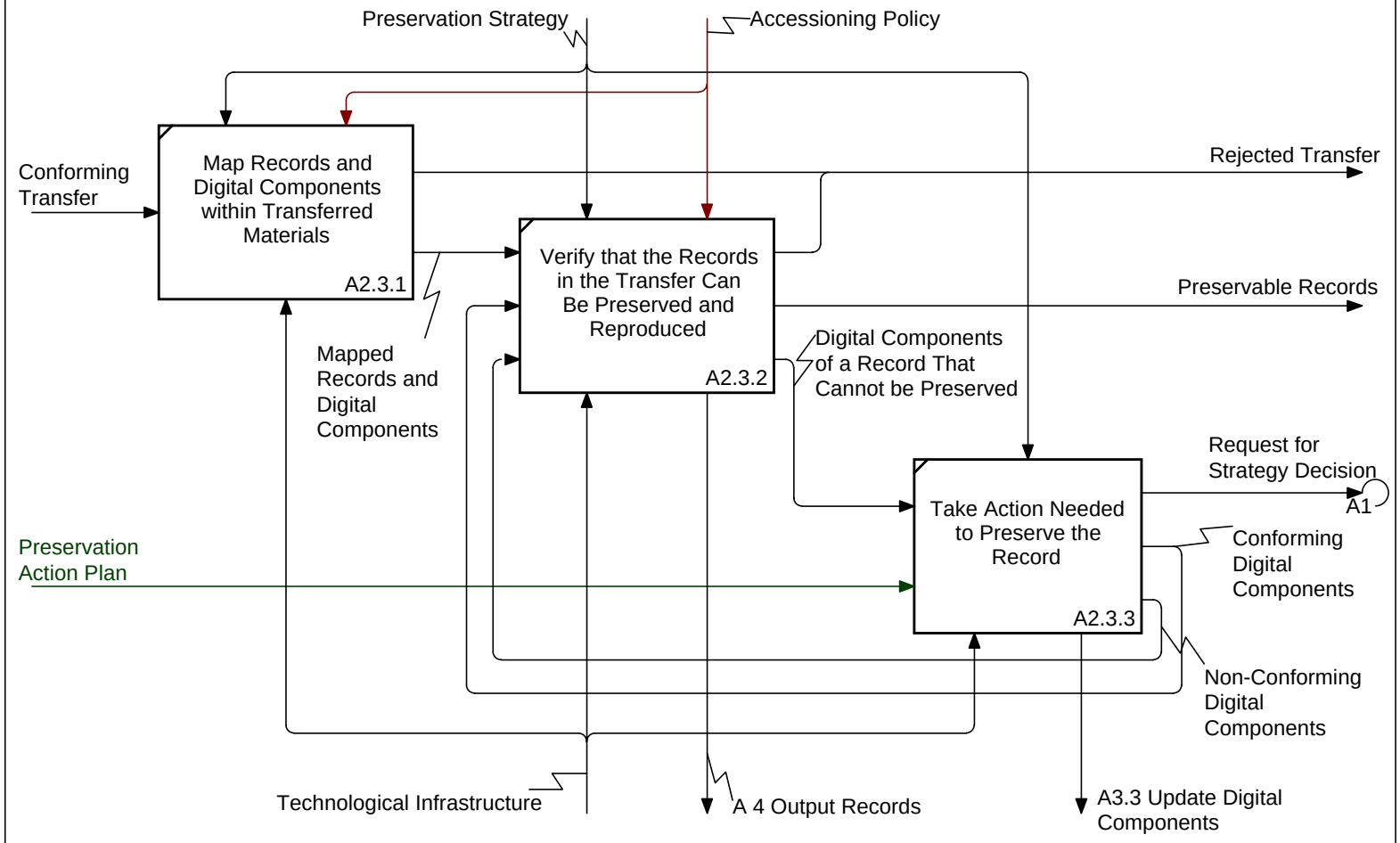
NODE:	TITLE:	NUMBER:
A1.3	Specify Preservation Strategy and Plan	v 5.1

USED AT:	AUTHOR: Preservation Task Force	DATE: 2/18/2000		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 9/30/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A0



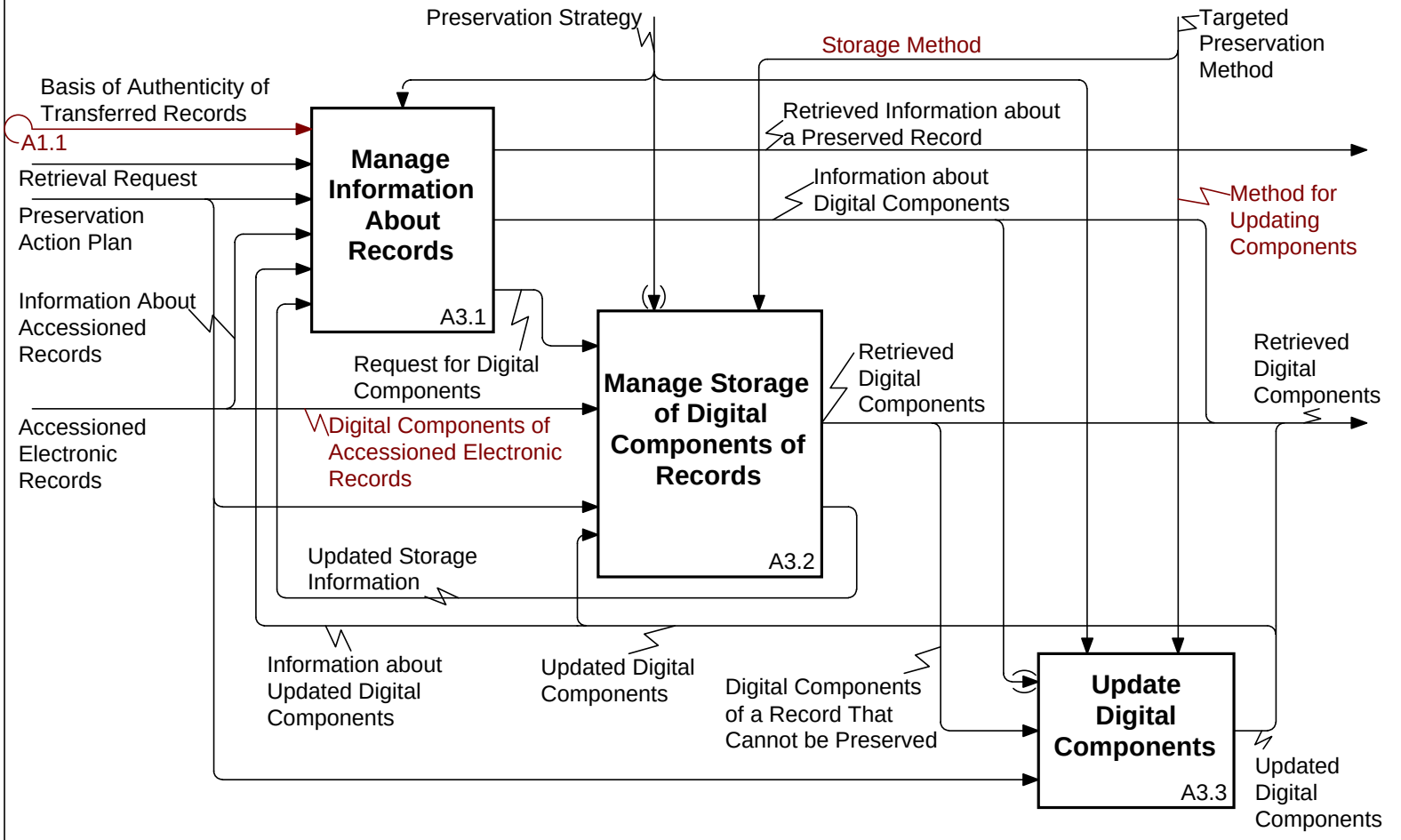
NODE:	TITLE:	NUMBER:
A2	Bring in Electronic Records	v. 5.1

USED AT: PTF WOrkshop 8 2001	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 10/11/2000 REV: 9/30/2001		WORKING DRAFT RECOMMENDED PUBLICATION	READER DATE	CONTEXT: A2



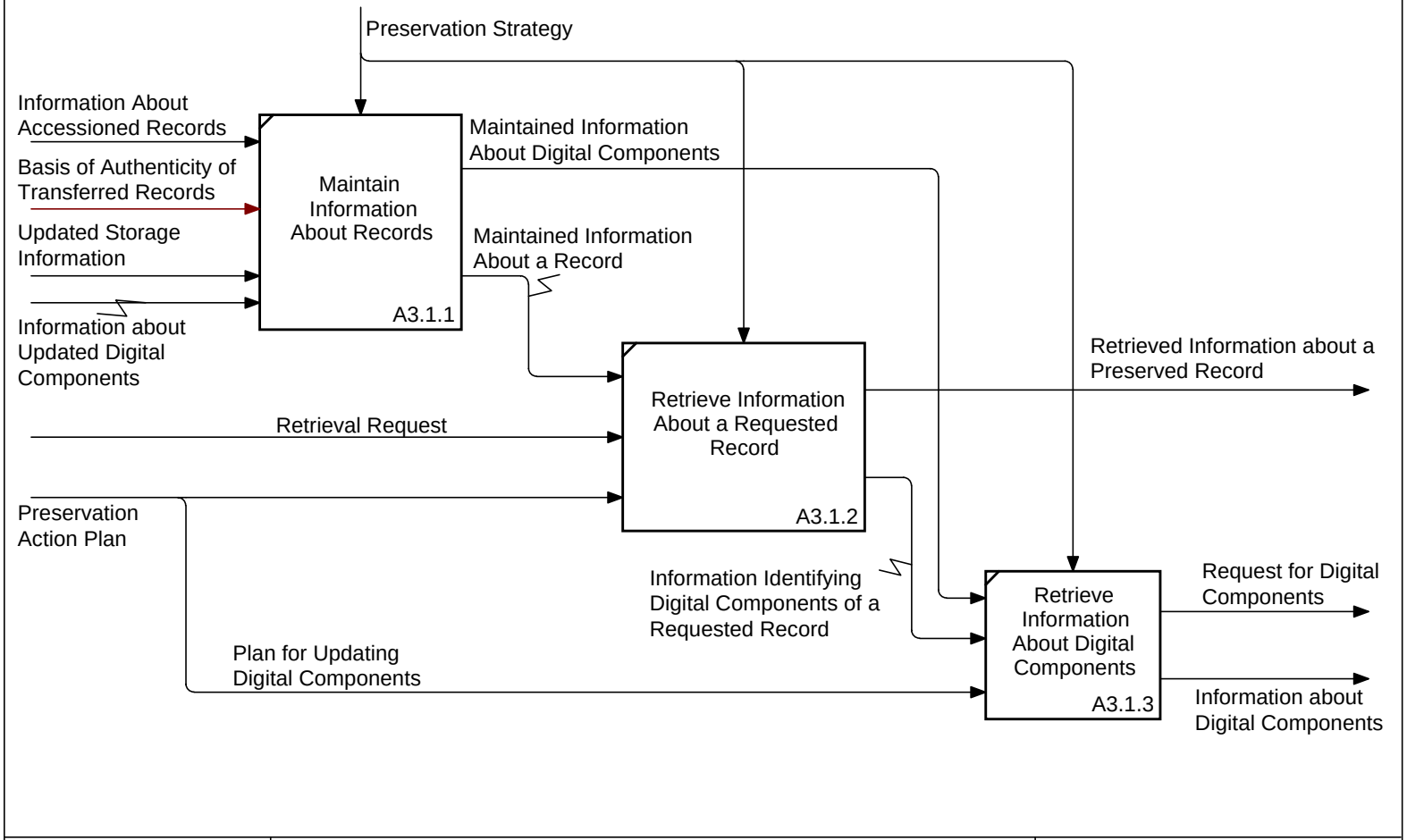
NODE:	TITLE:	NUMBER:
A2.3	Examine Electronic Records	v 5.1

USED AT:	AUTHOR: Preservation Task Force	DATE: 2/17/2000		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 9/30/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A0



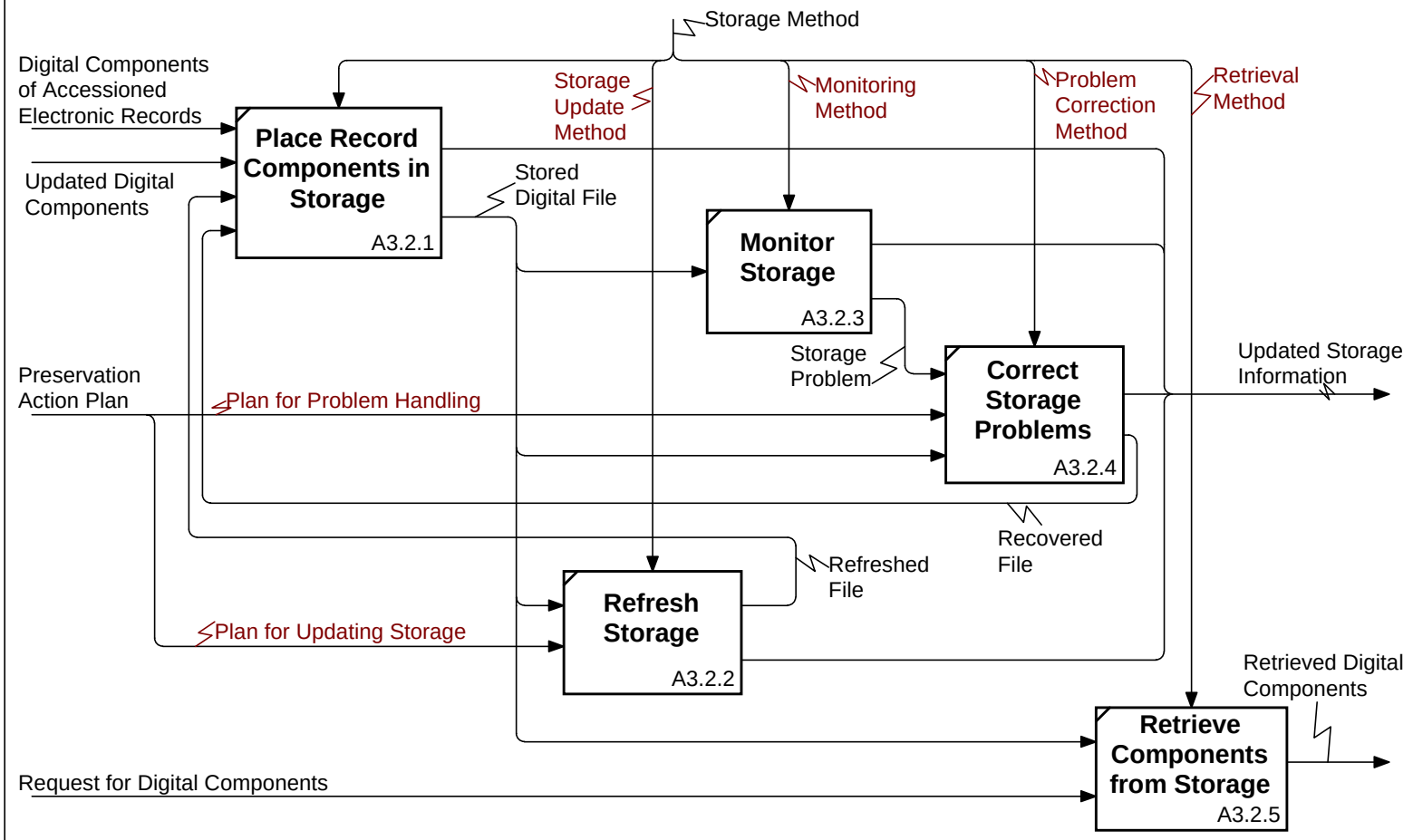
NODE:	TITLE:	NUMBER:
A3	Maintain Electronic Records	v 5.1

USED AT: PTF Workshop 8 2001	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 2/11/2001 REV: 9/30/2001	WORKING DRAFT RECOMMENDED PUBLICATION	READER DATE	CONTEXT: A3



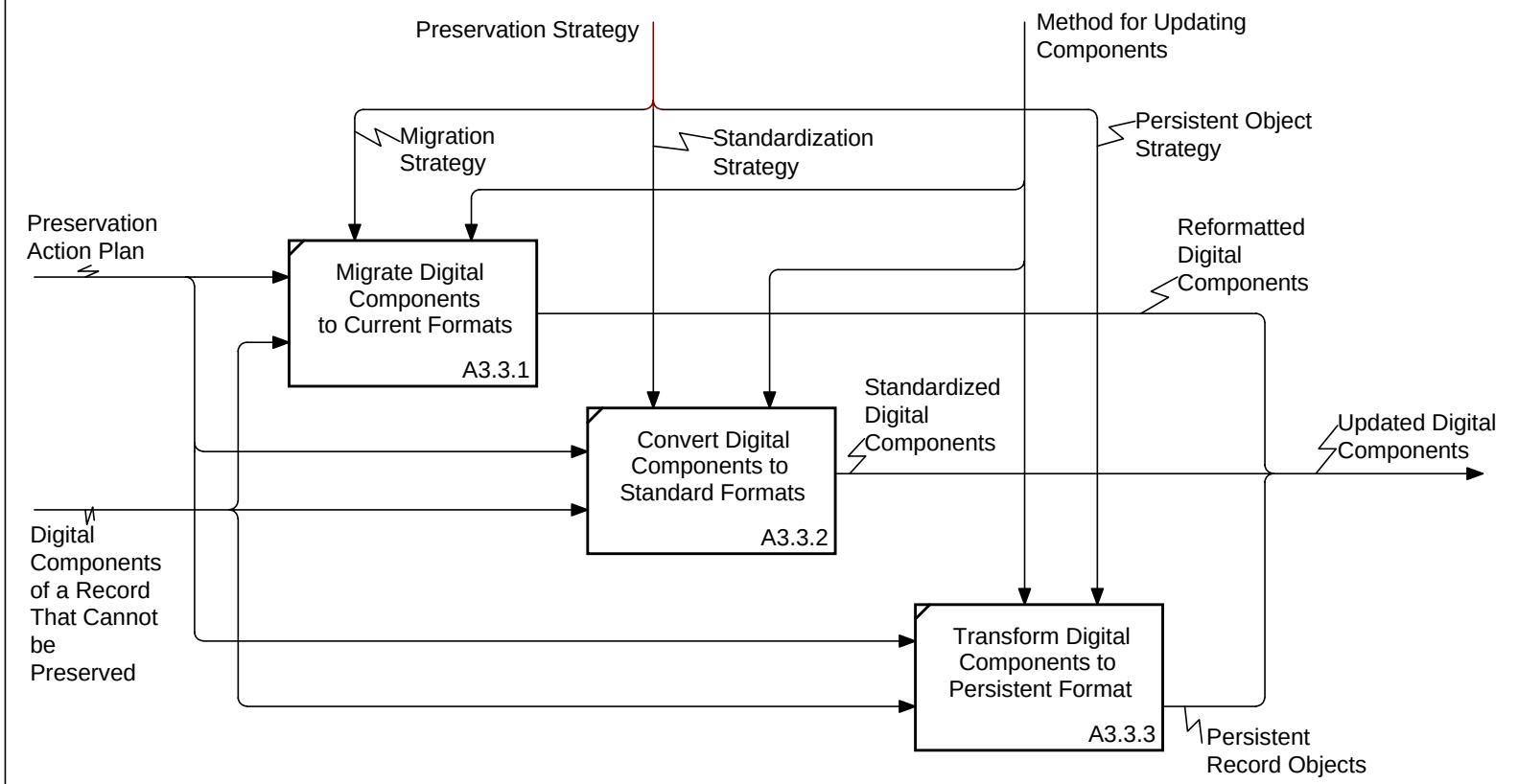
NODE:	TITLE:	NUMBER:
A3.1	Manage Information About Records	v 5.1

USED AT:	AUTHOR: Preservation Task Force	DATE: 6/25/2000		WORKING	READER DATE	CONTEXT:
after PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 10/9/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A3



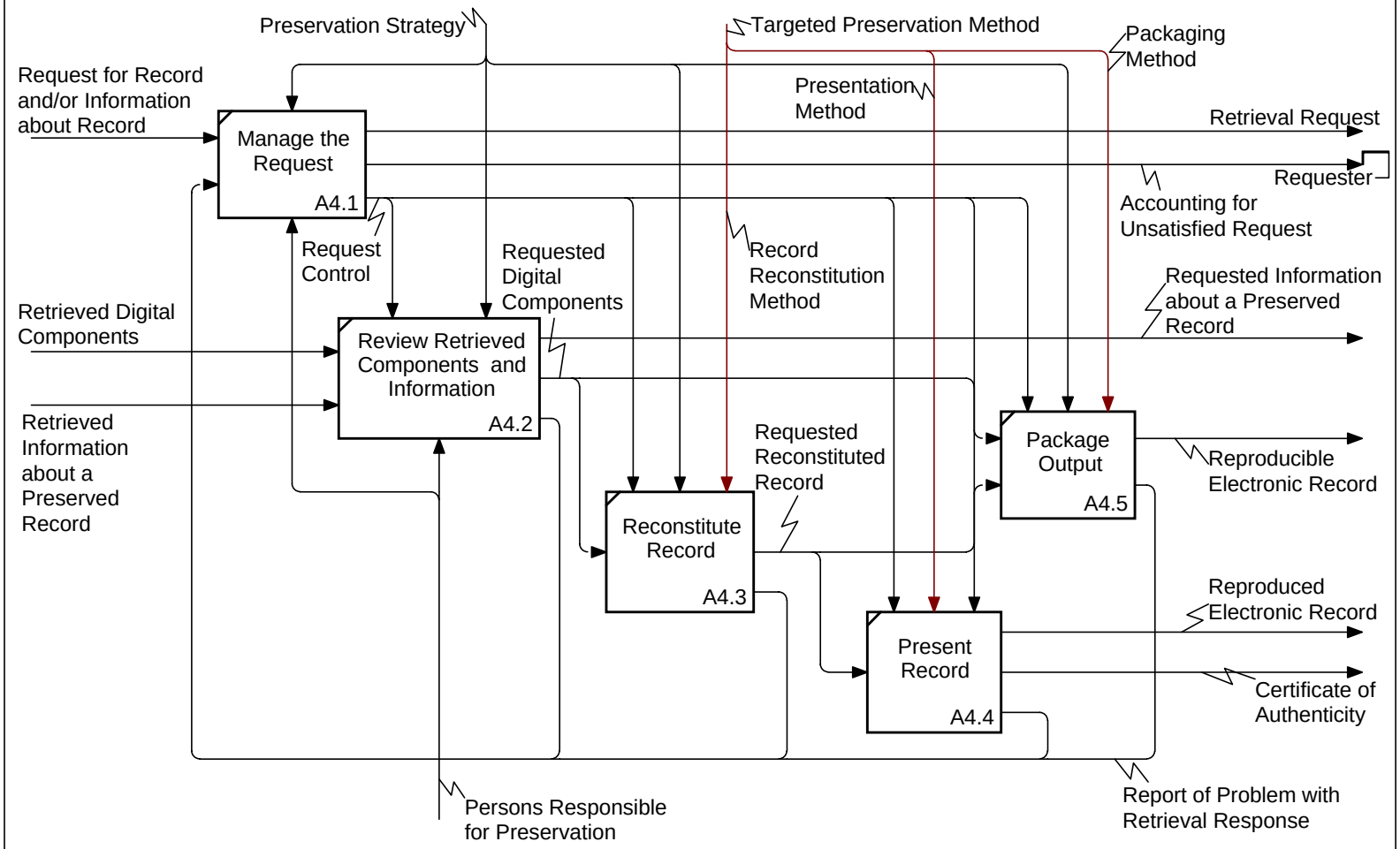
NODE:	TITLE:	NUMBER:
A3.2	Manage Storage of Digital Components of Records	v 5.1 draft

USED AT: PTF Workshop 8 2001	AUTHOR: Preservation Task Force PROJECT: InterPARES Project NOTES: 1 2 3 4 5 6 7 8 9 10	DATE: 7/11/2000 REV: 10/6/2001		WORKING DRAFT RECOMMENDED PUBLICATION	READER DATE	CONTEXT: A3



NODE:	TITLE:	NUMBER:
A3.3	Update Digital Components	v 5.1

USED AT:	AUTHOR: Preservation Task Force	DATE: 6/24/2001		WORKING	READER DATE	CONTEXT:
PTF Workshop 8 2001	PROJECT: InterPARES Project	REV: 10/11/2001		DRAFT		
			■	RECOMMENDED		
	NOTES: 1 2 3 4 5 6 7 8 9 10			PUBLICATION		A0



NODE:	TITLE:	NUMBER:
A4	Output Electronic Record	v 5.1

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Preservation Task Force Activity Definitions

Draft Version 5.1

September, 2001

Task Force Members:

Kenneth Thibodeau, U.S. National Archives and Records Administration, chair,
Richard Blake, Public Records Office, UK,
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

Activity Name	Activity Number	Activity Definition	Activity Note
Preserve Electronic Records	A0	Under the control of Archival and Institutional Requirements, and limited by the possibilities available within the State of the Art of Information Technology, preserve electronic records received in Transfers of Electronic Records. To preserve the records, the process uses Information about the Electronic Records Selected for Preservation. The process will also accept input of a Request for A Record or for Information about a Record and produce, in response, either the Requested Information About a Preserved Record, a Reproduced Electronic Record, or a Reproducible Electronic Record. The process will produce a Certificate of Authenticity from any Reproduced Electronic Record if requested. The process also produces Information about Preservation, including information about the process itself, the records being preserved, and how the process preserves their authenticity. Electronic records are preserved by Persons Responsible for Preservation using Information and Communications Technology and Facilities. The technology infrastructure and facilities are used in all subprocesses; therefore, they are implicit, rather than explicit, in all decomposition diagrams.	
Manage the Preservation Function	A1	Operating under the control of archival and institutional requirements and in light of knowledge of the state of the art of information technology, preservation is managed by producing a comprehensive preservation framework consisting of sets of preservation strategies and preservation action plans, each linked to a specific body of electronic records selected for preservation, along with the technological infrastructure and preservation methods needed to implement the action plans. These outputs are all used in execution of the preservation function. The 'manage' process also outputs information about the preservation function and about records being preserved and, on request, will produce a report on the authenticity of one or more records. The 'manage'	Preservation framework: Plans, principles, guidelines and objectives for the preservation of records, produced by interpreting external controls and applying them to the bodies of records appraised for preservation. [Proposed to Glossary Cttee. at W 7 2/2001] Prior definitions: Establishing the preservation framework, deriving preservation strategies from that framework, evaluating execution function and reporting. Use Information about Electronic Records Selected for Preservation and Management Information about the exercise of the preservation function

Activity Name	Activity Number	Activity Definition	Activity Note
		process produces its output using information about bodies of records selected for preservation, received from appraisal, and information about transferred records, received from the ingest process, and management information received as feedback from all processes involved in execution of the preservation function. Management activities are accomplished by persons responsible for preservation.	
Determine Preservation Requirements	A1.1	Determine the archival requirements for preserving and reproducing an electronic record by interpreting and applying external controls on the preservation function to information received from the Select Electronic Records function about the records to be preserved, and information about transfers and accessions received as feedback from the "Bring In," "Maintain," and "Output" Electronic Records processes. This entails identifying the classes of objects that must be preserved, including types of records and ordered groups of records, and specifying, for each class, the attributes and methods that must be preserved, as well as the requirements for certifying that any reproduced record is authentic. Determination of archival requirements is guided by evaluation of prior experience in applying such requirements to records that have been transferred to the archives. The evaluation will not alter the external requirements, but will improve their articulation and application. The result of this process will be specified archival requirements for preservation, where the specification consists in identifying what archival and institutional requirements apply to what records and how each applicable requirement is to be implemented.	Old definition: Establish the set of rules, norms, policies, and standards which interpret external controls on the preservation function and specify how those controls are to operate, given information received from the Select Electronic Records function about the records that are to be preserved, and information about transfers and accessions received as feedback from the "Bring In" and "Maintain" processes. Identify classes of objects that must be preserved, including types of records, archival sets of records, and digital components of records and specify, for each class, the attributes and methods that must be preserved, as well as the requirements for certifying that any reproduced record is authentic. Review the terms and conditions for transfer established in appraisal and define tests that will be used to determine if these terms and conditions are met. [Workshop 8, 6/21/2001]
Identify types of archival properties that must be preserved	A1.1.1	Limited by the State of the Art of Information Technology, use Information about Electronic Records Selected for Preservation -- and also Information about Transferred Records when, on examination, their properties are found to be different than what had been determined in appraisal -- to identify the	Old definition: Identify the types of records, defined by documentary form, that have to be preserved. Use information on all electronic records selected for preservation, including information from appraisal and from exercise of the preservation function

Activity Name	Activity Number	Activity Definition	Activity Note
		types of archival properties that must be preserved. This activity will identify the Classes of Records that exist within a body of records selected for preservation. For each Class of Record, determine how Archival Science, the preserver's Institutional Requirements (tunneled into this diagram), determinations made in appraisal, and/or the records management practices of the records creator, indicate specific archival properties that must be preserved, including intrinsic or extrinsic elements of form. For the body selected for preservation, specify the Type(s) of Arrangement of records established by the records creator and identify how archival bonds are expressed. For each archival property that must be preserved, determine the parameters, measures, or other evidence that will be used to identify the property and to verify that it has been preserved intact.	for records that have been transferred to the archives, to identify the classes of records. For each type of record, determine whether archival science, appraisal, the records management practices of the records creator, and/or institutional requirements specify intrinsic or extrinsic elements of form or methods necessary for preserving authenticity. For each such property, determine the parameters, measures, or other evidence that will be used to identify the property and verify that it has been preserved intact. [KT 11.04.2001] Examples: For e-mail records, the preserver must ensure that the sender, the recipients, any qualifications of recipients (e.g., carbon copy, member of a distribution list), the date sent, the date opened, the subject assigned by the sender, the message body, and any attachments are identified and that the preserver can either verify that none of these elements have changed or account for any changes in the course of preservation. Institutional requirements might specify that, if an archival set includes e-mail and web pages, any URLs or e-mail addresses included in either the subject line or body of an e-mail message must be specifically identified in order to enable recognition of records that are related by such explicit cross-references. [KT 11.04.2001]
Determine How Records are Composed from Digital Components	A1.1.2	Limited by the State of the Art of Information Technology, use Information about a body of Electronic Records Selected for Preservation, Information about previously Transferred and Accessioned Records from that body, and about Classes of Records, to identify the digital components of each class of record and how a record is composed	Old definition: Identify the methods through which a record which contains more than one digital component is constructed from those components. Use information on all electronic records selected for preservation, including information from appraisal and from exercise of the preservation function

Activity Name	Activity Number	Activity Definition	Activity Note
		from its digital components in order to articulate the requirements for reconstituting a record from its digital components and presenting the record in the documentary form stipulated for that class of record. The result is a set of Record Reproduction Requirements for each class of records, including the parameters, measures, or other evidence that will be used to verify that a record has been reproduced properly.	for records that have been transferred to the archives, along with information on types of records and classes and groups of digital components and the preservation requirements applicable to each, to identify the methods. For each identified method, determine if and under what conditions a method may be substituted for another. Determine the minimal set of methods that must be available within the archives for reconstituting a record from its components, and identify the record, or sets of records, as appropriate, to which each method applies or will apply. For each method, determine the parameters, measures, or other evidence that will be used to verify that it functions properly when records are reproduced. . [KT 19.04.2001] [amended MS 9/26/01] Examples of methods include embedding one digital component in another, linking from one digital component to another stored elsewhere, composing records via database selections, queries, reports, or views. [KT 19.04.01]
Determine How Records Are Arranged	A1.1.3	Guided by Archival Science and Institutional Requirements (tunneled to this diagram), and limited by the State of the Art of Information Technology, use Information about a body of Electronic Records Selected for Preservation -- and also Information about Transferred Records when, on examination, their properties are found to be different than what had been determined in appraisal -- along with information on Type(s) of Record Arrangement to determine how to identify an archival aggregate and identify the records or other aggregates that belong in it and specify the ordering of the members within the aggregate and of the aggregate within the archival fonds. Define criteria that will be used to determine if records and aggregates have been properly arranged	Old definitions: For each archival aggregate selected for preservation, determine what is needed to identify it, identify the records that belong in it, and specify the ordering of the records, and how it is related to other aggregates. [KT 11.04.2001] Identify the methods that must or could be used to implement the arrangement of records and of aggregates, and determine whether any of these are the same as methods for composing individual records from digital components. Determine if the methods used to reconstitute archival aggregates vary depending on the types of records they

Activity Name	Activity Number	Activity Definition	Activity Note
		when the aggregate is reconstituted. This activity will produce Requirements for Arranging Records.	include, especially when archival sets include more than one type of records. Define criteria that will be used to determine if records and aggregates have been properly arranged when the set is reconstituted. [KT 19.04.2001] Examples For a traditional filing system, the highest level set would be all the records filed in that system. the second highest level would be all the records grouped together in a transfer conforming to selection criteria established in the appraisal of the records. The lowest level set would be a single file, unless there were some subdivision of individual files specified in the filing system. Within the set constituted by a transfer, intermediate level sets would correspond to all defined levels of the filing system. In this case, the highest level set would be uniquely identified by its provenance and the name of the filing system, and characterized by its inclusive dates. A transfer set might be identified by an unique accession number assigned by the archives and characterized by (a) by inheritance its provenance and the name of the filing system and (b) explicitly by its inclusive dates. Intermediate levels would be identified by a combination of the accession number and the name or code assigned to them in the filing system. Ordering of the members would be chronological for transfer sets, alphabetically or numerically by filing code or name for intermediate levels, and finally within files either chronologically, empirically as found at the time of transfer, or according to explicit rules articulated by the records.
Determine how Archival	A1.1.4	Guided by Archival Science and Institutional	[KT 9 2001]

Activity Name	Activity Number	Activity Definition	Activity Note
Bonds Are Expressed		Requirements (tunneled to this diagram), and limited by the State of the Art of Information Technology, use information about the Types of Archival Bonds in a body of electronic records selected for preservation to determine how archival bonds are identified, what techniques are necessary to preserve contextual relationships and technical dependencies and which technologies are necessary for the reproduction of authentic relationships of records both within and between archival aggregates. This activity will produce the Requirements for Instantiating Archival Bonds.	
Synthesize Requirements for Preservation	A1.1.5	Guided by Archival Science and Institutional Requirements (tunneled to this diagram), synthesize information about Classes of Records contained in a body of electronic records selected for preservation, Record Reproduction Requirements applicable to those classes, Requirements for Arranging Records, and Requirements for Instantiating Archival Bonds within that body of records, producing a comprehensive and coherent set of Specified Requirements for Preservation of the selected records. These requirements will then serve as a principal control to Determine the Basis for Authenticityn, as well as guiding the functions of technology selection, preservation planning and evaluation.	Old definitions: Given generic archival requirements for the preservation of authentic electronic records, ensure that the total set of preservation requirements specified for any record or group of records selected for preservation is complete and, as appropriate, internally consistent. [KT 05/2001] The preservation requirements for preserving archival properties of classes of records, for reconstituting record from their digital components, and for preserving the original ordering and archival bonds among records are reviewed altogether to ensure that they are complete, consistent and would support certification of the authenticity of reproduced records [7/2001]
Determine Basis for Authenticity	A1.1.6	Guided by Archival Science and Institutional Requirements (both tunneled to this diagram) and the Specified Requirements for Preservation applicable to a body of records selected for preservation, determine the basis for asserting the authenticity of the records. This basis will have two parts: information supporting the presumption of the authenticity of the records as transferred from the creator and information about how the preserver satisfies the applicable Specified Requirements for Preservation after the records are transferred to the preserver. The information	Old definition: Given baseline archival requirements for presuming that the records creator maintained the records authenticity, use information about the records creator's practices to assess the extent to which the creator maintained authentic records prior to transfer to the archives. The assessment becomes a general control on the certification of authenticity of copies reproduced by the archives. Given archival requirements concerning the

Activity Name	Activity Number	Activity Definition	Activity Note
		requirements to support a presumption of authenticity of records up to their transfer is received as part of Information about Electronic Records Selected for Preservation and Information about Transferred and Accessioned Records, and will be designated for maintenance along with other information about the records. The types of information required to support an assertion that the preserver has preserved and reproduced authentic records becomes a control on subsequent preservation activities.	identity and integrity of records, and information about the types of records and classes of archival sets and digital components of records, determine the basis for certifying the authenticity of copies of each type of record. Analyze whether each type of record is sufficiently complete with respect to an action to be determined authentic in itself, or whether one or more requirements for authentication are only implicit in the record and need to be supplied either from other records to which it is bound archivally or from the records system. [KT 5 2001] This includes analyzing whether a record is sufficiently complete with respect to an action to be determined authentic in itself, or whether one or more requirements for authentication are only implicit in the record and need to be supplied either from other records to which it is bound archivally or from the records system. Requirements for certifying authenticity that depend on activities within the Preserve Electronic Records model should be part of the synthesis of requirements for preservation produced in activity A1.1.5. The authenticity requirements that apply to how records are composed and arranged and how their archival bonds are expressed should be parts of the outputs of activities A1.1.2, A1.1.3, and A1.1.4, respectively. Therefore this activity can be deleted.
Select Preservation Technologies	A1.2	Take into account Information about Electronic Records Selected for Preservation and Information about Transferred and Accessioned Records in order to select Preservation Methods that will be used to preserve the electronic records. Acquire Information and Communications Technology providing both the	Institutional requirements that apply to the selection of methods, but not to determining archival requirements include the need to conform to the institution's information technology architecture, data standards and related procedures, security requirements,

Activity Name	Activity Number	Activity Definition	Activity Note
		Preservation Methods themselves and the Technological Infrastructure necessary to execute those methods. Determine the applicability of each method by specifying the target bodies of records, types of electronic records, and/or classes of digital components to which it will apply, along with conditions for its application to each target group, thus defining a Targeted Preservation Method. The process is controlled by Institutional Requirements, such as the institution's information technology architecture and standards, data standards, and related procedures, security requirements, access restrictions, and performance objectives and measures. It is also controlled by the Specified Requirements for Preservation of a body of electronic records, applicable Authenticity Requirements for Preservation, and Evaluation of the Execution of the preservation methods that have been used to date. The selection is limited by the State of the Art of Information Technology. The scope of this process includes all methods which could impact the preservation of the records from the time of transfer. That includes methods for checking the integrity of the transfer process, methods related to storage of digital components in digital files and on physical media, methods to reconstitute and reproduce records, and methods to enable others to reproduce the records.	access restrictions, and performance objectives and measures. [KT 4 2001]
Identify Preservation Options	A1.2.1	Given Information about the Digital Components of Electronic Records to be preserved, determine what Preservation Options exist within the State of the Art of Preservation Technologies that might satisfy the Specified Requirements for Preservation to preserve and reproduce the record, properly ordered within the archival aggregate(s) to which it belongs.	Old definition: Given information about the technical characteristics of electronic records selected for preservation and about the state of the art of information technology, identify the methods or techniques that could be applied to or used with the records to satisfy the archival requirements for preservation and reproduction of authentic copies of the records.
Evaluate Preservation Options	A1.2.2	Evaluate the identified Preservation Options applicable to a body of electronic records selected for	Old definition: Given information about records selected for preservation, and an

Activity Name	Activity Number	Activity Definition	Activity Note
		preservation according to their ability to satisfy the applicable Specified Requirements for Preservation and Authenticity Requirements for Preservation, taking into account the Evaluation of Execution of preservation methods used to date and the Institutional Requirements of the preserver related to factors such as its information technology standards and architecture, and its procedures and criteria for selection and implementation of information technology. This activity will produce Evaluated Preservation Options.	evaluation of the success of current preservation methods as applied to similar records that have already been accessioned into the system, evaluate available preservation methods that could either complement or replace methods currently being used to preserve electronic records against the archival requirements for preservation.
Select Method to Apply to a Class of Preservation Objects	A1.2.3	From Evaluated Preservation Options, select the methods necessary and appropriate for preserving a body of electronic records selected for preservation, given Information about the Digital Components of the Records. The method(s) selected should best satisfy the Specified and Authenticity Requirements for Preservation, as well as Institutional Requirements related to the selection of information technology. Specify the domain to which each Selected Preservation Method applies. This domain will include one or more bodies of records, classes of records, types of digital components, types of arrangement of records, methods for expressing archival bonds, and/or presentation methods.	Old definition: Having evaluated current and available preservation methods, select those that best satisfy archival and instutional requirements, and identify the archival sets and/or types of records and classes of digital components to which the selected methods should be applied.
Acquire Capability to Apply Selected Preservation Method	A1.2.4	Within the limits of the State of the Art of Information Technology and consistent with Institutional Requirements concerning the acquisition of information technologies, acquire and adapt, configure, or enhance as necessary, the hardware, software, media, services, and support needed to apply the Selected Preservation Methods for preserving and reproducing electronic records. The technology acquired in this activity will operate at two different levels: (1) Technological Infrastructure will provide the basis or platform on which (2) Targeted Preservation Methods can be applied. In general, Technological Infrastructure will consist of commercial off-the-shelf products such as computers, servers,	Old definition: Acquire and adapt, configure, or enhance as necessary, the hardware, software, facilities and support needed to implement the method adopted for preserving and reproducing an electronic record or record aggregate. Develop or modify procedures and rules for use of technology. [version 4 2001] [singularized 9-5-01, MS]

Activity Name	Activity Number	Activity Definition	Activity Note
		peripheral devices, operating systems, database management systems, communications software, digital storage media, etc. Targeted Preservation Methods may include commercial off-the -shelf products, for example, software to be run in emulation, but may also include applications developed specifically for the preservation and reproduction of records.	
Specify Preservation Strategy and Plan	A1.3	Guided by the Specified and Authenticity Requirements for Preservation and Evaluation of the Execution of Preservation processes and methods, use Information about Records Selected for Preservation, and Information about Transferred and Accessioned Records, along with Preservation Technology Specifications concerning both Infrastructure Technology and Targeted Preservation Methods, to develop a Preservation Strategy for preserving a body of records and producing authentic copies of the records, and to articulate one or more Preservation Action Plans to implement that Strategy. The goal of a Preservation Strategy is to ensure the preservation of authentic electronic records from the point at which they are transferred to the archival system, through the maintenance of their digital components over time, to the delivery of certifiably authentic reproducetions of the records.	The domain of a preservation strategy could be a set of records, a class of sets, a type of records, or a class of digital components. A stratgegy applicable to a set of records would include, for example, a prescribed method for ensuring that archival aggregates, such as files and series, could be preserved over time. The domain of a strategy applicable to a class of componets might be that of e-mail, or databases, web pages, or digital photographs. An example of a preservation action that would need to be taken only when specified conditions are encountered would one entailed by a change in applicable methods for reproduction of records. The new methods would have to be invoked only in processing a request for reproduction of the records. Old definition: In this process, the archival and institutional requirements which have been distilled into a set of comprehensive Archival Requirements for Preservation, which are both specific for a given body of records selected for preservation and consistent with requirements articulated for similar records and archival sets, together with the Preservation Methods selected for that body of records and evaluation of prior application of such methods control the development of preservation strategies and

Activity Name	Activity Number	Activity Definition	Activity Note
			action plans aimed at ensuring the preservation and reproduction of authentic electronic records from the point at which they are transferred from the current records system to the archival system, through the maintenance of their digital components over time, to the delivery of reproduced copies of the records which are certifiably authentic. Information about records selected for preservation, transferred to the archives or maintained by it are combined with specification of the information technology acquired to implement the Selected Preservation Methods are the inputs used to develop the preservation strategies and plans [KT 5 2001]
Articulate Preservation Strategy	A1.3.1	Controlled by the Specified and Authenticity Requirements for Preservation applicable to a body of electronic records selected for preservation, and limited by the State of the Art of Information Technology, use Information about this body of Electronic Records Selected for Preservation and Information about Transferred and Accessioned Records from the same body of records, along with Preservation Technology Specifications, which described the Targeted Preservation Methods applicable to these records, to specify and output a comprehensive Preservation Strategy for preserving the body of records.	
Plan for Implementing Preservation Strategy	A1.3.2	Within the parameters of the applicable Preservation Strategy, take into consideration Information about (a body) of Electronic Records Selected for Preservation, Information about Transferred and Accessioned Records from that body of records, and Preservation Technology Specifications, develop one or more Preservation Action Plans defining actions to take when the records are transferred, while they are being maintained, and in responding to requests for a record	

Activity Name	Activity Number	Activity Definition	Activity Note
		from that body of records. The Preservation Action Plan for transfer of records will include implementing Terms and Conditions for Transfer specified in agreement with the Appraiser responsible for selecting the records for preservation, writing digital components expected in the transfer to storage media, and any modifications of such components needed to conform to the applicable preservation strategy.	
Assess Strategy and Plan	A1.3.3	Using Information about (the application of the preservation strategy and implementation of the preservation action plan to) Transferred and Accessioned Records, determine whether and to what extent the Preservation Strategy and Preservation Action Plan(s) applicable to a body of electronic records selected for preservation have succeeded in satisfying Specified and Authenticity Requirements for Preservation. Use this assessment to produce an Updated Strategy and/or Updated Action Plan(s). If monitoring of electronic records selected for preservation prior to their transfer produces significant Updated Information about Electronic Records Selected for Preservation, revise the strategy or plan accordingly. Similarly, if applicable preservation methods have been changed, revise the strategy and/or plan to reflect the Updated Preservation Technology Specifications.	
Evaluate Execution of Preservation	A1.4	Consistent with Institutional Requirements for management analysis, and guided by applicable Preservation Strategies and by both Specified and Authenticity Requirements for Preservation, analyze Management Information About Preservation in order to determine how well goals and objectives are being achieved. The analysis addresses whether records are being preserved successfully. If not, it identifies the causes of problems and determines how they might be resolved or avoided. Identify cases where rules or objectives conflict, are unclear, insufficient, inappropriate or misdirected., as well as possible	Old definition: Analyze management information about preservation output from the "Bring In" and "Maintain" functions in order to determine whether records are being preserved successfully, resolve or avoid problems, and identify areas for improvement. Includes identifying cases where rules or objectives conflict, are unclear, insufficient, inappropriate or misdirected, as well as possible improvements in operation, and produce the evaluations of both performance strategies

Activity Name	Activity Number	Activity Definition	Activity Note
		improvements in operation. Assess whether the set of Preservation Strategies in effect are coherent and not conflicting. Determine if Preservation Methods are operating as intended and whether the Technological Infrastructure is adequate and appropriate. Produce evaluations of both performance strategies and the overall preservation framework. Review Management Information about the Output of Electronic Records to determine if requirements for authentic copies are being satisfied and to characterize customer satisfaction. Produce an Evaluation of Execution appropriate to guide other management processes. Output Information About Preservation for external stakeholders. Produce a Report on Authenticity of Records to respond to any challenge to the adequacy and efficacy of the preservation process. This process may be triggered by a Request for Strategy Decision.	and the overall framework. Output external reports on the exercise of the preservation function and information on the results of evaluation that serves to control other management activities so that mistakes are not repeated, problems are avoided, and/or overall performance is improved. [KT 4 2001]
Bring in Electronic Records	A2	Following direction established in the preservation strategy for a given body of records selected for preservation, the 'bring in' or ingest function applies preservation method(s) targeted to that body of records to implement the preservation action plan for those records by processing each transfer of electronic records into accessioned electronic records. The ingest process also produces information about each transfer of electronic records, which is used in the 'manage' process to confirm or revise the preservation strategy and action plan(s) applicable to those records, and also management information which is used to evaluate execution of the ingest function. The process is carried out by persons responsible for preservation, using infrastructure technology.	Old defintion: The process of bringing electronic records and related information into the preservation function and establishing control over them.
Register Transfer	A2.1	Following the Registration Procedure defined by the Preserver's Accessioning Policy, register the transfer of electronic records by capturing information about the transfer, such as submitter's name, record creator's name, and the date of receipt of the transfer	OAIS: "The Receive Submission function provides the appropriate storage capability or devices to receive a SIP from the Producer (or from Administration). Digital SIPs may be delivered via electronic

Activity Name	Activity Number	Activity Definition	Activity Note
		in a Record of the Transfer, and establish basic control over the materials transferred by identifying what has been transferred and where it is located. The registration record forms the basis for identifying and tracking materials in the transfer and information about it in all other preservation processes. The registration process should also inspect what was received in order to ensure that the physical transfer has been accomplished correctly. This inspection provides quality assurance of the physical transfer, but does not address any questions related to the specific records reportedly contained in the transfer. Registration produces a Notification of Receipt, which is sent to the submitter. This notification should describe any problem identified in receiving the transfer, such as network errors in transmission, missing media volumes, or obvious damage, and request the submitter to send a new transfer or otherwise correct the problem. The Registered Transfer is forwarded for verification that it is authorized.	transfer (e.g., FTP), loaded from media submitted to the archive, or simply mounted (e.g., CD-ROM) on the archive file system for access. Non-digital SIPs would likely be delivered by conventional shipping procedures. The Receive Submission function may represent a legal transfer of custody for the Content Information in the SIP, and may require that special access controls be placed on the contents. This function provides a confirmation of receipt of a SIP to the Producer, which may include a request to resubmit a SIP in the case of errors resulting from the SIP submission. OAIS: "The Quality Assurance function validates the successful transfer of the SIP to the staging area. For digital submissions, these mechanisms might include cyclic redundancy codes (CRCs) or checksums associated with each data file, or the use of system log files to record and identify any file transfer or media read/write errors." QA is connected solely to "Receive Submission."
Verify that the Transfer is Authorized	A2.2	Acting under the Preserver's Accessioning Policy, determine if the transfer is authorized; that is, it comprises the records that have been selected for preservation, and those records have been submitted either by the records creator or an agent acting for the creator. Verification that a transfer is authorized is based on comparing the terms and conditions for transfer, established as part of the Preservation Strategy during appraisal, with information accompanying the Registered Transfer, This information is reviewed to determine if it indicates that the transfer was sent by an authorized person; it comprises records specified for transfer; it includes necessary information about the records, their digital components, and the basis for asserting the	

Activity Name	Activity Number	Activity Definition	Activity Note
		authenticity of the records as received; and that the materials transferred are of the correct types and in the specified formats. This verification may include steps that are specific to the records reportedly contained in the transfer whenever such specific tests are set out in a Preservation Action Plan related to the records. If the terms and conditions of transfer have been satisfied, the Conforming Transfer is passed to the next step, where its contents are examined. Otherwise either the transfer is rejected outright or the submitter is asked to address any problems identified.	
Examine Electronic Records	A2.3	Acting under the preserver's Accessioning Policy and in accordance with the Preservation Strategy established for the records reportedly included in the transfer, examine the digital files and digital components of records in the transfer, along with accompanying information to: 1. determine if the transfer actually includes all records and aggregates of records specified in the terms and conditions of transfer and that these records and aggregates are adequately and accurately described in the accompanying information to enable their preservation, reproduction in authentic form, and interpretation; 2. identify any actions required to preserve both the individual records transferred and the archival sets in which these records belong; 3. initiate technical or other preservation actions that should be taken immediately and schedule preservation actions that should be taken at a later date. This examination will include any specific tests included in a Preservation Action Plan for the records. If on examination, any record, digital component, or aggregate is found to have unexpected properties that would make it difficult or impossible to follow the established Preservation Strategy, or would make it questionable whether following that strategy would satisfy archival requirements, the situation should be reported to the 'Manage' function for evaluation. If the evaluation results in a change in the applicable	

Activity Name	Activity Number	Activity Definition	Activity Note
		Preservation Strategy, the Manage function will issue a Preservation Action Plan specifying how to resolve the problem(s) identified in the examination. If the results of the examination are positive, the records in the transfer are deemed to be suitable for preservationm, and the Preservable Records are forwarded to be accessioned. If the examination reveals unresolvable critical problems, the transfer is rejected and returned to the Submitter. If it might be possible for the Submitter to correct problems that the Preserver could not resolve, the Rejected Transfer includes a request for corrective action and resubmission.	
Map Records and Digital Components within Transferred Materials	A2.3.1	In accordance with an institution's Accessioning Policy and the applicable Preservation Strategy, using the Technological Infrastructure in place, determine how the records, their digital components, and the information about them included in a Conforming Transfer are identified, and where they are located in the digital files and other materials received. Identify any records or components that should be in the transfer but are not found and determine if there are any records or components in the transfer that should not have been transferred. If there are critical deficiencies found, terminate processing and output the Rejected Transfers. If no critical problems are encountered, output the Mapped Records and Digital Components, along with related information.	Old definition Determine how the records, digital components, and the information about them that should be included in the transfer are identified, and where they are located in the digital files and other materials rerceived. Identify any records or components that should be in the transfer but are not found and determine if there are any records or components that should not have been transferred. [6/21/2001]
Verify that the Records in the Transfer Can Be Preserved and Reproduced	A2.3.2	Ensure that transferred records can be preserved and reproduced in accordance with the applicable preservation strategy. Determine if if all of the records that should be in the transfer can be reconstituted and presented; if all digital components of these have been received and are in the formats stipulated in the terms and conditions of transfer; if archival aggregates established by the records creator can be reestablished; if the archival bonds among records can be expressed or instantiated, and also what basis exists for asserting the authenticity of the records as	

Activity Name	Activity Number	Activity Definition	Activity Note
		transferred. Records that can be preserved and reproduced are output as Preservable Records. In the case of a record that cannot be preserved in accordance with the preservation strategy, identify the digital component(s) that prevent or impede implementation of the strategy and forward them for actions which will enable preservation of the record. Digital components modified as a result of such actions are returned for verification of the success of the modification. If it was necessary to modify the applicable preservation strategy to enable preservation of the records, the secondary verification will be on the basis of the revised strategy. The verification process will take into account any cases where digital components were not successfully modified, along with other problems discovered with the transfer, to determine whether, under the institution's accessioning policy, the preservable records – including records deemed to have only minor problems -- should be accessioned or the transfer should be rejected.	
Take Action Needed to Preserve the Record	A2.3.3	When it has been determined that a record in a transfer cannot be preserved in accordance with the applicable preservation strategy, take the steps indicated in the relevant preservation action plan with respect to any digital component of the record which needs to be modified to conform with the strategy. Effect such modifications by invoking process A3.3, "Update Digital Components." After the update, output the Conforming Digital Components. If the action plan cannot be implemented, or if it will not produce the desired result, refer the action to the "Manage Preservation" process as a request for a strategy decision. This request should trigger a revision in either the preservation strategy or action plan. When a revised strategy and/or plan is received, effect the necessary modification(s) as indicated above. If the result of a request for a strategy decision is not to modify either the strategy or	Given information about transferred records, determine if they can be preserved according to preservation strategies for the same records established on the basis of information received from appraisal or developed in the course of bringing in earlier transfers of the same records. If current preservation strategies are found to be applicable and adequate, determine if they entail any need to update any of the digital components of the records. If so, mark those components for updating in the 'Maintain' function. If not, report that the records can be accessioned. If currently applicable preservation strategies are found to be inapplicable or inadequate, report the problem(s) to the 'Manage' function, which will return a Preservation Action Plan. The

Activity Name	Activity Number	Activity Definition	Activity Note
		plan, output the problem components as Non-Conforming Digital Components.	Action Plan will indicate whether to reject the transfer, accept the problematic digital components 'as-is', or to mark those components for updating in the 'Maintain' function. If problematic components are to be accepted as in, the management plan will indicate if they are to be treated as exceptions to established preservation strategy or are to be associated with a new or revised strategy. [KT 4 2001] OAIS: "The Generate AIP function transforms one or more SIPs into one or more AIPs that conforms to the archive's data formatting and documentation standards. This may involve file format conversions, data representation conversions or reorganization of the content information in the SIPs. The Generate AIP function may issue report requests to Data Management to obtain reports of information needed by the Generate AIP function to produce the Descriptive Information that completes the AIP. This function sends SIPs or AIPs for audit to the Audit Submission function in Administration, and receives back an audit report." "The Generate Descriptive Information function extracts Descriptive Information from the AIPs and collects Descriptive Information from other sources to provide to Data Management. This includes metadata to support searching and retrieving AIPs (e.g., who, what, when, where, why) and could also include special browse products (thumbnails, images) to be used by Finding Aids." "The Coordinate Updates function is responsible for transferring the AIPs to Archival Storage and the Descriptive Information to Data Management. Transfer of the AIP includes a

Activity Name	Activity Number	Activity Definition	Activity Note
			storage request and may represent an electronic, physical, or a virtual (i.e., data stays in place) transfer. After the transfer is completed and verified, Archival Storage returns a storage confirmation indicating (or verifying) the storage identification information for the AIP. The Coordinate Updates function also incorporates the storage identification information into the Descriptive Information for the AIP and transfers it to the Data Management entity along with a database update request. In return, Data Management provides a database update response indicating the status of the update. Data Management updates may take place without a corresponding Archival Storage transfer when the SIP contains Descriptive Information for an AIP already in Archival Storage."
Accession Electronic Records	A2.4	Acting in accordance with the Preserver's Accessioning Policy, formally accept responsibility for preserving a transferred body of records that have been determined to be preservable. Create a Record of the Accession and forward the Accessioned Electronic Records to the 'Maintain' process. If Preservable Records are determined not to satisfy a requirement of Accessioning Policy not addressed in verifying the authority for the transfer or examining the records, the accession may be rejected. Requirements that might be applied in the accessioning process include resource constraints, where it might be determined that the Preserver cannot afford to preserve the records, or demands for or restrictions on access which the Preserver cannot satisfy.	
Maintain Electronic Records	A3	Following direction established in the preservation strategy for a given body of records selected for preservation, apply preservation method(s) targeted to	Old definition: Manage information about records, putting record components into storage, manage storage of record

Activity Name	Activity Number	Activity Definition	Activity Note
		that body of records to implement the preservation action plan for those records by maintaining the digital components of accessioned electronic records, along with related information necessary to reproduce the records, certify their authenticity, and enable correct interpretation of the records. This maintenance activity enables the output, in response to a retrieval request, of the digital components of a record, along with information about that record, or, if the request is only for information, the requested information The 'maintain' process also produces management information which is used to evaluate execution of the ingest function. The process is carried out by persons responsible for preservation, using infrastructure technology.	components, maintain the ability to reproduce records, and retrieve digital components.
Manage Information About Records	A3.1	The 'Manage Information About Records' process collects and maintains information necessary to carry out the Preservation Strategy for a body of electronic records being preserved, including information about their digital components, the archival aggregates they comprise, their authenticity, their interpretation, and the preservation activities performed on them.. Carrying out actions specified in the Preservation Action Plan, information about Accessioned Electronic Records is collected when they are accessioned. It is combined with Storage Information identifying the files, locations, and other relevant data about the digital components of the Accessioned Electronic Records when they are placed in storage and subsequently when storage parameters are changed. When a Preservation Action Plan entails any modifications to digital components, Information About those Digital Components is provided to ensure that all affected components are updated appropriately and, after the modification, Information about the Updated Digital Components is also updated. In response to a Retrieval Request for information, Retrieved Information About a Preserved Record is provided. In response to a Retrieval	

Activity Name	Activity Number	Activity Definition	Activity Note
		Request for a record, information identifying the digital components of the record and their storage location(s) is retrieved to produce a Request for Digital Components, which is used to retrieve those components from storage; Information About those Digital Components and Retrieved Information About the Preserved Record is output to support reproduction of the record and, if needed, certification of its authenticity.	
Maintain Information About Records	A3.1.1	As dictated by the Preservation Strategy, Information about Accessioned Records, and information providing the basis for asserting the authenticity of the records as transferred is collected and maintained over time. This information is kept up to date, by the input of Updated Storage Information, to reflect changes in the storage of the digital components of the records, for example, when storage media are replaced or defective files are recovered. Similarly, when changes in the applicable Preservation Strategy lead to updating of any digital components, the relateded information is updated through input of Information About the Updated Digital Components. This activity outputs Maintained Information About a Record and Maintained Information About Digital Components.	
Retrieve Information About a Requested Record	A3.1.2	Following the applicable Preservation Strategy, respond to a Retrieval Request for information about a record by outputting Retrieved Information About a Preserved Record. When the Retrieval Request is for the digital components of a record, output Information Identifying the Digital Components of a Record. Output the same information when a Preservation Action Plan requires updating of the digital components of a record.	
Retrieve Information About Digital Components	A3.1.3	In accordance with the applicable Preservation Strategy, upon receipt of Information Identifying the Digital Components of a Requested Record, retrieve Maintained Information About Digital Components and output the requested Information About Digital	

Activity Name	Activity Number	Activity Definition	Activity Note
		Components along with a Request for Digital Components to be retrieved from storage. Follow the same procedure when triggered by a Plan for Updating Digital Components.	
Manage Storage of Digital Components of Records	A3.2	In accordance with the Preservation Strategy established for a body of records, and applying the Storage Method selected to implement that Strategy, place the digital components of Accessioned Electronic Records into storage, taking the specific steps defined in the Preservation Action Plan for these records and maintain them. In response to a Request for Digital Components, retrieve the requested components and output them. When digital components are output for updating in accordance with a Preservation Action Plan, place the Updated Digital Components in storage and, as provided by the Action Plan, either maintain or delete the older versions of these components. Provide to the 'Manage Information' process Updated Storage Information about the identities, locations and other relevant parameters of stored digital components whenever components are updated or other changes, such as media refreshment, are made in storage.	Old Definition: Facilitate efficient, secure storage and provide ongoing protection of electronic records components in storage for subsequent retrieval. UBCS + The information about the records together with records components constitutes an AIP.
Place Record Components in Storage	A3.2.1	When electronic records are accessioned , place the Digital Components of the Accessioned Electronic Records into one or more Stored Digital Files in the storage system prescribed by the preserver's Storage Method, and provide Updated Storage Information about those components. When any of the stored components are updated in response to a Preservation Action Plan, place the Updated Digital Components into storage and either replace or retain prior versions of those components, as dictated by the applicable Preservation Strategy (tunneled to this diagram). Provide Updated Storage Information about the updated components and about any stored components that have been deleted or superseded. When a Stored Digital File is copied to new storage media, place the Refreshed File into storage, delete	Old definition: Move the physical files containing the electronic records components in an accession from where they were held during registration, examination and accessioning to the storage system where they will be maintained over time. Replace stored files which have been determined to have unrecoverable read errors. Replace storage media which have reached the end of their expected usefulness with new media containing copies of the same files. Replace stored files whose components have been migrated, reformatted to standard form, or converted to persistent form. OAIS: "The Receive Data function

Activity Name	Activity Number	Activity Definition	Activity Note
		the older copy of the file, and provide Updated Storage Information about the current location of the digital components stored in that file. When a problem discovered in a Stored Digital File is corrected, place the resultant Recovered File in Storage, delete the problem file, and provide Updated Storage Information about the current location of the digital components stored in the file and about any data loss or other residual problems with any of those components.	receives a storage request and an AIP from Ingest and moves the data to permanent storage within the archive. The transfer request may need to indicate the anticipated frequency of utilization of the data objects comprising the AIP to allow the appropriate storage devices or media to be selected for storing the AIP. This function will select the media type, prepare the devices or volumes and perform the physical transfer to the Archival Storage volumes. On completion of the transfer this function sends a storage confirmation message to Ingest including the storage identification of the AIPs."
Refresh Storage	A3.2.2	When triggered by a Plan for Updating Storage, use the prescribed Storage Update Method to copy a Stored Digital File to new storage media. When indicated by the plan, replace one or more components of the storage subsystem, ensuring that the Stored Digital File(s) involved in, or affected by, such replacement are carried forward without any inappropriate alteration. Document the process as part of the preservation history of all records whose digital components are contained in these digital files. Send any Refreshed File to be placed in storage and provide Updated Storage Information about any digital components affected by the process.	OAIS: "The Migrate Media function provides the capability to reproduce the AIPs over time. Within the Migrate Media function the Content Information and Preservation Description Information (PDI) must not be altered. However, the data constituting the Packaging Information may be changed as long as it continues to perform the same function. The migration strategy must select a storage medium, taking into consideration the expected and actual rates of errors encountered in various media types, their performance, and their costs of ownership. If media-dependent attributes (e.g., tape block sizes, CD-ROM volume information) have been included as part of the Content Information, a way must be found to preserve this information when migrating to higher capacity media with different storage architectures. Anticipating the terminology of section 5, this function may perform 'Refreshment', 'Replication', and 'Repackaging', but not 'Transformation'."
Monitor Storage	A3.2.3	In accordance with the Preservation Strategy	OAIS: "The Manage Storage Hierarchy

Activity Name	Activity Number	Activity Definition	Activity Note
		(tunneled to this diagram) for the storage system, apply the prescribed Monitoring Method to monitor the operation of the storage system, the media on which Stored Digital Files are recorded, the files themselves, and the facilities where the system and files are located. Provide Updated Storage Information about the problems identified and the stored digital components they affect.	function positions the contents of the AIPs on the appropriate media based on storage management policies, operational statistics, or directions from Ingest via the storage request. It will also conform to any special levels of service required for the AIP or any special security measures that are required and ensure the appropriate level of protection for the AIP. These include on-line, off-line or near-line storage, required throughput rate, maximum allowed bit error rate, or special handling or backup procedures. This function also provides operational statistics to Administration summarizing the inventory of media on-hand, available storage capacity in the various tiers of the storage hierarchy, and usage statistics." "The Error Checking function provides statistically acceptable assurance that no components of the AIP are corrupted during any internal Archival Storage data transfer. This function requires that all hardware and software within the archive provide notification of potential errors and that these errors are routed to standard error logs that are checked by the Archival Storage staff. The PDI Fixity Information provides some assurance that the Content Information has not been altered as the AIP is moved and accessed. Similar information is needed to protect the PDI itself. A standard mechanism for tracking and verifying the validity of all data objects within the archive may also be used. For example, cyclical redundancy checks (CRCs) could be maintained for every individual data file. A higher level of service such as Reed-Solomon coding to support combined error detection and correction

Activity Name	Activity Number	Activity Definition	Activity Note
			could also be provided. The storage facility procedures should provide for random verification of the integrity of data objects using CRCs or some other error checking mechanism."
Correct Storage Problems	A3.2.4	Upon notification of a Storage Problem, apply the Problem Correction Method prescribed by the Preservation Strategy (tunneled to this diagram) to take the actions indicated by the Plan for Problem Handling to eliminate the problem. If the Storage Problem affects any Stored Digital File, take action to copy the data stored in that file to a Recovered File and generate Updated Storage Information on each digital component affected by the process, including the identity of the Recovered File where the component is written, the success of the copy process, and any data loss or residual, uncorrected problem.	OAIS: "The Disaster Recovery function provides a mechanism for duplicating the digital contents of the archive collection and storing the duplicate in a physically separate facility. This function is normally accomplished by copying the archive contents to some form of removable storage media (e.g., digital linear tape, compact disk - recordable) but may also be performed via hardware transport or network data transfers. The details of disaster recovery policies are specified by Administration."
Retrieve Components from Storage	A3.2.5	In response to a Request for Digital Components, apply the Retrieval Method specified in the Preservation Strategy (tunneled to this diagram) to retrieve the Stored Digital File(s) in which the requested components are written from storage and output copies of the Retrieved Digital Components.	OAIS "The Provide Data function provides copies of stored AIPs to Access. This function receives a data request that identifies the requested AIP(s) and provides them on the requested media type or transfers them to a staging area. This function also sends a notice of data transfer to Access upon completion of an order."
Update Digital Components	A3.3	As indicated by the Preservation Strategy established for a given body of electronic records (tunneled to this diagram), take the steps indicated in the applicable Preservation Action Plan, applying the Method(s) for Updating Components prescribed by the strategy to update Digital Components of a Record That Cannot be Preserved because of technological obsolescence, changes in Preservation Strategy, or similar factors. Examples of update processes include migration, standardization, and transformation to persistent form. Return the Updated Digital Components to Storage, providing Information about the Updated Digital Components to the 'Manage Information' process. If	

Activity Name	Activity Number	Activity Definition	Activity Note
		the Updated Digital Components belong to a record that is the subject of a Retrieval Request, also send the components, along with related information, to the Output Electronic Record process. However, if the updating was done only to satisfy conditions of a Retrieval Request and was not required to conform to Preservation Strategy, the Updated Digital Components are sent, along with related information, to the Output Electronic Record process, but they are not sent to storage. This process may be invoked directly when records in a transfer are being examined and it is determined that there is a need to take action to preserve a record, before the components are sent to storage.	
Migrate Digital Components to Current Formats	A3.3.1	Following the migration schemes of the overall preservation strategy, and employing the preservation action plan for a record, migrate the digital components of a record that cannot be preserved as is to a specified current format in which they can be preserved. This activity will result in reformatted and updated digital components.	This preservation method counteracts digital format obsolescence by migrating record components to newer formats required for accessing and processing the components using outdated versions of the original software used to produce or store the components. It might also use functional equivalents of versions of the original software.
Convert Digital Components to Standard Formats	A3.3.2	Using the standardization strategy specified in the overall preservation strategy and the specified method for updating components, convert digital components of records that cannot be preserved as is to a standard format according to the record's preservation action plan. This activity will produce standardized digital components.	This preservation method aims at reducing the variety of problems involved in preserving electronic records by converting their formats from ones which presumably are dependent on specific hardware and/or software products to more generic, 'standard' formats; e.g. converting natural language components to plain text; converting databases to relational table form
Transform Digital Components to Persistent Format	A3.3.3	Using the persistent object strategy specified in the overall preservation strategy and the specified method for updating components, transform digital components of records that cannot be preserved as is to a persistent format according to the record's preservation action plan. This activity will produce	This preservation method aims at transforming electronic records components into forms that are independent of specific technology. In contrast to the standardization method, which reduces differences among records, persistent

Activity Name	Activity Number	Activity Definition	Activity Note
		persistent record objects.	object transformations rely on specifically expressing all the properties of records that must be preserved and explicitly capturing those abstracted properties.
Output Electronic Record	A4	Following direction established in the preservation strategy for a given body of records selected for preservation, apply preservation method (s) targeted to that body of records to implement the preservation action plan for producing an authentic copy of a records in response to a request for it. If specified in the request for a copy of the record, produce a certificate attesting to the authenticity of the copy. Alternatively, if requested, produce a reproducible electronic record; that is, the digital component(s) of the record along with instructions for producing an authentic copy of the record and information necessary to interpret the record. In the case of a request only for information about a record, deliver the response. To produce all of these outputs, translatesan external request for a record or for information about a record into a retrieval request to the 'maintain' function. Also produce management information which is used to evaluate execution of the 'output' function. The process is carried out by persons responsible for preservation, using infrastructure technology. The Output Electronic Record process may also be invoked when records in a transfer are being examined in order to verify that the records in the Transfer can be preserved and reproduced.	
Manage the Request	A4.1	Following provisions in the applicable Preservation Strategy, register an incoming Request for a Record and/or Information about a Record. Translate the request into terms that can be executed in the preservation system, and send the Retrieval Request to the Maintain Records process. Define Request Controls to ensure that the request is fulfilled and accounted for. These controls govern the process of responding to a request. If any problem is	Old definition: Register incoming requests, translate into preservation system terms and define controls to ensure that requests are fulfilled and accounted for. (08-29-01)

Activity Name	Activity Number	Activity Definition	Activity Note
		encountered in fulfilling the request, a Report of Problem with Retrieval Response is sent as feedback to this process. If the request cannot be satisfied, produce an Accounting for Unsatisfied Request and send it to the requester.	
Review Retrieved Components and Information	A4.2	Under the control of the applicable Preservation Strategy and Request Control, receive Retrieved Digital Components and/or Retrieved Information about a Preserved Record and determine whether all components and information necessary to satisfy the request for records have been received and can be processed for output. If the request entails producing a copy of a record, send the Requested Digital Components forward to Reconstitute and Present the Record. If a request for a record does not require reproduction of the record within the preservation system, send the Requested Digital Components forward to be packaged with related Information and delivered to the requester. If the request is only for information, output the Requested Information about a Preserved Record. If the request cannot be satisfied in accordance with the Request Control, produce a Report of Problem with Retrieval Response.	
Reconstitute Record	A4.3	Under the control of the applicable Preservation Strategy and Request Control, apply the appropriate Targeted Preservation Method to Retrieved Digital Components to link or assemble the components as necessary to reproduce the record and output the Requested Reconstituted Record. If the record cannot be reconstituted, produce a Report of Problem with Retrieval Response.	
Present Record	A4.4	Under the control of the applicable Preservation Strategy and Request Control, apply the appropriate Targeted Preservation Method to Retrieved Digital Components to the Requested Reconstituted Record to present the record with the appropriate extrinsic form. If requested, produce a Certificate of Authenticity for the Reproduced Electronic Record. If the process of reproducing the record is unsuccessful,	

Activity Name	Activity Number	Activity Definition	Activity Note
		produce a Report of Problem with Retrieval Response.	
Package Output	A4.5	Under the control of the applicable Preservation Strategy and Request Control, apply the appropriate Targeted Preservation Method to Retrieved Digital Components to combine Requested Digital Components with Information, including instructions on how to reproduce the record, into a package suitable for reproducing the record on an external system designated by the Requester. If the Request for a Record entailed reconstituting the record within the preservation system, package the Requested Reconstituted Record suitably for presenting the record on an external system designated by the requester. If the process is unsuccessful, produce a Report of Problem with Retrieval Response.	

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Preservation Task Force Arrow Definitions

Draft Version 5.1

September, 2001

Task Force Members:

Kenneth Thibodeau, U.S. National Archives and Records Administration, chair,
Richard Blake, Public Records Office, UK,
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

Arrow Name	Arrow Definition	Arrow Note
A 4 Output Records	A call arrow invoking the process for outputing electronic records, as delineated in A4	
A3.3 Update Digital Components	A call arrow invoking the process for updating digital components of records, as delineated in A3.3.	
Accessioned Electronic Records	A body of electronic records selected for preservation, transferred to the preserver and accepted by the preserver for preservation.	[KT 9 2001] Information necessary to preserve and reproduce the records, to assess their authenticity, and to certify the authenticity of reproductions of the records may be necessary as a condition for acceptance of the transfer, but need not be included in the definition of an accession.
Accessioning Policy	The policy of the institution or person responsible for preservation with respect to accepting responsibility for records transferred for preservation. Includes standards and specifications for acceptable and unacceptable deviations from standards, such as when records that should be in a transfer are missing or when information that should accompany the transfer is missing, inappropriate or unclear.	
Accounting for Unsatisfied Request	An explanation of why a Request for a Record and/or Information abut a Record could not be satisfied in whole or in part.	[KT 9 2001]
Archival Requirements	Requirements derived from archival science, diplomatics, best practices and prevalent standards within the archival community.	Benchmark and baseline requirements for authenticity are archival requirements
Authenticity Requirements for Preservation	Criteria that the preserver must satisfy and document to support an assertion that it has preserved and reproduced authenticy records, in conformance with the Baseline Requirements for Authenticity of preserved records.	
Basis of Authenticity of Transferred Records	The information that supports a presumption that records transferred to the preserver are authentic	[KT 9 2001]
Certificate of Authenticity	An attestation by the person responsible for preservation that one or more records are authentic.	Def. proposed to Glossary Cttee IPW 7 2/16/2001 Extensive def: A document, an attachment, or an annotation which attests to the authenticity of one or more records
Classes of Records	A class of records is a set of records with common attributes and methods determined on the basis of their documentary form.	
Conforming Digital	A conforming digital component is a digital component	

Arrow Name	Arrow Definition	Arrow Note
Components	which can be processed using current preservation methods in order to preserve and reproduce an electronic record.	
Conforming Transfer	A transfer of electronic records that satisfies the terms and conditions stipulated for the transfer.	
Digital Components of a Record That Cannot be Preserved	A Digital Component of a Record That Cannot be Preserved is a digital components of an electronic record that cannot be reconstituted or presented, or whose archival bonds cannot be expressed, or whose arrangement in archival aggregates cannot be reestablished using the Preservation Methods specified in the applicable Preservation Strategy.	
Digital Components of Accessioned Electronic Records	The digital components of the electronic records included in a transfer and accepted by the preserver for preservation.	[KT 9 2001]
Evaluated Preservation Options	Preservation Options that have been evaluated according to the institutional requirements of the preserver, the requirements for preservation specified for a body of records selected for preservation, and the evaluation of execution of preservation.	
Evaluation of Execution	The result of evaluating management information about the execution of preservation process preservation to determine whether requirements were satisfied, preservation strategies are effective, and action plans meet objectives and performance targets.	[6/21/2001] arrow named changed from "Evaluation" at PTF workshop 8-28-01
Facilities	Locations where digital preservation technologies are installed and operates, and locations where electronic records components are stored.	
Information About Accessioned Records	Information identifying records that have been accessioned, their digital components, and the preservation strategies that will be applied to them.	
Information about Digital Components	Metadata or other information retrieved in response to a request for a record or for information about a record, or in response to a plan for updating the digital components and used to satisfy the request or to carry out the plan.	
Information about Digital Components of an Electronic Record	Technical information concerning a digital component or a class of digital components of electronic records necessary to store and retrieve the digital components of an electronic record and to reconstitute the record from the	[KT 9 2001]

Arrow Name	Arrow Definition	Arrow Note
	components and to present it in authentic form, properly ordered with respect to related records.	
Information about Electronic Records Selected for Preservation	Information output from the appraisal process identifying and characterizing records which are to be preserved, including what information about the records should accompany the transfer, the basis for asserting the authenticity of the records as maintained by their creator, and the terms and conditions of transfer.	[revised KT 9 2001] ApTF Def: A record or records providing the necessary information about electronic records to maintain them continuously in authentic form, including the terms and conditions of transfer.
Information About Preservation	Reports on, data about, or accounting for the exercise of the preservation function, the preserved records, and the authenticity of copies of those records. The coverage of such information may range from the preservation function as a whole, to one or more processes within the function.	
Information about Preservation Technologies	Information about methods that are available withing the state of the art of information technology and that could be used to preserve electronic records.	
Information about Transferred and Accessioned Records	Information about electronic records, their arrangement, and their digital components, which accompanies the transfer of electronic records and which is developed in the process of bringing them under the preserver's control.	[6/21/2001] This information is input to to the managemenet processes of determining archival preservation requirements, selecting preservation methods and specifying preservation strategies where it supplements, updates, and as needed corrects information about the same records accumulated or developed in the process of selecting them for preservation. [6/21/2001] Information about previously accessioned records needs to be input to these processes under three conditions: 1. When there is a change in archival requirements affecting the records; for example, when additional transfers of records from the same archival fonds have characteristics that require modifications in the archival requirements for preserving the records; 2. When Evaluation determines that previous characterizations of archival requirements for preservation were erroneous or inappropriate; or 3. When changes in applicable preservation methods entail a change in the classification of digital components. [KT 4 2001] This information is analyzed in the processed leading to the specification of preservation strategies for the records, and some of it will be embedded in the strategies. In contrast,

Arrow Name	Arrow Definition	Arrow Note
		information that is produced in the evaluation process acts as a control on the other management processes. If, for example, evaluation identifies deficiencies in selected preservation methods, the selection process must act on this determination.
Information about Updated Digital Components	Information about changes that have been made to a digital components of an electronic record in the process of updating it, about any problems that occurred in the process, and about the identity and location of the component on storage media and in a storage system or facility.	[KT 9 2001]
Information and Communications Technology	Digital hardware, software, and storage and communications media	Digital information and telecommunications technology is obviously a mechanism necessary for the preservation of electronic records. It is shown in this context diagram as a mechanism supporting the entire process. However, as shown in decomposition diagrams A1, and A1.2, technology is actually acquired in the management process which selects from the available technologies those which will be used for preservation of records. This selection of Preservation Methods entails the acquisition of the Technology Infrastructure necessary for the implementation of the Preservation Methods and the execution of the preservation processes.
Information Identifying Digital Components of a Requested Record	Information which specifies all the digital components necessary to reproduce a record and their unique identifiers.	[KT 9 2001]
Institutional Requirements	External legal, regulatory, societal, and cultural constraints imposed on the institution responsible for preserving records, together with the preserver's internal standards, policies, procedures, goals, objectives and criteria applicable to records in general or of electronic records specifically. An external legal, regulatory, societal, and/or cultural constraint imposed on the institution responsible for preserving records, and/or an internal policy, procedure, goal, or objective, applicable to the preservation of records in general or of electronic records specifically. [Proposed to Glossary Cttee 2/2001 at W7]	{Revised KT 9 2001] Need to correlate with Appraisal Task Force & Domain IV. Definition should also reflect specifics articulated through decomposition of the arrow at leaf levels.
Maintained Information About	Information supporting the preservation, reproduction and	

Arrow Name	Arrow Definition	Arrow Note
a Record	interpretation of a record being preserved and the certification of the authenticity of a copy of that record	
Maintained Information About Digital Components	Information identifying a stored digital component, and the record(s) that comprise that component, and information enabling the reproduction of the record(s)	[KT 9 2001]
Management Information About Preservation	Information about an electronic record that is transferred for preservation and about the processes of bringing in, maintaining and reproducing the record. This information enables management to evaluate those activities and the preservation strategies and plans they implement, and to determine what changes may be needed in requirements or strategies.	[6/30/01] [singularized 9 2001, MS]]
Mapped Records and Digital Components	A mapping of information about electronic records reportedly transferred for preservation to the digital file(s) that were transferred, identifying and locating the digital components included in the file(s), linking them to the records that they constitute, and determining whether the information about the records and the digital components in the transfer is appropriate and sufficient for preservation and reproduction of the records.	
Method for Updating Components	A Targeted Preservation Method used to migrate, transform or otherwise modify digital components of electronic records in accordance with the preservation strategy applicable to those records.	
Migration Strategy	A preservation strategy entailing the reformatting of one or more classes of digital components, defined by data type, to a new storage representation.	[KT 9 2001]
Monitoring Method	A preservation method for determining whether a storage system is properly maintained and functioning or whether storage media are intact and free from problems that would interfere with reading the data written on the media.	[KT 9 2001]
Non-Conforming Digital Components	A non-conforming digital component is a digital component on an electronic record in a format that prevents or impedes the reproduction of the record in accordance with the applicable preservation strategy.	[KT 9 2001]
Notification of Receipt	A record sent to the submitter acknowledging that the preserver has received the transfer and, if needed, asking the submitter to address any problems identified in registering the transfer.	Stipulated in OAIS

Arrow Name	Arrow Definition	Arrow Note
Packaging Method	A Targeted Preservation Method used to wrap or encapsulate either the Digital Components of an electronic record or a Reconstituted Electronic Record in a form that will enable reproduction of the record on a designated target system, and to combine the resultant digital object with instructions for reproducing the record on the target system.	[KT 9 2001]
Persistent Object Strategy	A strategy for preserving electronic records by making explicit the specific attributes and methods of the records that must be preserved and imposing those specifications on the maintenance or transformation of the digital components of the record and on the process of reproducing the record.	[KT 9 2001]
Persistent Record Objects	Electronic record components, or proxies for the arrangement of electronic records in an archival set of records in a form that is independent of any specific hardware or software.	
Persons Responsible for Preservation	Persons authorized and charged with carrying out the preservation function or processes within that function.	[6/21/2001]
Plan for Problem Handling	An action plan that specifies what to do if a problem is encountered in carrying out a preservation strategy.	[KT 9 2001]
Plan for Updating Digital Components	A preservation action plan that entails executing a process to change a digital component of an electronic records, for example through migration or persistent object transformation, in order to enable the record to be reproduced in accordance with the currently applicable preservation strategy.	[KT 8 2001]
Plan for Updating Storage	A plan established by the Manage Preservation function to update one or more components of the storage system or to move physical files to new digital storage media.	
Presentation Method	A Targeted Preservation Method used to present an electronic record in appropriate documentary form.	
Preservable Records	An electronic record comprised of digital components which all conform to the applicable preservation strategy.	[singularized 04.09.2001, MS] Pre-singular: Records all of whose digital components conform to applicable preservation strategies.
Preservation Action Plan	A plan for one or more preservation actions to be taken for the transfer of a record to the archives, in accessioning the record, or for a record being maintained. The plan is formulated in accordance with the Preservation Strategy	[singularized 9-4-01, MS]

Arrow Name	Arrow Definition	Arrow Note
	applicable to the record in question.	
Preservation Options	Technological possibilities for performing one ore more processes necessary to preserve a given body of records, or any subset of that body, or class of digital components in that body.	[KT 9 2001]
Preservation Strategy	A coherent and comprehensive approach for preserving a body of records selected for preservation, derived from archival and institutional requirements, taking into account Evaluation of Execution of current and prior Preservation Strategies and reflecting the State of the Art of Information Technolody. A Preservation Strategy includes objectives for maintaining components of electronic records and related metadata and information over time, and for reproducing the records in authentic form, in the order imposed by the records creator, and criteria for evaluating execution of the Preservation Strategy. The strategy includes specifications for handling exceptions to its standards, and identifies the targeted preservation methods to be used.	[KT 9 2001] Definition modified 9 2001 to eliminate inclusion of preservation methods in strategy, because they are shown as a separate output of activity A1.2 Old definition: A coherent and comprehensive approach for preserving a body of records selected for preservation, including methods for maintaining components of electronic records and related metadata and information over time, and for reproducing the records in authentic form and in the order originally imposed by the records creator, and criteria for evaluating execution of the preservation strategy. [as proposed on 8-29-01]
Preservation Technology Specifications	Information about a Targeted Preservation Method	
Problem Correction Method	The method stipulated in a preservation strategy for correcting problems of a specified type.	[KT 9 2001]
Record Composition Requirements	Stipulations as to how a record is to be reconstituted from the digital components that contain its content and/or specifications about the arrangment of the content within the record	[KT 9 2001]
Record of Accession	A record documenting the preserver's acceptance of responsibility for preserving a set of electronic records.	[KT 9 2001]
Record Reconstitution Method	A Targeted Preservation Method used to assemble or link the Digital Components of an electronic record to enable its presentation in appropriate documentary form.	[KT 9 2001]
Recovered File	A physical or logical file which has been successfully read, or reconstituted, and placed back into appropriate storage after one or more storage problems affecting the file were found.	
Refreshed File	A physical or logical file which has been copied from an older storage medium or system to a newer when the older medium or system.	singularized 9-5-01, MS

Arrow Name	Arrow Definition	Arrow Note
Registered Transfer	A transfer is determined as authorized if and only if it comprises a record that has been selected for preservation and that the record has been submitted either by the record's creator or an agent for the creator.	[singularized 9/4/2001, MS] Pre-singular: A transfer is determined as authorized if and only if it comprises records that have been selected for preservation and that records have been submitted either by the records creator or an agent for the creator.
Registration Procedure	The procedures and recordation required by the preserver to identify and track a transfer of electronic records.	[KT 9 2001]
Rejected Accession	A set of electronic records selected for preservation and transferred to the person responsible for preservation, but not accepted for preservation by the preserver.	
Rejected Transfer	A transfer of electronic records which does not satisfy requirements for being accessioned or preserved.	If a transfer is rejected as a result of specific criteria established in appraisal, information about the rejection should be sent to the Appraisal Function as well as the submitter. KT 12 Feb 01 At A-0, this arrow is an unresolved output because in A2 it has been changed to external reference to submitter. KT 12 Feb 01 Pre-singular: A set of transferred electronic records which does not satisfy requirements for being accessioned or preserved.
Report of Problem with Retrieval Response	Information describing why a request for a record or for information about the record cannot be satisfied in whole or in part.	[KT 9 2001]
Report on Authenticity of Records	An account of preservation activities related to a given record or aggregate of records to support the assertion or certification that reproductions of the record(s) are authentic.	
Reproduced Electronic Record	An authentic representation or other version of a record reconstituted from its digital components, along with information supporting the interpretation of the record.	(PTF 08-29-01) [singularized 09-04-01, MS] In this context, a version of a record is a reproduction which is not authentic.
Reproducible Electronic Record	Digital components of an electronic record, the technical information necessary to reproduce the record from the digital components, information about the authenticity of the record, criteria for certifying the authenticity of the reproduced record, and information that supports interpretation of the record.	(KT 09-17-01)
Request Control	A specification of the deliverable(s) to be produced in response to a request for a record or for information about a record, and the criteria to determine whether the request has been satisfied.	

Arrow Name	Arrow Definition	Arrow Note
Request for Digital Components	An instruction to retrieve the digital components of a record.	[KT 9 2001]
Request for Record and/or Information about Record	A request to output a preserved record, or to provide information about a record that cannot be found in or derived from archival description or finding aids, but requires either retrieval and processing of the preserved record or of information that is created or specifically maintained in the preservation system.	[singularized 9-4-01, MS]
Request for Strategy Decision	A request for a management decision formulated when it is determined that an electronic records cannot be preserved in accordance with the applicable preservation strategy because one or more digital components of the electronic record do not conform with the specifications of that strategy. The request asks for a determination of whether the preservation strategy should be changed or, alternatively, the component(s) should simply be determined to be non-conforming digital components.	[KT 9 2001]
Requested Digital Components	All of the digital components necessary to satisfy a request for a record.	
Requested Information about a Preserved Record	The information provided in response to a request for information about a preserved record.	[KT 9 2001]
Requested Reconstituted Record	The digital components of a requested electronic record linked or reassembled to enable reproduction of the record.	[KT 8 2001]
Requirements for Arranging Records	Stipulations as to how the original order of records is to be respected in the physical or logical structuring of sets or archival aggregates of records.	[KT 9 2001]
Requirements for Instantiating Archival Bonds	Stipulations as to how archival bonds are identified, and how they are to be instantiated when a set of related records is retrieved and reproduced.	[KT 9 2001]
Retrieval Method	The preservation method to be used to retrieve the digital components of an electronic record from storage.	[KT 9 2001]
Retrieval Request	A request for retrieval of a record and/or information about the record.	[singularized 9-4-01, MS] Pre singular: A request for a retrieval of records and/or information about the records. [PTF meeting 8-28-01]
Retrieved Digital Components	The digital components of an electronic record retrieved from storage in response to a request.	[KT 9 2001]
Retrieved Information about a	Information retrieved from storage in response to a	

Arrow Name	Arrow Definition	Arrow Note
Preserved Record	retrieval request.	
Selected Preservation Method	A preservation method selected as best satisfying archival and instutional requirements for one or more classs of preservation objects. The domain in which a Selected Preservation Method applies will include one or more bodies of records, classes of records, types of digital components, types of arrangement of records, methods for expressing archival bonds, and/or presentation methods.	[KT 9 2001] Need to define class of preservation objects for IP Glossary
Specified Requirements for Preservation	Complete and consistent requirements for preserving a specified body of records in a manner which will enable reproduction of the records and certification of the authenticity of the reproduced records.	[7/2001]
Standardization Strategy	A preservation strategy entailing the migration of a class of digital components to a standard format.	[KT 9 2001]
Standardized Digital Components	Digital components converted to a standard format supported in a preservation strategy of standardization.	
State of the Art of Information Technology	The state of the art of the technology with respect to its ability to satisfy archival preservation requirements, the state of the underlying computer science with respect to its ability to develop relevant capabilities not within the state of the technology, and the existence and prevalence of applicable standards.	
Storage Method	A Targeted Preservation Method used to place and maintain digital components in storage, to retrieve them from storage, or to remove them from storage.	[KT 9 2001]
Storage Problem	A problem with storage media, storage formats, a storage system or facility that could impact on the continued preservation of records.	
Storage Update Method	A method used to ensure that stored digital components are completely and correctly brought forward when any component of a storage subsystem is changed or when digital files are moved or migrated to newer storage media.	[KT 9 2001]
Stored Digital File	A digital file placed in a storage system on digital media.	[KT 9 2001]
Targeted Preservation Method	Software used to implement a preservation strategy or strategies. A preservation method is targeted by specifying the bodies of records, types of electronic records, and/or classes of digital components to which it will apply, along with conditions for its application to each	[9 2001] The scope of a target may range from all digital objects being preserved to a single type of record, class of digital component, or form of presenation.

Arrow Name	Arrow Definition	Arrow Note
	target group.	
Technological Infrastructure	Hardware and common services software, such as operating systems, communications software, data base management systems, electronic mail applications, etc., necessary for the implementation of preservation methods and the execution of the preservation processes defined in the Preserve Electronic Records model.	[KT 9 2001]
Terms and Conditions for Transfer	The specifications governing the transfer of a body of electronic records selected for preservation to the preserver. These specifications include, at a minimum, when the records should be transferred, whether the entire body of records should be transferred at one time or incrementally, what information should accompany the transfer, label and format of transfer files, and what physical medium or telecommunications channel should be used to effect the transfer. [KT 9 2001]	May be considered as a special type of preservation action plan. [KT 9 2001]
Transfer of Electronic Records Selected for Preservation	The digital components, the accompanying information related to preservation and reproduction of the related record, and transmittal information.	[Proposed to Glossary Cttee. at W 7 2/2001] [singularized 04.09.2001, MS]
Types of Archival Bonds	A type of archival bond is a specific manner in which the archival bond of a record is implemented or indicated	
Types of Record Arrangement	A type of record arrangment is a logical or physical method of expressing the ordering of records within an archival aggregate of records.	[KT 9 2001] For example, a series of files may be ordered alphabetically by subject or numerically by identifier, or chronologically. Each of these three is a type of arrangement.
Updated Action Plan	A preservation action plan modified as a aresult of an assessment of its effectiveness in achieving the objectives of a preservation strategy or because of a change in the strategy.	[KT 9 2001]
Updated Digital Components	An updated digital component is a component that has been modified under a preservation action plan.	Must include information about updated components.
Updated Information about Electronic Records Selected for Preservation	Revised information identifying and characterizing records which are to be preserved.	[KT 9 2001]
Updated Preservation Technology Specifications	Information about a Targeted Preservation Method that has been modified or discontinued, or that has been selected to replace another method.	
Updated Storage Information	Information indicating a change in the location of a digital	[KT 9 2001]

Arrow Name	Arrow Definition	Arrow Note
	component in storage, the occurrence of a storage problem, the action taken to correct a storage problem, the results of such action, or the copying of digital files from older to new storage media.	
Updated Strategy	A revised preservation strategy	[KT 9 2001]

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

HOW TO PRESERVE AUTHENTIC ELECTRONIC RECORDS

Draft

July 25, 2001

Task Force Members:

Kenneth Thibodeau, U.S. National Archives and Records Administration, chair,
Richard Blake, Public Records Office, UK,
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

HOW TO PRESERVE AUTHENTIC ELECTRONIC RECORDS

DRAFT

Report of the InterPARES Preservation Task Force

July 25, 2001

1 Introduction

This report communicates the results of the work of the InterPARES Preservation Task Force. The Preservation Task Force was chartered to identify and develop the procedures and resources required for the implementation of the conceptual requirements articulated in the InterPARES research on authenticity and appraisal of electronic records[2] To achieve this goal the Preservation Task Force formulated and analyzed the problem of preserving authentic electronic records in order to articulate a detailed, in-depth understanding of this problem. The Task Force followed two principal paths in articulating this understanding: (1) an international survey of current practices and plans in the preservation of electronic records and (2) a formal modeling of the function of preserving electronic records. The American InterPARES Research Team independently contributed to the Task Force's work by developing a bibliography on digital preservation[3]and a report on digital storage media.[4]

The analysis has produced a functional model of the processes necessary to preserve electronic records selected for preservation, a model of the information needed to support the preservation function, a glossary defining the terms in the models, and a report on a survey of current practices and plans in the preservation of electronic records.

Address empirical bases, depth and precision of analysis, expertise of members.

1.1 Acknowledgements

To be developed

[2] http://www.interpares.org/researchplan.htm
[3] http://is.gseis.ucla.edu/us-interpares/bib_pres.htm
[4] P. C. Hariharan. Media, A Presentation for the InterPARES Panel at UCLA. December 2, 1999.

2 Electronic Records

Any approach to the preservation of electronic records has to start with clear recognition of the basic characteristics of such records, to what extent these basics are the same as those of traditional records, and how they differ.

2.1 Recording Information

Recording information enables it to be transmitted across time and space and between or among persons or organizations. Any recording of information requires some way to represent that information on a physical medium. Textual information, for example, is represented through the use of alphabets or character codes, punctuation marks, abbreviations, fonts or handwriting styles, page layout, etc. When this information is recorded on paper, ordinarily there is absolutely no difference between the way the message is represented on the medium and the way it is presented to humans for interpretation and use. But when the information is recorded digitally, there is an inevitable difference between the way it is represented on a medium and the way it is presented for use. In part, this difference derives from the fact that different types of media are used for storage and presentation; for example, digital information is typically stored on magnetic or optical media, but displayed on CRT or LCD screens. More important, however, is the basic difference between the digital encoding of information in binary values for storage, transmission and processing by computers and the translation of that encoding into a form that can be used by humans.

> *Whether a textual document is stored digitally as a scanned image of a paper document or in an character-based representation, such as ASCII or Unicode, it is necessary to transform that representation into a very different one to make the document readable. For example, in character mode, every single character must be presented in the chosen font, but in storage font may be indicated only by a special code that precedes an entire block of text to which it applies.*

It is not possible to store an electronic record in the form of a record. The 'form of a record' is the documentary form which enables it to achieve its intended purpose in the first place and to communicate the same information over time. Electronic records are stored in forms which substantially differ from those in which they can serve their intended purpose as records.

> *This report was created using word processing software on a PC. Essential to its form as a record is the visual presentation of natural language text in lines, paragraphs, and sections. But this report cannot be stored, in the computer's memory, on its hard drive, or on any digital medium, in this form. It can only be stored as one or more sequences or strings of bits. Each textual character (letter, space, and punctuation mark) in this report is stored as a sequence of eight bits. Special sequences of bits are used to indicate breaks in the flow of text, such as the separation of paragraphs and section headings, and the*

indentation of this paragraph, and other presentation features, such as different type sizes, bold, and italics.

The difference between the way digital information is represented in storage and the way it is presented for use occurs even when the use is within or between computer systems and does not involve humans.

> When an individual uses a check card to make a purchase, the store's computer system reads the magnetically encoded data on the check card to identify the customer's bank and account. The stores computer then sends a message to the bank's computer asking for transfer of the amount of the sale. The bank's computer then checks it checking account database to see if the requested funds are available and, if so, records the transaction, deducting the sale total from the account balance. Then it notifies the store's computer, which creates its own record of the transaction. In this single transaction, there are several transformations of the way the information involved is represented digitally. Transformations occur, for example, when the computer reads the magnetic strip on the check card; when the store sends the message to the bank over a communications channel; when the bank's computer receives the message and stores it in its memory; when the computer invokes the checking account database; when the database management system invokes its optimizer; when the optimizer interacts with the storage subsystem to request and retrieve the account data, etc. Then there are additional transformations when the sales information is displayed on the cash register and printed as receipts for the store and customer.

Differences between the storage and use versions of electronic records are not limited to the way the data is inscribed on phsyical media. There may also be substantial differences between the units in which digital information is stored, that is digital files, and the units in which that information is organized for use. There is no necessary relationship between digital files and the archival units of record, file of records, record series, etc. A single record could be stored in one or more digital files. A single digital file could contain one or more records, and a single digital file, such as one containing the specifications for the layout of a form or report, could be used in thousands of records. In business applications which rely on database systems, the database often consists of several thousand logical files each of which contains part of the contents of potentially millions of records.

Often the digital file in which an electronic record is 'stored' does not in fact contain the entire record. Ordinarily, if a document was created using, say, word processing software and stored as a single digital file in the native format used by the software, we consider that this file is or contains the record. However, in current technology, the word processing file is insufficient to reproduce the record as it was meant to appear. Data which is essential to correct rendering of the presentation features of an electronic

record is usually stored outside of the file, or files, which contain the content of the record.

> *In order to present this report, or any word processing file, on a video display device, it is necessary to pull in data from one or more other files. While the word processing file contains bytes representing all of the text characters in the document, the word processing application needs to use data that the Windows operating system stores in separate files called dynamic link libraries, or 'dll' files, in order to display these characters in the font or fonts used by the writer. Such font files are used with all user created files where the pertinent fonts are selected. From an information technology perspective, the font files are extensions of the Windows operating system, rather than parts of the user created files. Nonetheless, from an archival perspective, a font file must be treated as a digital component of any record which relies on the font for proper presentation. While some font files may contain bitmapped images of each character, most often the 'dll' file contains data about the characters that the software uses dynamically to synthesize or form the characters for display. Similar processes occur for other types of content, such as graphics, and in other types of applications which contain such data types.*

In sum, the relationship between archival units and digital files may be one-to-one, one-to-many, many-to-one, or many-to-many.

Differences between storage representations and use presentations do not only occur among different records and digital files. The same presentation of a record can be produced from a variety of storage representations. Conversely, a single storage representation may be processed to output a variety of presentations.

> *One technique that is used to ensure the fixity of both the content and the appearance of digital documents is to convert them from word processing files to formats, such as "portable document format" or "pdf" files, which include the data necessary to form font characters properly, eliminating dependence on dynamic link libraries.*
>
> *While the data content of a record may be stored in many different tables in a relational database, the data must be brought together in order to present the record in its proper documentary form. Once the record is reconstituted from its different data components, it may be possible, and in fact simplest, to save it as a single digital file, for example as a word processing file, without in any way impacting its identity or integrity.*

Because there is no necessary mapping between digital files and records, in many cases it is possible to change the way the record is stored in digital files without changing the record itself.
The inevitable, and often repeated, changes between the storage representation and the presentation for use of digital information creates paradoxical elements in the preservation of electronic records.

2.2 Keeping Electronic Records

When a recording of information is intended to serve as a *record* of an action or state of affairs, it is essential that the message it transmits be fixed. An authentic record is one that is what it purports to be and that is free from tampering or corruption. Determining that it is what it purports to be means establishing its identity. Determining that it is free from tampering or corruption means demonstrating that its integrity remains intact through space and time. In the case of records on paper and other 'hard' media, the authenticity of a record over time rests on the assumption that the physical object that embodies the record has not changed in any way that would affect the message it was intended to communicate. Thus, in traditional archival practice, preference was given to the 'original' record and to an 'unbroken chain of custody.' The 'original' is preferable because any process of copying it introduces an opportunity for alteration. The principle of the unbroken chain of custody stipulates that, throughout their lifecycles, records should be in the custody of known parties who can be trusted to preserve them intact. Continuous custody is important because any break in control over the record also creates risk of deletion, alteration or substitution; furthermore, any discontinuity in custody may make it impossible to demonstrate that a record has not been altered.

Probably the most basic aspect of preserving records on hard media is placing and keeping them in storage. Ideally, the environment in which the records are stored should not include any elements which would damage the records or cause them to deteriorate; furthermore, if needed, the environment should be designed to reduce or retard any deterioration that is intrinsic to the physical media on which the records are stored or to the physical means used to inscribe the records on the media, for example, by controlling temperature and humidity. Active conservation measures are taken to prevent or recover from any damage or deterioration. The goal is to justify faith that a record retrieved from storage is the same in all essential respects as the record previously placed in storage. Traditionally, the preservation of records has focused on ensuring their fixity through the processes of conservation and maintenance, where maintenance refers to keeping records in places and under conditions which protect them from harm and minimize or reduce any innate tendencies towards deterioration, and conservation refers to interventions to repair damage and to prevent deterioration which has a high risk of occurring.

Keeing electronic records is more complex and difficult. Like all records, an electronic records must transmit the message intended by its creator; however, the fixity of the message carried by an electronic record is at risk because of the changes between the way it is represented in storage and the way it is presented for use. Both placing an electronic record in storage and retrieving it for use entail transforming the way the content, structure and appearance of the record are inscribed on a physical medium. Both storage and retrieval transformations create risks that the record may be altered. These risks are compounded by the software, hardware, and the physical media used, including both the storage media and the media on which the records are presented in record form. While maintenance and conservation of the stored information remain essential, the integrity of an electronic record depends on guaranteeing that none of the

changes between storage representation and presentation for use, in either direction, has altered the message the record was intended to convey from the time it was first filed as a record to the point of any subsequent use. The necessity of ensuring that transformations between storage and use do not corrupt the records adds a new focus to the preservation of electronic records. An electronic record cannot be said to have been preserved unless it can be delivered in authentic form.

The process of preserving electronic records extends over the entire lifecycle of the records from creation to disposition and, in the case of records that are preserved for posterity, to the reproduction of those records. The overall process of preservation must be continuous. If there is ever a point where we cannot reasonably assert that the record continues to carry its original message intact, we can never thereafter assert that it is authentic. It is important to recognize that while the process must be continuous over time, the activities that constitute the process are discrete steps. Each instance where the way the information is represented changes – whether moving between storage and use or between storage media or subsystems – is a potential point of failure, a weak link where the entire chain could be broken. The process of preserving electronic records extends to and includes interactions between computer systems and human users and interoperations between computer systems, subsystems and applications.

2.3 Foundation Concepts

This review of differences between traditional and electronic records leads to the articulation of several basic concepts about the preservation of electronic records which distinguish it from preservation of traditional records.

2.3.1 Digital Components of Electronic Records

The most basic concept demanding attention in preserving electronic records is that, in addition to all the intrinsic and extrinsic elements of form that make up any record, an electronic record also comprises one or more digital components. A digital component of an electronic record is a digital object that *is*, or *is part of*, an electronic record, or that *contains* one or more parts of one or more electronic records, and that has *specific methods* for storage and reproduction. The complexity of this definition derives from the unlimited cardinality in the relationships between electronic records and their digital components. While difficult to define, the concept of digital component is relatively easy to describe and to grasp empirically.

Every electronic record has at least one digital component: a stream of bits representing information contained in the record. In this simple case, the record and the digital component are congruent: the component *is* the record. However, the contents of a

record may be stored in several bit streams; for example, a single record may be stored as a compound document, with different parts of the record stored in different digital files. In such cases, the record has as many digital components as it has bit streams. A bit stream may also contain data which indicates how the information content is to be presented for use; for example codes indicating fonts, type sizes, line endings, paragraph indentations, etc. However, codes indicating presentation features may also be stored in separate bit streams, such as dynamic link libraries, report templates, or style sheets. Each bit stream indicating presentation features is an additional digital component of the record. When a record has more than one component, each digital component *is part of* the record.

These examples suggest that digital components are distinguished from one another based on the fact that they are stored separately. While that may be true, it is not always the case. A record may consist of many digital components, but several components may be stored together in one physical file. If a physical file contains all of the components of an electronic record, it *contains* the record. If a physical file contains some, but not all of the components, it *contains parts of* the record. The basis on which digital components are distinguished, then, is not simply physical storage. It would be more accurate to say that each component is a logical or physical object that the system processes as an unit. In other words, each component has a specified method or *methods*. In fact, when several components are stored together in a physical file, there must be specified methods for locating and extracting the components. Thus, in accordance with the definition of digital component, the physical file is itself a component.

The containment relationship among components may be based on factors other than physical storage.[5] A textual record may contain non-textual content, such as a spreadsheet, an image or even a voice annotation. In this case, it does not matter whether the different types of content are stored together or separately. Whether the units of non-textual content constitute separate digital components depends on the format or data-type used to represent the contents. In a word-processing file, a picture would constitute a distinct digital component because the word-processing content is character-type data while the picture is binary or raster data. However, if the textual record was a scanned document image, both the picture and the text would be binary image data and constitute only a single digital component. Thus, data-type appears as a second criterion for distinguishing digital components. However, this criterion needs to be refined. A word processing file containing two digital photographs has only two data-types, but three digital components. A word processing document and a spreadsheet both used character data, but each requires different software for proper processing and presentation. Conversely, alphanumeric characters and linear graphics are two different types of data, but most word processing applications are capable of vertical and horizontal lines; therefore, a word processing document containing such lines has only one digital component determined by data-type. Data-type is a criterion for

[5]The containment relationship may be either physical or logical, or both. A digital file may contain one or more distinct digital components stored entirely within that file (physical), or it may only contain links pointing to distinct components stored elsewhere (logical).

distinguishing digital components of records if and only if different data-types have different *methods* associated with them.

As this discussion shows, there is no necessary relationship between the elements of form of an electronic record and its digital components. In fact, in some cases the relationships between the record and its digital components can be changed without significant impact on the record as such. For example, a textual record originally created as a word processing file could be changed to a binary image or portable document format without impacting any of its essential characteristics as a record. However, changes in the digital components of a record could corrupt the record. Therefore any such changes need to be under preservation control.

2.3.2 Preservation Control

A technological boundary exists between any two states of a system or of interoperating systems when the transition from one state to another does, or can, entail significant changes in the attributes or methods of a digital object. For records, significant changes are those which affect identity or integrity. Technological boundaries exist at macro and micro levels. Macro level boundaries occur at the interfaces between systems, subsystems or applications, such as during system, media, or data format migrations or in transfers between the 'live' systems in which the records are created, and other applications in which they are transmitted over space or stored over time. Micro level boundaries occur when a record is decomposed into separate digital components or is reconstituted from its components, and when different methods are invoked to process distinct components. Transitions from storage representation to presentation for use can involve both macro and micro boundaries.

Preservation control is critical in transitions across technological boundaries. Preservation control consists of actions, conditions, and constraints designed to ensure the preservation of records and their continued authenticity. While preservation controls during maintenance of the records in storage must be adequate and effective, the risks of corruption or loss of records are more frequent and complex during transitions across technological boundaries. Thus preservation controls can be divided into two types. Steady-state controls are those which ensure that records remain unchanged, either in storage, in active systems, or in communications. Dynamic controls are those which ensure that records remain authentic across technological boundaries.

Preservation control, in most instances, will be accomplished through technical means, but it must be determined according to archival principles and criteria.

2.3.3 Archival Requirements for Preservation

Naively, preservation may be seen as a process which keeps records free from change. However, it can be easily shown that it is practically impossible for any record to remain absolutely unaltered or immutable over time. More importantly many changes that occur naturally or accidentally do not impact the authenticity of the record. Paper darkens. Ink fades. Microfilm scratches. While such changes may indicate threats to the continued existence of the records, they do not necessarily make them inauthentic. One of the best known records in the world is the original of the Constitution of the United States of America. The paper has yellowed and the ink has faded considerably over the years. Facsimiles of this document have been produced which arguably look more like the document originally did than the original currently does; however, no facsimile or other copy can ever approach the value persistently attributed to the original. Hundreds of thousands of people regularly visit the National Archives Building in Washington to view the original record. In fact, the physical changes in the original evince its authenticity because they result from the operation of the laws of nature: the original record should not look as it did two hundred years ago.

The requirement for records to remain unchanged, thus, needs to be qualified. Even in the case of hard copy records, this requirement effectively means that the record should not be changed in any way that relates to its essential record nature, rather than to its existence as a physical object. For electronic records, this qualification has been aptly stated in the InterPARES report, Requirements for Ensuring the Authenticity of Electronic Records Over Time: "When we refer to an electronic record, we consider it essentially intact and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered."[6]

The archival requirement for integrity, and therefore for authenticity, depends on the message intended by the record creator. The interpretation of a record depends on the reader as much as the creator; therefore it is beyond the control of the record preserver. Whatever interpretation is made, it must be consistent with the creator's intended communication, and that undeniably involves the information content of the record, and it may involve the way the content is presented. Here again, neither requirement is absolute. If an accident, such as water damage, resulted in a few words in a document being lost or blurred (content or presentation), we would say that the record is damaged, but not that it is inauthentic. The strength of the requirement for unaltered content and presentation depends on the intended use of the record. The requirement is greatest when the user wants to see the 'original,' regardless of interpretation. But for many valid uses, it is sufficient if the content – of all or part of the record – is substantially intact and unaltered and its presentation is basically consistent with the original.[7] If this were not true, it would be impossible for historians, political scientists, and others to cite *original* sources in analytic works.

[6] InterPARES Authenticity Task Force. Requirements for Assessing the Authenticity of Electronic Records. July 2001.

[7] In this case, 'original' refers to the state of the record at the moment of its creation, rather than to a durable physical object.

As stated in the Requirements for Ensuring the Authenticity of Electronic Records, the requirements for authenticity of copies of records depend on the purposes for which the copies are made. In the case of electronic records, all access after the records have been stored is to reproductions of the records.

2.3.4 'Original' Electronic Records

The transformations entailed by storage, retrieval and presentation of electronic records make the concept of 'original record' of uncertain applicability in the domain of electronic records. The traditional concept of an original record is tightly coupled with its inscription on a specific physical medium. The same record inscribed on any other unit of physical medium is not considered to be an original, but a copy. Given that electronic records are not stored even on the same type of medium used to present them to humans,[8] and that the physical inscription of the record on a storage medium is fundamentally different that the inscription on a display device, a strict application of the traditional concept of original record would mean that the original ceases to exist at the moment it is committed to storage and deleted from the video device the writer used to create it.[9] Therefore, the closest we can come to an electronic original, once the record has been set aside, is a copy in the form of the original. Given that electronic records are not stored in their original form, to produce a copy in the form of the original we need to maintain information about that form and also about the methods which are needed to translate between the storage representation and the presentation for use.

In the digital environment, there is an important distinction to consider in the concept of 'the form of the original.' With records in hard copy, the form of the original is effectively the form in which the record was inscribed on a medium, because in inscribing the record the writer fixed the information content in a determined form. The inscription expresses the writer's intent. With electronic records, concern is often expressed about preserving the 'look and feel,' that is the presentation features, of the record; however, there are elements of extrinsic form that the writer cannot fix in an immutable form, but be changed at whim by any user.

> *Depending on the software used, simply changing the size of the window in which a document is viewed can change properties such as character size or line*

[8] E.g., magnetic and optical storage media are drastically different than cathode ray tubes or LCD displays.

[9] Even in the narrow context of the writer keeping the record displayed on the same device, the traditional view of an original record cannot be sustained for electronic records. Without changing any of the binary data in the record – including both the informational content of the record or special codes indicating presentation features – the writer can effect signficant changes in the appearance of the records by a variety of simple steps, such as changing the size of the window in which the record is displayed, changing the magnification within the window, or switching between draft and page-oriented views of the record. The impact of such actions on the presentation of the record may be beyond the writer's control. For example, with some software changing window size changes the flow of text across lines, while with other software it changes the type size. In hard copy, substantial changes in the appearance of a record can only be achieved by producing new and different copies of the record, but in a digital environment such changes involve only altering the way the record is present, not altering anything 'in' the record.

length. Similarly, changing the magnification or 'zoom' ratio, or switching between draft and print image modes in word processors alters the appearance of a document on a screen.

Unless there is evidence that the writer intended to fix the presentation of the record, changes, such as window size or magnification, would not be regarded as producing a different record, especially when the software that permits such changes also makes it easy to reverse them. Such variations do not alter a record, as such, any more than viewing it under a magnifying glass changes a record on paper.

2.3.5 The Need to Reproduce Electronic Records

The transformations entailed by storing, retrieving and presenting an electronic record lead to the recognition that, in literal terms, you cannot preserve an electronic record, you can only preserve the ability to reproduce the record. A logical corollary to this assertion is that the only real way to prove that an electronic record has been preserved is to reproduce it. While the production of copies is usually seen as part of the archival reference or communication function, in the case of electronic records it is also within the scope of preservation. These functions overlap at the point of reproduction. The emphasis of the reference function is that the copies produced respond to the interests and requests of users, while the preservation function emphasizes the production of certifiably authentic copies.

Reproducing the record involves both its intrinsic and extrinsic elements of form. With respect to the bit streams that are maintain in storage over time, an electronic record contains one or more digital components. The first step in reproducing an electronic record is to reconstruct it by assembling all of its digital components in the proper arrangement. The second step is to present or render each of the components individually and all of the components collectively in the proper documentary form. The final process in reproducing the record is to reestablish its immediate context. The immediate context of a record is its archival bond: the relationship between a record and other records. This is a two step process. The first step consists of reestablishing the structure of the set of records in which the record belongs, the second step is to populate that structure with the relevant records.

2.3.6 The 'Chain of Preservation'

An electronic record 'in storage' is simply not the same as it was either before being stored or after retrieval. To justify faith that an electronic record retrieved from storage is the same in all essential respects as the record previously placed in storage, the rationale that is applied to hard copy records; namely, that a physical object has been under continuous control that prevented it from being altered is not sufficient. Given that the storage and retrieval processes for electronic records inevitably entail physical

and representational transformations, the traditional concept of an unbroken chain of custody needs to be expanded to encompass the processes that are necessary to ensure that an electronic record is transmitted over time without inappropriate alteration. This expanded concept can be called the unbroken chain of preservation: the entire process of committing an electronic record to storage, maintaining it in storage, retrieving, and presenting it must adequately preserve all its essential attributes in order to support a credible claim that the retrieved electronic record is authentic. In addition to what is entailed in the chain of custody, the chain of preservation will include information about the records creator's practices to support a presumption of authenticity, in accordance with the benchmark requirements for authenticity, information about the processes of bringing the records into the archives and maintaining them over time, and information about the reproduction of records, in accordance with the baseline requirements for the production of authentic copies of records. In the digital environment, where records are not affixed in stable fashion to durable media in the forms in which they are presented for use, it is necessary to invoke an additional principle: the unbroken chain of preservation. It is not sufficient to hold on to the records. We must also ensure that any action that affects the way the records are represented or presented protects their integrity.

2.3.7 Preservation, Conservation and Maintenance

These concepts reflect a substantial departure from prevailing thinking about digital preservation. Most attention in this area has been focused on overcoming technological problems of obsolescence and media fragility. The focus of the Preservation Task Force is not on dealing with technological problems, but on achieving the positive objective of transmitting authentic electronic records over time and generations of technology. It is this objective, and the archival requirements attendant to it, which define the parameters and criteria for selecting among technological alternatives and evaluating the success of approaches and actions for preserving the records. While it is possible to compare the merits of different approaches to obsolescence and media fragility from a purely technological perspective, technology alone cannot determine what is the best choice simply because what is best depends primarily on the purpose for which an action is taken, not on the method of acting.

Within this view, steps taken to counteract obsolescence and media fragility may be termed 'conservation actions' and measures taken to avoid or minimize the effects of obsolescence and media fragility may be regarded as maintenance activities. Conservation and maintenance are parts of preservation. However, preservation activities are not limited to solving or avoiding technological problems. Preservation actions – such as media migration, storage system updating or replacement, use of different software, and even changing the data formats in which the digital components of the records are stored – may be taken not only to solve problems but also simply because new and better alternatives have been developed in information technology.

3 The Process of Preserving Electronic Records

3.1 The "Preserve Electronic Records" Model

The InterPARES Preservation Task Force has focused its efforts on the articulation of a formal model of the process of preserving electronic records. The process model was articulated using the Integration Definition (IDEF) methodology adopted by the International Team. Specifically, the model was articulated in accordance with the IDEF(0) standard for function modeling.[10]

In IDEF(0), "A function model is a structured representation of the functions, activities or processes within the modeled system or subject area." 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. It should be interpreted as describing a subject area, rather than a specific system. That is, 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 roadmap which 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.

The basis for the content of the preservation process model is the Open Archival Information System (OAIS) Reference Model, which is currently a draft ISO standard.[11] "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

[10] Draft Federal Information Processing Standards Publication 183. Integration Definition For Function Modeling (IDEF0). December 21,1993.

[11] Consultative Committee for Space Data Systems. Reference Model for an Open Archival Information System (OAIS). Red Book. May 1999. http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf

information objects comprising electronic records and archival aggregates of such records.

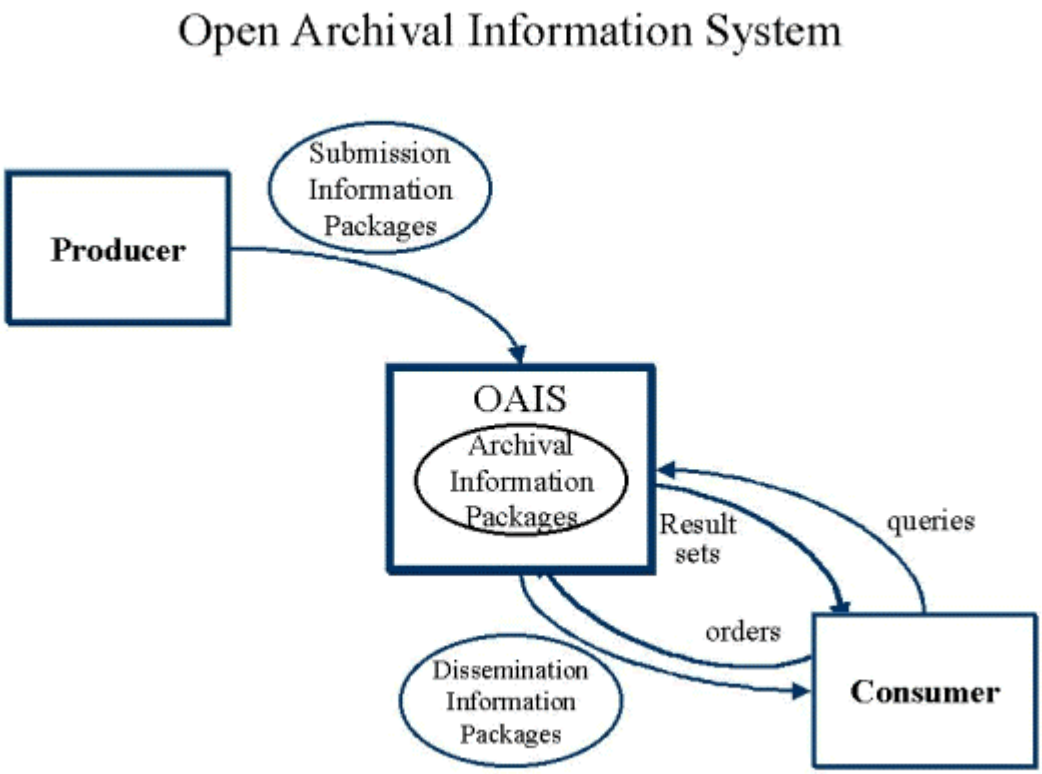


Figure 2. Open Archival Information System

Here again it is necessary to distinguish between the function described by the 'Preserve Electronic Records' model and a system which would implement that model. The preservation function might be carried out by a system which provides only the functionality described in the model. But it might equally well be implemented in a system which includes 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, but 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.

The boundaries of the preservation function model derive from the viewpoint according to which the model is constructed. IDEF(0) models "functions (actions, processes, operations), functional relationships, and the data and objects." 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 important in one viewpoint may not even appear in a model presented from another viewpoint of the same subject."

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 which have been selected for preservation after they are no longer needed for the practical purposes for which they were originally created. The scope of the InterPARES project corresponds to that of the OAIS box in the center of figure 1. 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 which 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 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. 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, according to the concept of preservation as described in section 2 above.

In contrast to the OAIS model, the viewpoint of the 'Preserve Electronic Records' model only includes those aspects of submission and preservation that relate directly to transforming Submission Information Packages into Archival Information Packages, and it only includes those aspects of dissemination which relate to reproducing electronic records or providing requesters with the wherewithal to reproduce the records themselves. While the OAIS model includes determining what will be submitted to the OAIS and who is the designated customer community, these activities are beyond the scope of the 'Preserve Electronic Records' model.

The viewpoint largely determines the relationships between the appraisal and preservation models. Naively, one may assume that preservation follows appraisal because records must be selected for preservation before they are preserved; however, the relationship between the appraisal and preservation models is not that of a simple sequence, but rather reflects two different viewpoints on the same overall archival process. Each of the two models includes activities which do not appear in the other, but such activities may be related through their inputs, outputs, or controls. For example, the selection of what records are to be preserved is not itself a preservation activity; therefore, selection does not appear as a process in the preservation model. However, the records selected for preservation are a major input to the preservation model. Conversely, maintaining the digital components of the selected records in storage is not a selection activity and does not appear in the appraisal model. The preservation function selects methods for preserving records comprising different types of digital components. This selection determines feasibility of preserving different types of components; therefore, it acts as a control on the appraisal function.

There are activities which appear in both appraisal and preservation models. For example, disposition of records is modeled as an appraisal function. While the term, disposition, does not appear explicitly in the preservation model, the transfer of records to the preservation system is a disposition action which is included in the preservation model. Similarly, establishing the terms and conditions for transfer of records is a feedback loop between the appraisal and preservation models.

Following its charge, the preservation model is intended to articulate the procedures and resources required for preserving authentic electronic records. However, the requirements for preserving authentic electronic records were being developed by the Authenticity and Appraisal Task Forces, working in parallel to the Preservation Task Force. Because these task forces' results were not available, the preservation function model has been articulated up to this time so as to be neutral with respect to requirements for authenticity. The requirements for ensuring that the preserved records remain authentic can, and should be incorporated in the model. This is work that needs to be done in the future.

3.2 Preservation Overview

Figure 2 is the context diagram for the 'Preserve electronic records' model. It does not show any detail about the process itself. Instead it shows what goes into and comes out of the process. Three different kinds of things go into the process: Controls, which govern how the process is carried out, Mechanisms, which enable the process to happen, and Inputs, which are the things acted on in the process. Following IDEF(0) convention, in all diagrams, controls are shown going into the top of a process box; mechanisms are shown going into the bottom of the process box; and inputs enter the process from the left. What comes out of any process are its Outputs. In IDEF(0) Outputs always come out of the right side of a process box.

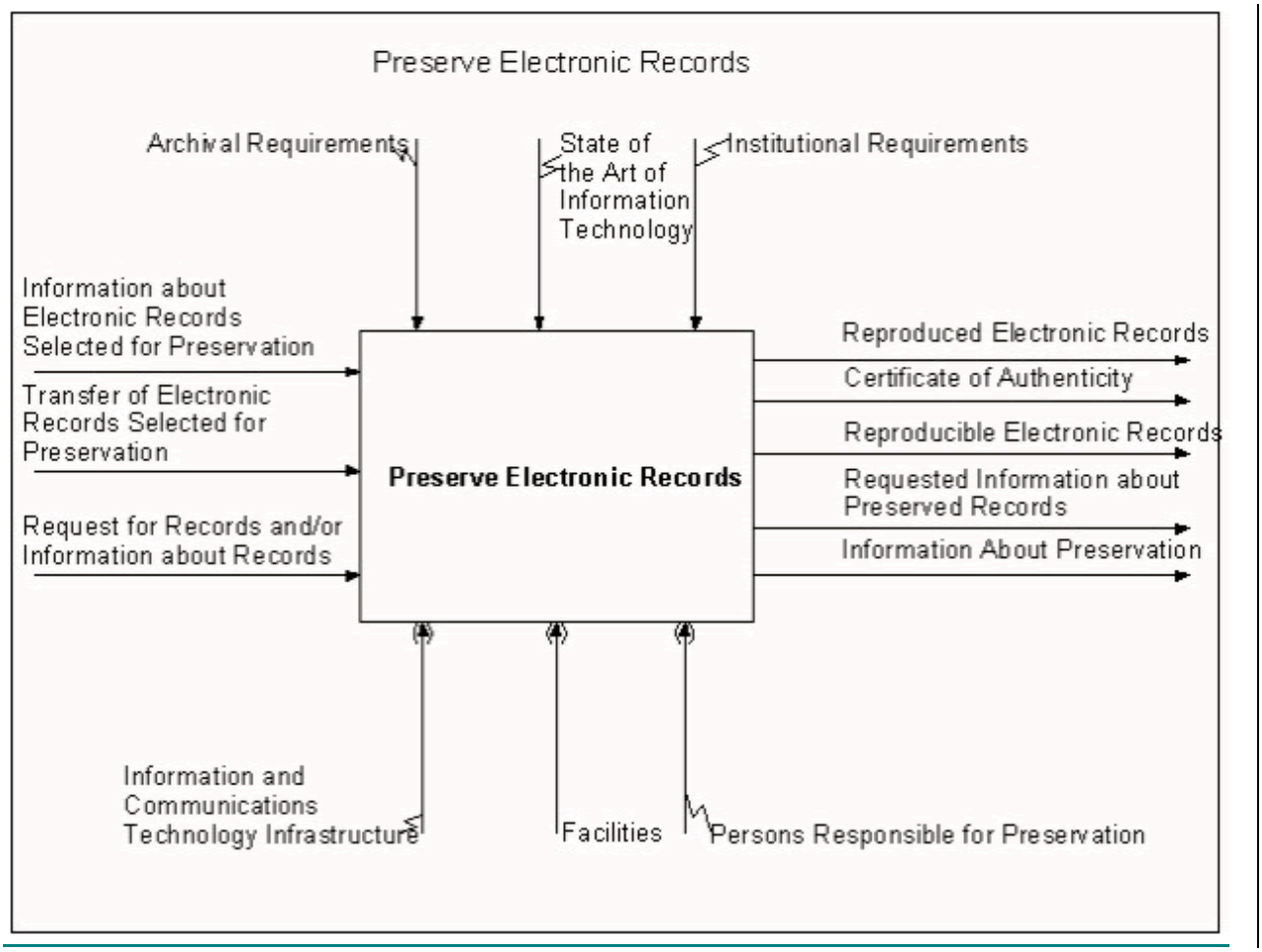


Figure 3. Preserve Electronic Records Context

The context diagram shows that three things control the process of preserving electronic records. In order to preserve records, and especially to preserve them as authentic, we need to know what are the requirements for doing so. These requirements derive from archival science and principles and related standards and best practices for managing records. These requirements are labeled as 'archival requirements in the diagram. Preserving electronic records entails using digital information technology. The possibilities for doing so are limited by the state of the art of information technology, which constitutes the second type of control on the preservation process. Finally, the exercise of the preservation function will also be governed by requirements of the institution in which this function is carried out.

The diagram shows three mechanisms that are necessary to perform the preservation process. They are an information and communications technology infrastructure, facilities where the electronic records will be stored and processed, and persons responsible for the process. While the state of the art of technology determines what is possible and impossible to do, the technology infrastructure comprises the hardware, software and physical media used to store and process the digital components of electronic records. The brackets surrounding the points on the three mechanism arrows indicate that these mechanisms are used in all preservation activities; therefore, they are not shown in the more detailed diagrams that follow below.

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 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 authentic.

3.2.1 The Main Preservation Processes

Preserving electronic records involves four processes: managing the preservation function, brining records into the preservation system, maintaining them over time, and outputting them. These processes are depicted in figure 3. This diagram is rather complex, but can be easily understand by tracing the basic path that each of the inputs follows.

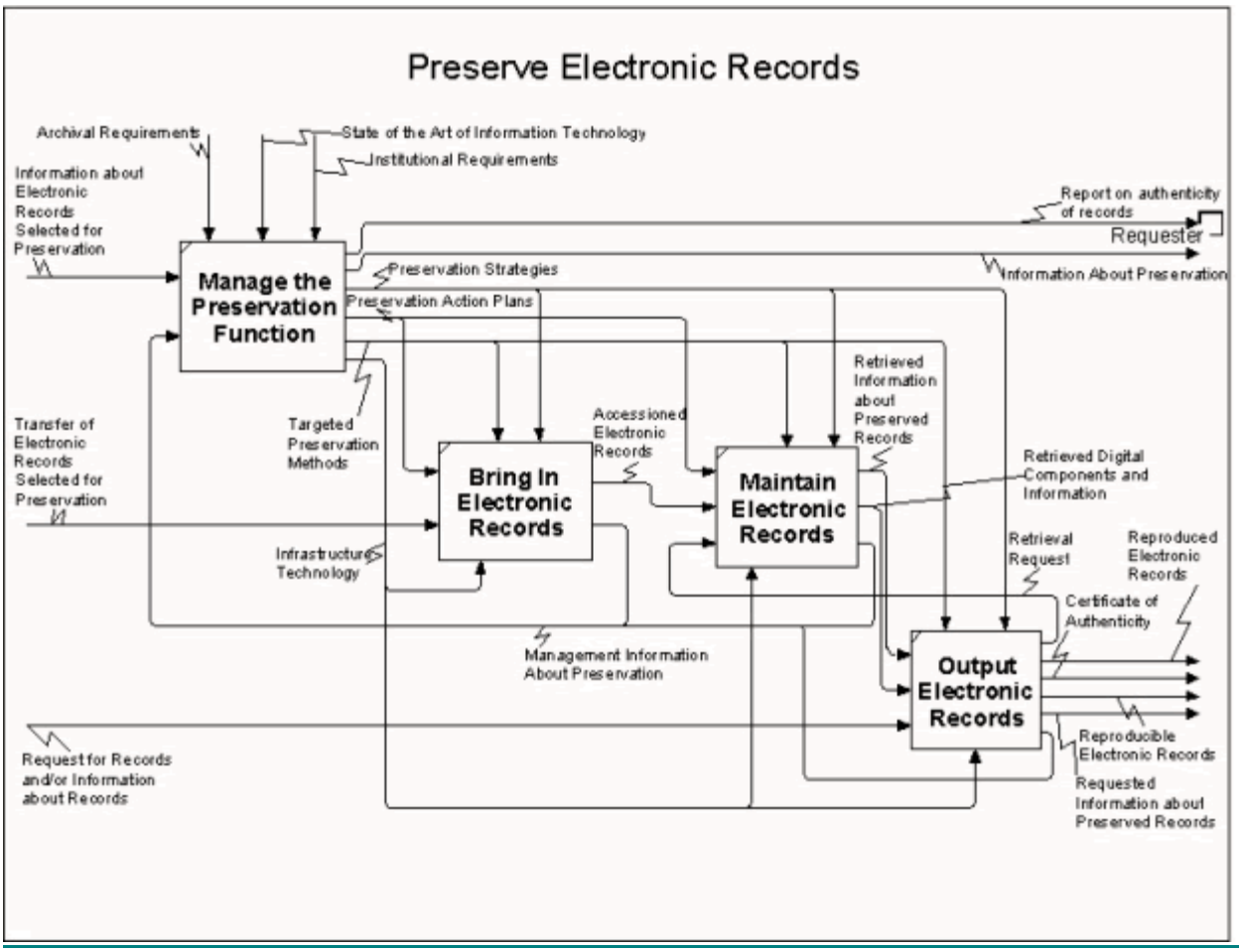


Figure 4. The Main Preservation Processes

The first input, in chronological sequence, is Information about Electronic Records Selected for Preservation. It is input to the process named,'Manage the Preservation Function.' This management process is unique in that it controls the other three basic processes. The management process takes what must be done and determines how it should be done and what the results should be. The other three processes carry out preservation activities according to the parameters established by management. The management process has the same basic relationship to all three execution processes, as depicted in figure 4.

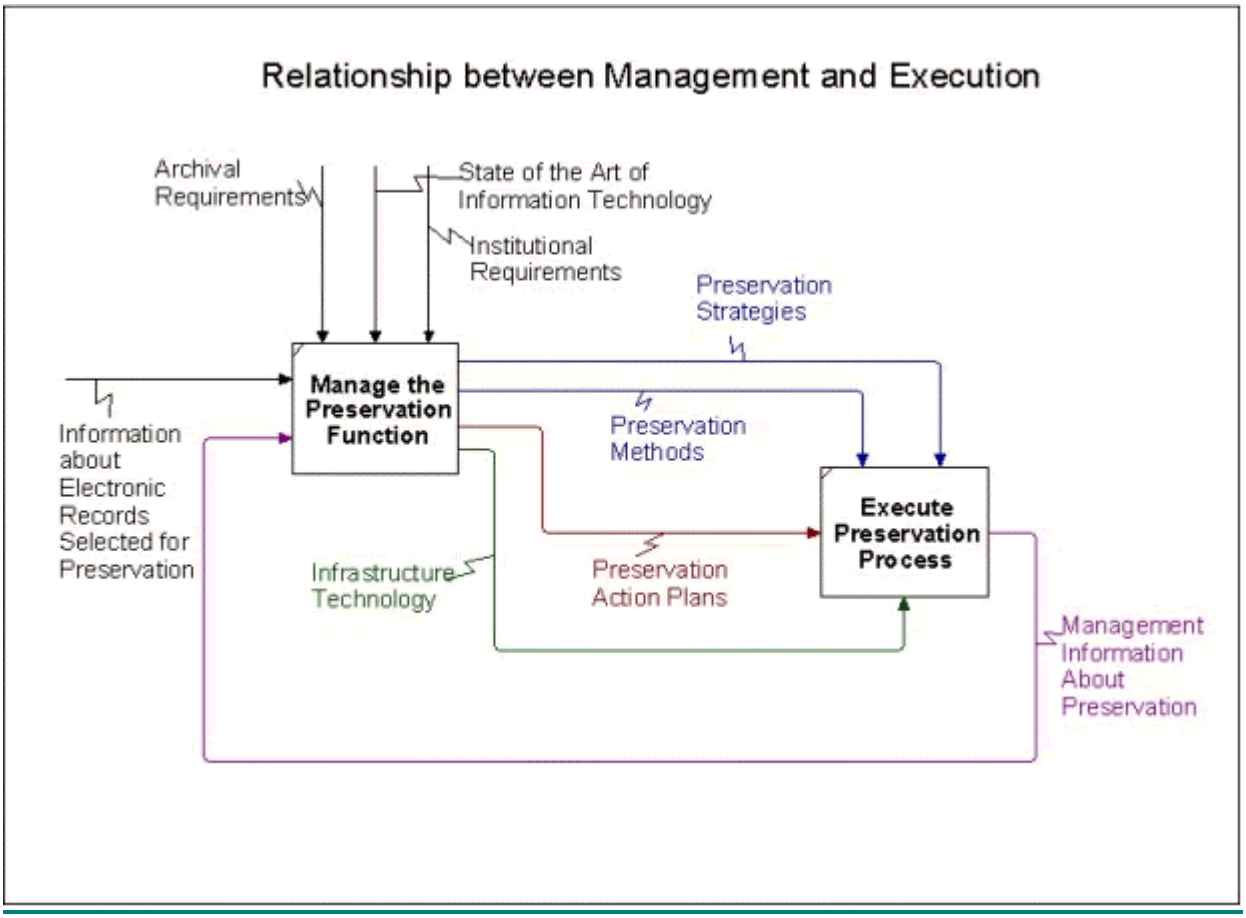


Figure 5. Management and Execution of Preservation Processes

Managing the Preservation Function: The management process synthesizes the external controls to determine how preservation should be accomplished and what the results should be. In each case of records selected for preservation, it articulates the archival and technical requirements for preserving the records taking into account their documentary form as records, their archival bonds, their digital components, and the requirements for producing authentic copies of the records; specifies procedures; selects and acquires technology appropriate for satisfying these requirements; and establishes criteria for documenting the preservation process and for determining if the process has been carried out successfully. The ensemble of specific requirements, means of meeting the requirements, and objectives in doing so constitute a preservation strategy for the records selected for preservation. The preservation strategy is thus an output of the management process which functions as a control on each of the execution processes. The preservation strategy for a body of records will specify software to be used for all processing of the records. This software constitutes preservation methods and also controls each execution process. The preservation methods require information technology infrastructure which includes hardware, media, and general purpose software, such as operating systems, database management systems, storage subsystems, and communications protocols. The management

function selects and acquires the information technology infrastructure which is used in carrying out the three execution processes.

The management function also sets up preservation action plans for actions that must be done in each execution process either at specified times or under specified conditions. For example, preservation action plans should specify how to determine, in each case, if the terms and conditions of transfer, established when the records were appraised, have been satisfied and what to do if not. The preservation action plans are input to each of the execution processes.

Each execution process should output management information about the process. This information is sent as feed back to the management function to enable it to evaluate the execution of preservation processes and determine if preservation strategies or methods, or infrastructure should be changed. The evaluation will also determine the feasibility of preserving different types of electronic records. This determination will be used during the appraisal process.

This description of the relationships between the management and execution processes applies to all three execution processes. It will not be repeated in the description of those processes, thus simplifying their descriptions.

Bringing Records Into the Preservation System: The second input to the preservation process comprises actual Transfers of Electronic Records Selected for Preservation. The transfers are input to the 'Bring in Electronic Records' process, as shown in figure 5. This process determines whether the transferred records are accessioned or rejected. If they are accepted, they are sent to the 'Maintain Electronic Records' process. If rejected, they are returned to whomever submitted them. In either case, information about the transfer is sent to the management function where it is combined with information received from appraisal about records selected for preservation in developing or modifying preservation strategies.

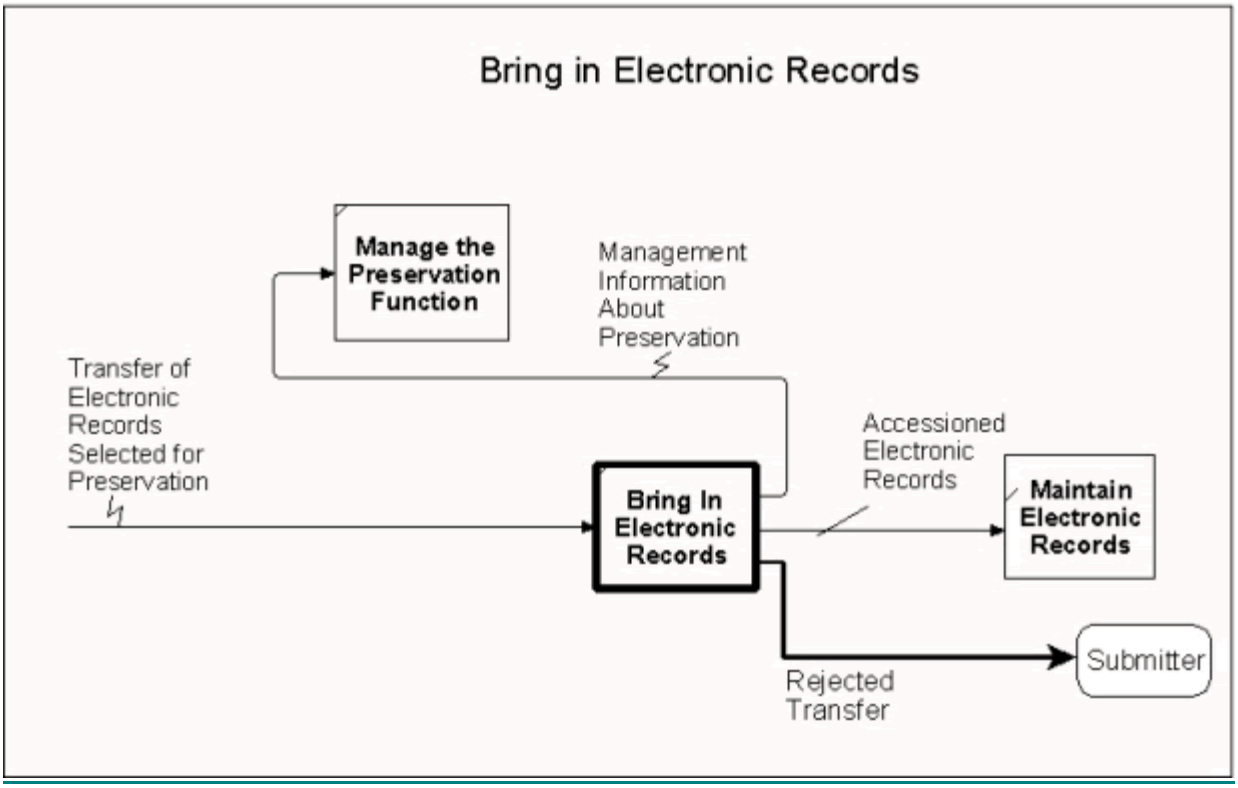


Figure 6. Bringing Records into the Preservation Process

Reproducing Electronic Records: The third input to the 'Preserve Electronic Records' process consists of Requests for Records and/or for Information about Records. Responding to such requests is the ultimate objective of preserving records; moreover, as explained in chapter 2, producing copies is the final step in the process of preserving electronic records. Given that the model is constructed from the viewpoint of the person responsible for preserving the records, this final step is included in the model.

However, the model does not include all facets of responding to requests for records or information. It is assumed that certain preliminary steps, and possible subsequent ones, related to the requests are taken by persons responsible for access to the records. For example, the persons responsible for the access function will help requesters to identify the records or information in which they are interested, and will determine if they have a right to receive the records or information they request. As with the appraisal function, there will be some overlap between the preservation and access functions. If a model of the reference or access function were articulated, it would include which coincide or overlap some of the activities within the 'Bring In' process because the examination of the records and related information during that process would provide the opportunity to develop the information needed for producing finding aids. Similarly, the reference function would probably determine the computer interface used to provide requesters with access to electronic records, but preservation would determine how the records need to be presented in that interface in order to guarantee that the reproduced records are authentic.

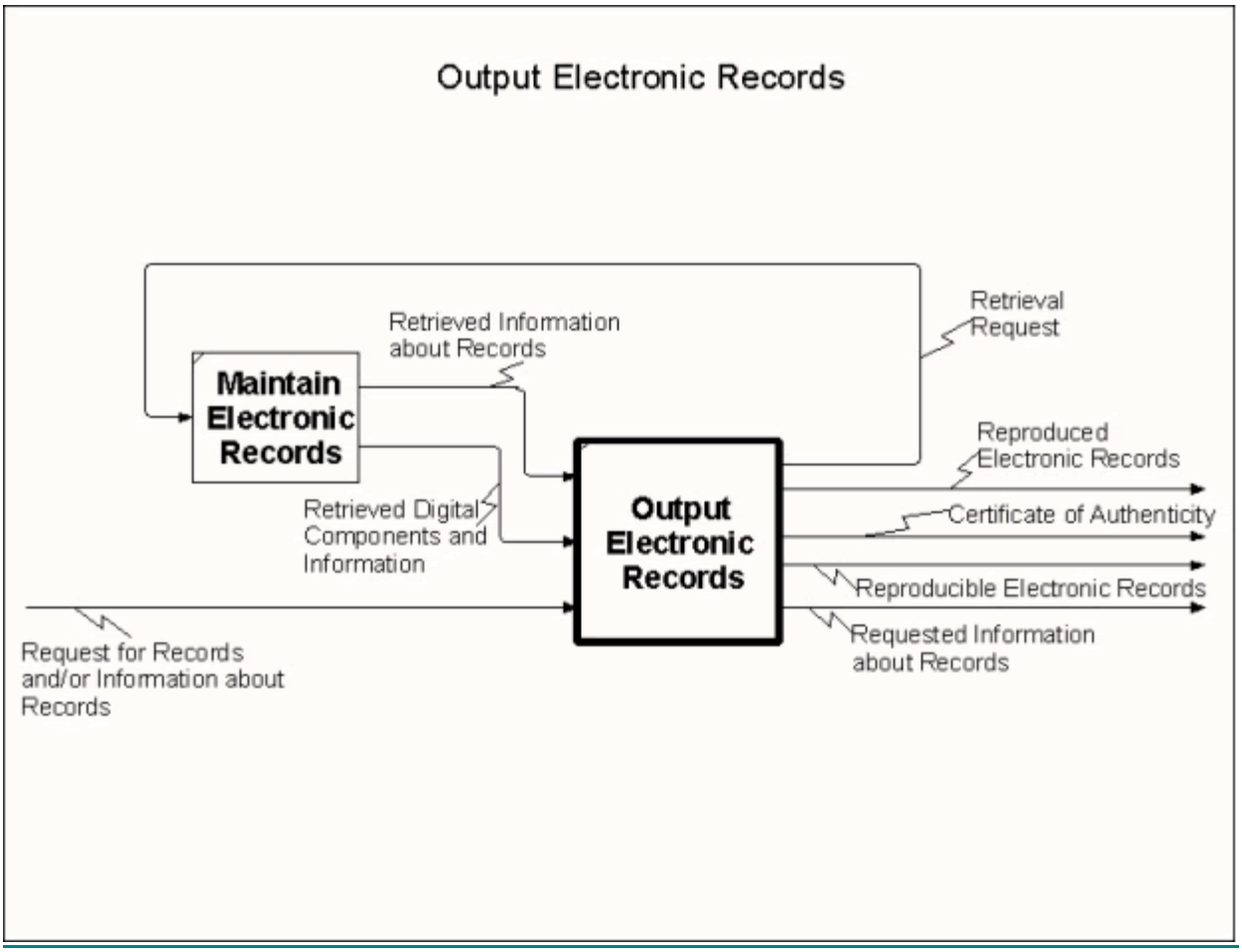


Figure 7. Reproducing Electronic Records

Figure 6 depicts the 'Output Electronic Records' process of the preservation model. The requests that are input to this process can be for records or for information about records, or both. While the reference function is responsible for helping requesters to identify the records of interest to them, descriptions and other finding aids will not contain all the detailed information about the records maintained by the preservation function. For example, a requester who needs a copy of the record that is certified as authentic may inquire whether such certification is possible before requesting the copy. Information about the records being preserved is maintained by the 'Maintain Electronic Records' process. When a request for information is received, the 'Output' process formulates the request in a way that the 'Maintain' process can respond to. For example, most of the information about the records is probably maintained in a database. The 'Output' process ensures that the request is in a format that can be executed as a query against the database. The 'Maintain' process retrieves the information and sends it to the 'Output' function. The 'Output' process determines if the retrieved information is complete and delivers it to the requester. The 'Output' process may also need to provide an explanation of the response.

> For example, a requester asks if it is possible to produce a copy of a record which can be certified as authentic. If the creator had kept the record in a system which had limited export capability, what was transferred to the archives may have been a plain text version of the record, losing the fonts, italics, and other aspects of presentation the record had in the creator's system. In such cases, the preserver could certify that a copy contained the complete and correct contents of the records, but that it could not certify that the appearance of the record was identical to the original.

When the request is for records, the same sequence of steps between the 'Output' and 'Maintain' processes is followed to return the components of the requested records to the 'Output' process. Then the records are reconstituted from their components. What happens after that point, however, depends on what the request specifies should be delivered. If the request is for an electronic copy of a record, in most cases that copy can only be produced on a system which is under the control of the preservation function. The final step in reproducing an electronic record is to render it with the appropriate presentation features. If the system on which the copy is presented is not under preservation control, there is no guarantee that it is properly presented.

In many cases, requesters will want to access electronic records on their own systems, either over the Internet or on digital media which can be read by their systems. In such cases, the 'Output' process will most often not be able to produce copies of the requested records. Rather it will deliver the digital components of the records along with instructions on who to reproduce the records from these components. The character of these instructions will vary depending on the delivery specifications in the request. For example, for records that will be rendered in a web browser, the instructions for reconstituting and rendering the records will be packaged together with the components and executed automatically by the requester's system. Even with such automatic reproduction, the preserver cannot guarantee that a copy on the requester's system is authentic. For example, the requester's system may not have all of the software needed to render the records properly. In cases where the requester want to bring electronic records into its own application, there may be a need to invoke middleware that mediates between the preservation system and the target application. The preserver should choose software mediators which protect end-to-end the integrity of the digital components that are transmitted between the two systems, but even here the preserver cannot certify the authenticity of copies produced in requesters' systems. In some cases, instructions for reproducing records from digital components will have to be in human-readable form.

These variant scenarios for the reproduction of copies of electronic records explain why the 'Output' process has the two distinct outputs: "Reproduced Electronic Records" and "Reproducible Electronic Records." The first of these outputs is produced when the records are presented on a system under preservation control. The second, consisting of digital components and instructions for reproducing the records, is produced for delivery to a system outside of preservation control.

The ‘Preserve Electronic Records’ process is designed to enable the production of authentic copies of electronic records. In some cases, requesters may want the preserver to attest to the authenticity of the copies. If so, the ‘Output’ process checks the preservation history kept by the ‘Maintain’ function to ensure that the chain of preservation for the requested records is intact; that is, that there was an adequate basis for presuming the records were authentic when transferred to the preserver’s custody and that the records have been properly preserved since that time, up to the point of reproduction. If these conditions have been satisfied, the ‘Output’ process issues a Certification of Authenticity. In cases where digital components are delivered to external systems, the instructions for reproducing the records should aim at the production of authentic copies. Accordingly, the ‘Output’ process should also provide the requesters with criteria they can use to determine if the copies are authentic.

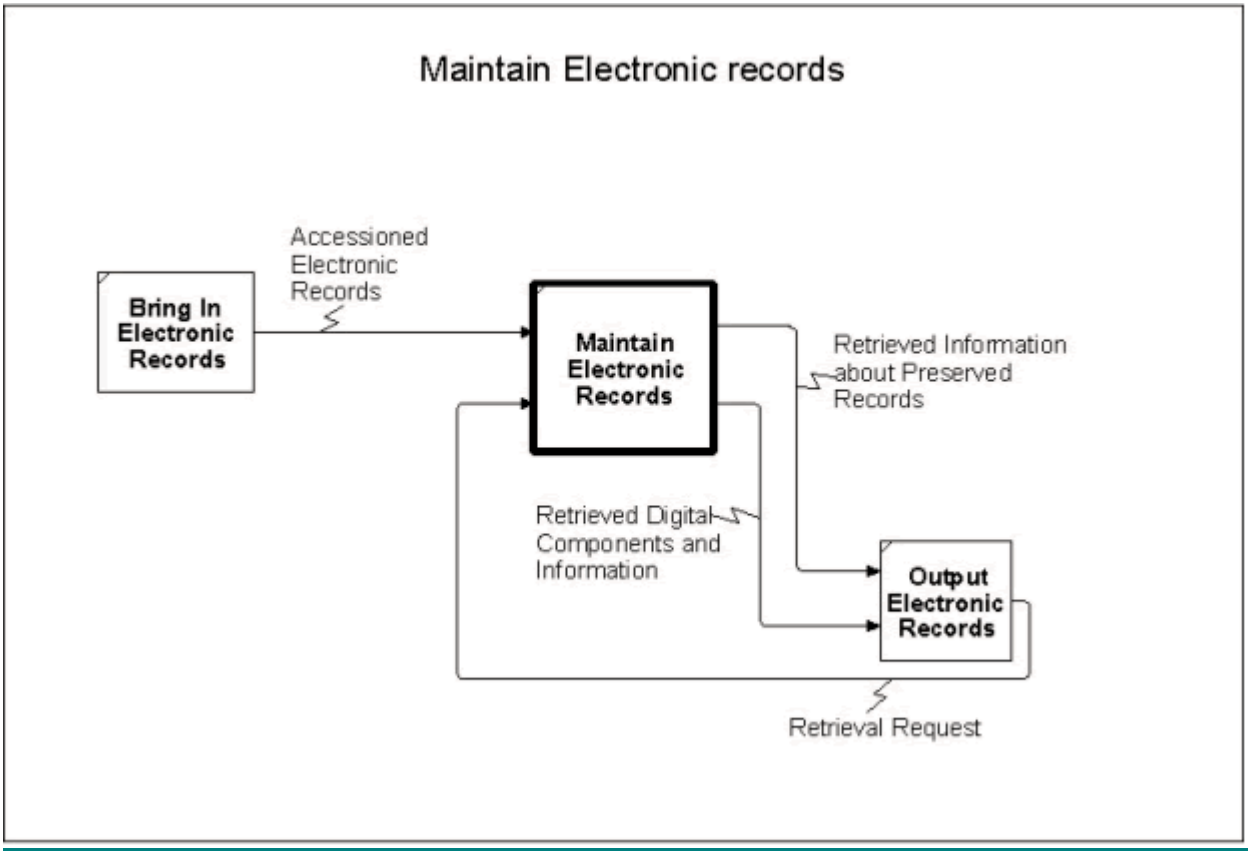


Figure 8. Maintaining Electronic Records

Maintaining Electronic Records Over Time: One of the four basic processes remains to be described, ‘Maintain Electronic Records.’ This is an internal process: none of the inputs that come into the preservation process from the outside go directly to this subprocess, and the outputs of the maintenance process go to other subprocesses within the function. Nonetheless, the ‘Maintain’ process is the core preservation process with respect to transmitting electronic records over time. It is depicted in figure 7. The ‘Maintain’ process is connected to both the ‘Bring In’ and ‘Output’ processes. The digital components of the records, along with information needed to reproduce

them, re-establish their archival bonds, and certify their authenticity are received from the 'Bring In' process. Requests for records and information are received from the 'Output' process and the results of executing such requests are returned to 'Output.'

The diagrams of the four preservation processes (figures 4 to 7) depict all of the entities that are in the comprehensive diagram of the four processes (figure 3), except for two outputs of the 'Manage' process: 'Information about Preservation' and 'Report on Authenticity of Records.' The first of these, 'Information about Preservation,' reflects the fact that the preserver will undoubtedly be responsible and accountable to others. In the case of an institutional archives, such as that of a corporation or university, the preserver will be responsible to the institution itself. In the case of government archives, not only will the preserver be responsible to higher levels of government, but it will also be accountable to the people. In any case, the preserver will need, and will want, to communicate information about its activities. The preserver will need to be able to produce a ''Report on Authenticity of Records' to justify its methods and procedures.

To this point we have examined the four preservation processes depicted in figure 3 as they relate to one another. In the following sections, we will consider the activities that take place within each of the four preservation processes.

3.3 Preservation Subprocesses

3.3.1 Manage the Preservation Function

Managing the Preservation Function determines how all other preservation processes will be carried out, and with what results, selects and acquires the necessary technology, and evaluates the execution of the function.

The management process consists of four sub-processes: Determine Preservation Requirements, Select Preservation Methods, Specify Outputs and Outcomes, and Evaluate Execution of Preservation.

Preservation requirements are determined by generic archival principles (such as provenance and authenticity of records) and specific institutional requirements (such as accessioning and dissemination policies and information technology standards), as well as information about the records to be preserved and knowledge derived from evaluation of the preservation of records already transferred. This process integrates and synthesizes external controls to determine the specific archival requirements for preserving and reproducing the records. This requires the specification of the attributes and characteristics of the records that must be preserved; a knowledge of how the records are composed from their digital components; an understanding of how records are to be grouped in the proper order; and a determination of how the authenticity of different classes of records will be certified.

Given these archival requirements for preservation, and working within the limits imposed by the state of the art of computer science and information technology, specific technological methods are identified, evaluated, and selected to be used in all aspects of preserving the different classes of records and archival aggregates selected for preservation: bringing them into the preservation system, maintaining them over time, and reproducing them in authentic form. The selection of these methods will be influenced by the institution's IT architecture, data standards and related procedures, security requirements, and access restrictions. Selecting preservation methods entails the identification and evaluation of available preservation methods; the selection of the method(s) which meets the archival requirements for preservation of each class of records; and the acquisition and adaptation, configuration or enhancement of the technology and other resources necessary to apply the selected method. Each preservation method will have a specified domain of application. Some methods, such as physical media and storage systems may be used for all records and digital components. Others will be specific to given classes of records or types of digital components.

A preservation method and its scope of application form the basis for a preservation strategy. The strategy is completed by specifying what outcomes will be produced by its application to the relevant domain. "Specifying Outcomes" sets the objectives and performance targets for the operation of the method and the means for identifying, measuring and reporting on the achievement of the objectives and targets. The outcomes should also encompass the results of risk assessments and indicate how to handle exceptional and problem cases.

In order to ensure that the objectives of the preservation function are being realized, and to maximize its performance, it is necessary to evaluate its execution. Evaluating Execution of the Preservation Function uses feedback from the processes of bringing in records, maintaining them, and reproducing them. Each of these processes is required to output "Management Information about Preservation." The evaluation sub-process measures performance against objectives, and identifies areas for improvement, at the micro level of objectives and targets and/or at the macro level of strategies and methods. Moreover, its output "Information about Preservation" acts as an input to the Appraisal function, transmitting information that will influence the " Determine the Feasibility of Preservation" sub-process of that function.

Taken as a whole, "Manage" governs the preservation function as a dynamic process in which the changing nature and volume of records to be preserved, shifts in institutional requirements, the evolving capacity of IT, and feedback from the function itself combine to produce preservation strategies, methods and objectives that evolve over time.

3.3.2 Bringing Electronic Records under Preservation Control

In accordance with the preservation strategies established in the preservation management process, electronic records are brought under preservation control. Bringing electronic records under preservation control includes four activities: registering the record transfer, verifying the authority for transfer, examining the records, and accessioning the records. The process starts with the transfer of records selected for preservation from the submitter. The person responsible for preservation will first determine that there is no evidence of problems occurring in the process of transfer. (If problems are identified, the submitter should be asked to resend the materials.) Registering the transfer captures information about the transfer such as submitter's name, record creator's name and current date which is contain in documentation accompanying transfer. Registration establishes basic control over the materials transferred by assigning a unique identifier to them. The next step involves verifying the authority for transfer, also based on information about the transfer. A transfer is determined as authorized if and only if it comprises records that have been selected for preservation and those records have been submitted either by the records creator or an agent for the creator.

Examining the electronic records that have been transferred is the principle means of bringing the records under preservation control. It serves four purposes: to determine whether the archives will preserve the records, to identify preservation strategies to be used, to determine when preservation interventions should occur, and to identify, produce and capture information necessary to assert the authenticity of the records. A transfer includes digital files and information identifying what records and digital components are contained in those files. The first step in examination is to determine where the records and components are located in the transferred files and how they are identified. This is achieved by reading the digital files to verify the information accompanying the records.

Once the records and components have been identified in the transferred material, they are examined to determine if the terms and conditions of transfer are satisfied. This includes determining that all components necessary to compose all the records that should be included in the transfer are present and intact; that the formats of the components and the methods necessary to reconstruct and render the records are known; and that there is sufficient data to reconstruct the files, series, and other archival aggregates in which the records are organized. The examination should also determine if any preservation intervention, such as migration or transformation to persistent form, is necessary to enable the records to be preserved using applicable preservation strategies. If so, the necessary interventions should be identified. They will be carried out in the process of maintaining the records.

The examination should also include a final review[12] of information that provides a basis for presuming that the records were maintained authentic. That information then becomes part of the preservation history of the records. It will be retained, and augmented in the course of maintaining the records over time in order to document the 'chain of preservation' and enable copies of records to be certified as authentic.

These activities enable the archives to accession the records or reject the transfer.

3.3.3 Maintaining Electronic Records Over Time

Electronic records are stored as *digital components*, which may be separate *digital files,* or contained in a single digital file. The preservation function aims to deliver records which convey the intellectual content and intent of the creator. To do this, it is necessary to reassemble the components to reconstitute the records and present them in their original documentary form.. This has to be done every time the record is accessed simply because an electronic record cannot be stored in the same form in which it is presented to humans. Strictly speaking, maintaining electronic records over time means maintaining the ability to output them. Maintaining electronic records, thus, requires storing their digital components and maintaining information about the records, such as what digital components they contain, how those components are related to each other and the records, and how the records should be presented. In order to be able to output authentic electronic records, it is also necessary to maintain information which justifies an assertion of authenticity.

The activities needed for maintenance of electronic records over time includes putting record components into storage, managing the storage of the record components and information about records, and maintaining the ability to retrieve components and reproduce the records. These functions and operations are governed by preservation strategies.

Maintaining Electronic Records requires managing information about the records and record components which, in the first instance, can be viewed as belonging to one of three classes: intellectual, technical, and administrative. The intellectual information includes: provenance, documentary context, and description of contents, structure and form. The technical information includes metadata about the components, their relation to the records, and the methods required to reconstitute the records and render them. on output media, which may be digital, such as a computer screen, or analog, such as paper or microform. Maintainability also requires administrative information such storage information (identification and location storage media, digital files and digital components; type of media, health of the media), and the history of actions taken to maintain the records and to prevent inappropriate alteration. The three types of

[12]This information is reviewed previously as part of the appraisal process.

information are used both in other processes related to maintaining electronic records and in responding to requests for records or information about records.

The second process necessary to maintain electronic records is to 'Manage Storage,' which involves putting the digital files containing the components into storage; monitoring storage both to identify any damage or deterioration that may occur and to determine when it is necessary to refresh or migrate storage media or storage systems; and correcting any storage problems occurring, including disaster recovery, retrieving components in need of maintenance or for reproduction. The basic objective in all processes involved in managing storage is to keep the bit streams which comprise the digital files and digital components intact and retrievable. Managing storage does not involve any preservation intervention which would change the digital components, such as by migrating them to new data formats. When corrective actions are taken to address problems which could alter the contents of one or more digital components, the affected records should be reproduced in order to prove that the corrective action was successful or to document the impact on the records being preserved. Managing storage entails reporting information about changes in the media or the location of the files and components, and about any actions taken to prevent or recover from storage problems. As appropriate, such information is used in managing information about the records or in managing the preservation function.

The third type of process covers actions taken is to maintain the ability to output electronic records. Such actions will be taken either when the 'Manage Preservation' process determines to change a preservation strategy affecting the reproduction of any records being preserved or when records are examined during the process of bringing them into the preservation system and it is determined that some technical intervention is needed to enable the records to be reproduced under applicable preservation strategies. The types of actions will depend on the methods selected in applicable preservation strategies. Under some strategies, such as migration and persistent object strategies, the actions may change the digital components themselves. Other strategies, such as maintaining the original computer systems and emulation will leave the digital components unaltered, but change the hardware or software used to reconstitute and render records. In all cases, however, preservation strategies determine both the formats in which digital components are stored and the methods applied to them to reproduce records, files, series, etc.

Thus, whenever a preservation strategy that affects the reproduction of electronic records is changed, the new strategy should be evaluated to ensure that the records can still be reproduced and to document any impact of the change in strategy. Similarly, whenever digital components are altered to conform to preservation strategy, the result should be evaluated. The evaluations should be performed at the level of individual records and, when sets of records, such as databases or case files, are impacted, at the aggregate levels as well.

Evaluating the success of such technical interventions requires reconstituting and rendering the record(s) and/or archival set(s) based on the reproduction strategies and

authenticity requirements. In both cases, any impact on the records should be documented. The results are needed both for managing information about the records and for managing preservation overall.

3.3.4 Reproducing Electronic Records

The ultimate objective of preserving electronic records is to transmit them over time to users who have a need for, or an interest in, them. The process of preserving an electronic record is 'complete' only at the point where the record has been reproduced in authentic form. Providing an electronic record entails a process of reproducing or reconstituting it from its digital component(s). The person responsible for preserving them needs to ensure that this process can be executed to output authentic records.

The process starts with the receipt of a request for a record, or records. It is presumed that, prior to a request coming into the preservation process, the access or reference function has determined that the requested records are available and that the requester has a right to them.

Reproducing an authentic electronic record requires retrieving the digital components of the record and executing the methods required to reconstitute the structure and content of the record from its digital components and presenting it in appropriate form. Certifying the authenticity of the reproduced record requires information supporting the presumption that the creator preserved the records authentically up to the time the records were transferred to the archives and the audit trail of their preservation since the time of transfer up to and including the process of reproduction.

The first step is to retrieve the digital components and related information. Then, the methods dictated by the applicable reproduction strategy can be applied to the components to reproduce the record. Depending on the request, this can be done by the person responsible for preservation, by a person responsible for access or dissemination, or by the requester.

When someone else will carry out the reproduction of records, the person responsible for preservation needs to provide the components and the information needed to reproduce the records and to support an assertion that the output record is authentic. In any case, the preserver should conform to the baseline requirements for the production of authentic copies of electronic records set out in the InterPARES Authenticity Task Force report, Requirements for Assessing the Authenticity of Electronic Records. If another person carries out the reproduction process, the preserver should provide that person either assurance or evidence that all requirements up to the reproduction process have been satisfied, along with specifications that apply to the reproduction itself. These specification will articulate the baseline requirements as they apply specifically to making authentic copies of the records in question.

4 Conclusion and Recommendations

Much attention on 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 a 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."[13] 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 which define the problem. It must be archival and records management criteria which 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." 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 which 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 both the mechanisms needed to carry out these processes and 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.

[13] http://www.interpares.org/researchplan.htm

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 which 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:

- Digital Components of Electronic Records,
- Preservation Control,
- Archival Requirements for Preservation,
- 'Original' Electronic Records,
- The Need to Reproduce Electronic Records, and
- The Chain of Preservation,

as set out in chapter 2 of this report. 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 brought into the archives will change. The preservation system must be capable of being expanded, adapted, and 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 will want to use the best available technology for access to the records. Preservation solutions should be designed 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 not only provided valuable products in response to the research questions that it originally posed, but 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. 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.

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Preservation Bibliography

Draft Appendix

October 12, 2001

Task Force Members:

Kenneth Thibodeau, U.S. National Archives and Records Administration, chair,
Richard Blake, Public Records Office, UK,
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

WORKING BIBLIOGRAPHY

The Preservation of Digital Records

The publications on this topic are proliferating rapidly in both electronic and paper versions. Unfortunately, the quality of the work is uneven, and there is much repetition of information. This bibliography will serve as a guidepost. In addition to up-to-date pieces that focus on digital issues, we have included several publications from the 1980s that deal with such core and enduring issues as selection for preservation, intellectual versus physical preservation, cost modeling, and relevant managerial functions. At the end of this bibliography is a list of other recent bibliographies.

ENGLISH

Articles

American Library Association. "Preservation Policy," in *ALA Policy Manual, 1992-93,* 52.2, p. 145.

This policy, written in 1991, is reprinted annually in the *ALA Policy Manual.* It also appears on the ALA web site, http://www.ala.org. In June 1993, the *Report on Implementation of the ALA Preservation Policy* was published; it also resides on the web. The 1993 report deals more particularly with electronic data issues.

Atkinson, Ross W. "Selection for Preservation: A Materialistic Approach." *Library Resources & Technical Services* 30.4 (October/December 1986): 341-353.

Atkinson establishes a typology for preservation decision-making that is organized around three classes: materials that must be preserved to protect their capital value; high-use materials; and lower-use research materials. He uses the Research Libraries Group (RLG) Conspectus as a model for distributing the responsibility of class 3 preservation around institutional subject strengths. (Contrast the Atkinson library-conspectus approach to the Coxes' archives-information appraisal approach.)

Bantin, Philip C. "Developing a Strategy for Managing Electronic Records – The Findings of the Indiana University Electronic Records Project." *The American Archivist* 61 (Fall 1998): 328-364.

Bantin, Philip C. “The Indiana University Electronic Records Project Revisited.” *The American Archivist* 62 (Spring 1999): 153-163.

Battin, Patricia. "From Preservation to Access: Paradigm for the Nineties." *IFLA Journal* 19.3 (1993): 367-73.

Battin posits that preservation is not the central function of librarianship because “[d]igital information implies continuous change accompanied by ever diminishing life cycles of the medium and the systems of access to the point where the preservation function must be considered at the point of creation of the knowledge.” Battin was early to recognize the key issues surrounding the preservation of digital media.

Bouché, Nicole. "Digitization for Scholarly Use: The Boswell Papers Project at the Beinecke Rare Book and Manuscript Library". *Council on Library and Information Resources* (March 1999): 1-15.

Bouché reports on the digitization of the personal papers of James Boswell, from forming a strategy to the scanning process. She evaluates the positive and negative aspects of such projects for scholar-editors.

Child, Margaret S. "Further Thoughts on 'Selection for Preservation: A Materialistic Approach." *Library Resources & Technical Services* 30. 4 (October / December 1986): 354-362.

Child agrees in principle with Atkinson’s views (see above) but suggests ways in which he could refine his classes. For example, Child prefers the archivists’ “intrinsic value” to Atkinson’s “capital value.” She fleshes out Atkinson’s ideas, suggesting practical means for their implementation. Finally, she argues for the importance of including documentary resources of all types -- print, manuscript, machine-readable files, etc. -- in a subject-based approach to preservation.

Cloonan, Michèle Valerie. “The Preservation of Knowledge.” *Library Trends* 41.4 (1993): 594-605.

Cloonan discusses the conceptual shift that must take place in the preservation field as library collections evolve from holding primarily paper-based materials to housing digital collections as well.

Cloonan, Michèle Valerie. “The W(h)ither Preservation?.” *Library Quarterly* 71.2 (April 2001): 231-242.

Conway, Paul. "Archival Preservation: Definitions for Improving Education and Training." *Restaurator* 10 (1989): 2, 47-60.

This article traces the emergence of what constitutes archival preservation and the implications of that definitiaon for education and training. He brings together the perspectives of the library and archival worlds.

Conway, Paul. *Preservation in a Digital World.* Washington, D.C.: The Commission on Preservation and Access, 1996.

An overview of issues and strategies particular to the preservation of digital records.

Cox, Richard J., and Cox, Lynn W. "Selecting Information of Enduring Value for Preservation: Contending with the Hydra-Headed Monster." *Rethinking the Library in the Information Age* 2 (April 1989): 115-130.

The authors characterize early preservation literature as advocating the preservation of all information. By the 1980s librarians acknowledged that not all information could be saved, and criteria for selection began to be determined through a collection management approach. Compares and contrasts selection and appraisal approaches, and promotes an "information value first" perspective where preservation follows information selection.

Cox, Richard J., "What is an Archival Record, and Why Should We Care?" *The American Archivist* 57 (Fall 1994): 592-594.

Crawford, Walt. "Bits is Bits: Pitfalls in Digital Reformatting." *American Libraries* 30.5 (May 1999): 47-49.

Intended for a general librarian readership, this article succinctly spells out the attendant issues surrounding the preservation of digital media: fragility of new media, obsolescence of technology, and so on.

Gilliland-Swetland, Anne J. "Archival Research: A 'New' Issue for Graduate Education." As presented at the Archival Educators' Working Meeting, University of Pittsburgh, August, 1999.

Graham, Peter S. *Intellectual Preservation: Electronic Preservation of theThird Kind.* Washington, D.C.: Commission on Preservation and Access, March 1994.

Graham addresses the point that with print materials preserving the artifact has meant preserving the information contained in it, while with digital materials preservation may merely signify copying. In the process of copying or migrating

documents, much information may be lost or altered. He discusses the issues of integrity and authenticity of information which, though central concepts to archivists, are not necessarily obvious to librarians. Finally, Graham considers technologies such as hashing and digital time-stamping as aids in authentication of records.

Green, Ann, JoAnn Dionne, and Martin Dennis. *Preserving the Whole: A Two- Track Approach to Rescuing Social Science Data and Metadata* Washington D.C.: The Digital Library Federation, June 1999.

An exploration of migration as a preservation strategy for social science data. It explores the options available for migrating both data stored in a technically obsolete format and their associated documentation stored on paper, which may itself be rapidly deteriorating. A very helpful glossary is included.

Hurley, Bernard J., John Price-Wilkin, Merrilee Profitt and Howard Besser. *The Making of America II Testbed Project: A Digital Library Service Model* Washington D.C.: The Digital Library Federation, December 1999.

The study defines a digital library service model that encapsulates the interaction of digital objects (including their metadata), tools, and services based on the principles of object-oriented design. The use of XML to encode objects and as transfer syntax is discussed. Also includes an extensive table of structural and administrative metadata elements.

International Federation of Library Associations. "IFLA Principles for the Care and Handling of Library Materials," rev. ed. The Hague: IFLA, 1999. Russian translation, Moscow, 1999 (ISBN 5-7380-0093-5).

As of August 1, 2001:
Available also in: Polish, Slovene.
Translations expected in Arabic, Bahasa* Malaysian, Croatian, French, Greek, Japanese, Portuguese, Romanian, Spanish and Turkish.

[*The website incorrectly spells it “Bhasa”. Bahasa means “language”.]

International Federation of Library Associations / UNESCO. “Survey on Digitisation and Preservation," compiled and edited by Sara Gould and Richard Ebdon under the direction of Marie-Thérèse Varlamott. London: IFLA, 1999.

Kenney, Anne. R. and Stephen Chapman. *Digital Imaging for Libraries and Archives.* Ithaca, NY: Cornell University Library, 1996.

Larsen, Poul Steen. “Books and Bytes: Preserving Documents for Posterity.” *JASIS* 50 (11): 1020-1027, 1999.

Lawrence, Gregory W., et al. “Risk Management of Digital Information: A File Format Investigation.” Washington, D.C.: Council on Library and Information Resources, June 2000.
Downloaded from: http://www.clir.org (October 2000)

Lesk, Michael. *Image Formats for Preservation and Access*. Washington, D.C.: The Commission on Preservation and Access, September 1992.

A primer on the logistics of electronic preservation. Lesk was one of the first writers to examine the technological and intellectual issues surrounding the reformatting of paper-based information to digital formats, and the migration of digital documents to newer digital formats. See also the chapter, "Collections and Preservation" in his *Practical Digital Libraries* (below).

O’Toole, James, M. “On the Idea of Uniqueness.” *The American Archivist* 57.4 (Fall 1994): 632-658.

Pederson, Ann. *Migrating Legacy Systems into the Electronic Environment*. Paper given at the *Electronic Records Seminar*. International Records Management Trust. June 1998.
Downloaded from: http://www.irmt.org/resources/ownlad2.html (November 9, 1999).

The Research Libraries Group, Inc. and The Commission on Preservation and Access. "Preserving Digital Information: Report of the Task Force on Archiving of Digital Information.” May 1, 1996
Downloaded from: http://www.rlg.org/ArchTF/tfadi.intro.htm

Rothenberg, Jeff. "Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation." *Council on Library and Information Resources* (January 1999): 1-35.

An extended discussion on computer emulation as a means for digital preservation. As Rothenberg is a strong proponent of emulation, this publication does not present a balanced view. His model is being used at The National Archives of the Netherlands and CURL Exemplars in Digital Archives Project (Cedars) 2, (Leeds, Oxford and Cambridge), University of Michigan sponsored by CURL (Consortium of University Research Libraries.

Rothenberg, Jeff. "Ensuring the Longevity of Digital Information". Originally in *Scientific American* 272 (January 1995) 1: 42-7; Revised February 1999.

This revision can be found at:
http://www.clir.org/programs/otheractiv/ensuring.pdf

Sanett, Shelby and Eun Park. "Authenticity as a Requirement of Preserving Digital Data and Records." *IASSIST Quarterly*, 24:1 (Spring 2000).

Smith, Abby. "Why Digitize?" *Council on Library and Information Resources* (February 1999): 1-13.

A straightforward overview of the issues relating to the digitization of collections.

Smith, Wendy. "The National Library of Australia's Pandora Project." *Libri* 47.3 (September 1997): 169-179.

A description of Pandora and other related Australian digital projects.

Trace, Ciaran and Shelby Sanett. "InterPARES: Securing the Future of Electronic Records." *Bulletin of the American Society for Information Science* 27:1 (October/November 2000): 24-26.

Monographs

Beagrie, N. and D. Greenstein. *A Strategic Policy Framework for Creating and Preserving Digital Collections* British Library Research and Innovation Report 107. London: The British Library, 1998.

Bennett, J.C. *A Framework of Data Types and Formats, and Issues Affecting the Long Term Preservation of Digital Material* British Library Research and Innovation Report 50. London: The British Library, 1997.

Bikson, T.K. and E.J. Frinking. *Preserving the Present: Toward Viable Electronic Records* The Hague: Sdu Publishers, 1993.

Consultative Committee for Space Data Systems. "Reference Model for an Open Archival Sytem (OAIS)". Washington, D.C.: National Aeronautics and Space Administration. 1999.

Cromwell-Kessler, Willy. "Crosswalks, Metadata Mapping, and Interoperability: What Does It All Mean?" In *Introduction to Metadata: Pathways to Digital Information*, edited by Murtha Baca. Los Angeles: Getty Information Institute, 1998.

Discusses the importance of creating maps of metadata systems and the "crosswalks" that result from analyzing and mapping those similarities of each metadata system which overlap. By use of maps and crosswalks, automated systems can interoperate in such a way that integrated access may be achieved. Also included is a very helpful chart, "A Crosswalk of Metadata Standards".

The Data Archive.*An Investigation into the Digital Preservation Needs of Universities and Research Funders: the Future of Unpublished Research Materials*, British Library Research and Innovation Report 109. The British Library, 1998.

Dollar, Charles M. *Authentic Electronic Records: Strategies for Long-Term Access.* Illinois: Cohasset Associates, Inc., 1999.

Duranti, Luciana. *Diplomatics: New Uses for an Old Science*. Maryland: Scarecrow Press, Inc., 1998.

Floyd, Barbara. An Archivist looks at Preservation In *Managing Preservation: A Guidebook.* Ohio: State Library of Ohio and the Ohio Preservation Council, 1992.

An overview of the common preservation problems faced by archivists while processing items at the collection level. Some possible solutions are provided. Also included is a brief discussion of the differences in span of responsibilty for preservation between the collection level and the repository level of archives.

Avrami, Erica, et al. *Values and Heritage Conservation*: *Research Report*. Los Angeles: The Getty Conservation Institute, 2000.

Gilliland-Swetland, Anne J. "Defining Metadata." In *Introduction to Metadata: Pathways to Digital Information*, edited by Murtha Baca. Los Angeles: Getty Information Institute, 1998.

Enlarges the concept of metadata beyond the descriptive function to include the range of activities which information professionals are involved in where metadata may be incorporated into digital information systems. The concept is broken down into several categories that reflect key elements of metadata functionality: administrative, descriptive, preservation, use, and technical. A very helpful glossary is included.

Green, A.; J. Dionne, and M. Dennis. *Preserving the Whole: A Two-Track Approach to Rescuing Social Science Data and Metadata.* Washington, D.C.:The Digital Library Federation, 1999.

Haynes, D., D. Streatfield, D. Jowett and M. Blake. *Responsibility for Digital Archiving and Long Term Access to Digital Data.* British Library Research and Innovation Report 67. London: The British Library, 1997.

Hendley, T. *Comparison of Methods and Costs of Digital Preservation.* British Library Research and Innovation Report 106. London: The British Library, 1998.

Historical and Philosophical Issues in the Conservation of Cultural Heritage. Edited by Nicholas Stanley Price et al. Los Angeles: The Getty Conservation Institute, 1996.

Kenney, Anne R. and Oya Y. Rieger. *Moving Theory into Practice: Digital Imaging for Libraries and Archives.* California: Research Libraries Group, 2000.

This timely monograph explores a number of topics, some of which include: a discussion of significant issues and projects regarding digitization; criteria for selecting material for digital conversion; digital image creation; digital benchmarking for decision-making; implementing quality control programs; and the practical aspects of image management and use.

Lesk, Michael. *Practical Digital Libraries: Books, Bytes, and Bucks.* San Francisco: Morgan Kaufmann, 1997.

In his chapter on "Collections and Preservation", Lesk explores the evolving challenges to preservation efforts from traditional preservation problems to issues and problems posed by electronic media. While there is some discussion about possible solutions, Lesk raises excellent points regarding practical and intellectual preservation issues and cost-effectiveness as a factor of decision-making. Also recommended is the chapter on "Economics" which discusses the need for a new model of cost modeling for digital libraries and explores economic ramifications of preserving electronic information.

Lyman, Peter and Howard Besser. *Time & Bits: Managing Digital Continuity: Time & Bits Conference* " Background Paper". February 1998.
Downloaded from: http://www.gii.getty.edu/timeandbits/tbpaper.html (August 1999).

Matthews, G., A. Poulter and E. Blagg. *Preservation of Digital Materials Policy and Strategy Issues for the UK: Report of a Meeting on the CPA/RLG Report, December 1996* Library Research and Innovation Report 41. London: The British Library, 1997.

National Institute for the Conservation of Cultural Property: Awareness Study. New Jersey: The Gallup Organization, 1996.

National Preservation Office. *Digital Culture: Maximising the Nation's Investment: A Synthesis of JISC/NPO Studies on the Preservation of Electronic Materials*, edited by Mary Feeney. London: The National Preservation Office, The British Library, 1999.

Includes a chapter (5) on estimating the costs of digital preservation.

Neavill, Gordon B. "Preservation of Computer-Based and Computer-Generated Records," in *Conserving and Preserving Materials in Nonbook Formats* edited by Kathryn Luther Henderson and William T. Henderson. 30th Allerton Institute. Urbana-Champaign, IL: University of Illinois, November 1998, pp. 45-60.

The conference actually took place November 6-9, 1988, in Champaign, Illinois.

Playback: A Preservation Primer for Video. Edited Sally Jo Fifer et al. San Francisco: Bay Area Video Coalition, 1998.

Includes chapters on aspects of preserving video art. Of particular interest is the chapter on "Establishing Priorities for Preservation" by Debbie Hess Norris. A glossary of video preservation terms is included.

Ross, Seamus and Gow, Ann. "Digital archaeology? Rescuing Neglected or Damaged Data Resources". London & Bristol: British Library and Joint Information Systems Committee, 1999.

Ross, Seamus and Higgs, E. (eds.) *Electronic Information Resources and Historians: European Perspectives*. St.Katharinen: Halbgraue Reihe zur historischen Fachinformatik, A20 and as British Library Report 6122, 1993.

Rothenberg, Jeff. *Using Emulation to Preserve Digital Documents*. Koninkliijke Bibliotheek: The Hague, July 2000.

Time & Bits: Managing Digital Continuity. Edited by Margaret MacLean and Ben H. Davis. Los Angeles: Getty Conservation Institute, 1998.

Westat. *Report on the Assessment of Electronic Government Information Products*. Prepared under contract for the United States National Commission on Libraries and Information Science. [Washington, D.C.]: United States Government Printing Office Superintendent of Documents, March 30, 1999.

Contains a short but useful section on preservation (pp. 59-61).

Conferences

Andersson, Ulf. "Short Version of the Sesam Report: Philosophy and Rules Concerning Electronic Archives and Authenticity." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Battin, Patricia. "Image Standards and Implications for Preservation." Talk presented at the *Workshop on Electronic Texts*, sponsored by the Library of Congress, 9-10 June 1992.
Downloaded from:
http://palimpsest.stanford.edu/byauth/battin/imagestd.htm

Besser, Howard. "Metadata, Images, & Digital Libraries." Paper presented at the *School for Scanning* . Chicago, 3 June 1999.

Busch, Joseph A. "The Value of Context for Data and Information." Position Paper presented at the *NSF Workshop on Data Archiving & Information Preservation*. Washington, D.C., 5 March 1999.
Downloaded from:
http://www.cecs.missouri.edu/NSFWorkshop/buschpp.html (July 1999).

Chapman, Steve. "Textual Scanning: Control and Costs." Paper presented at the *School for Scanning* . Chicago, June 1999.

Cloonan, Michèle Valerie and Shelby Sanett. "Comparing Preservation Strategies and Practices for Electronic Records." Paper presented at *Preservation 2000: An International Conference on the Preservation and Log Term Accessibility of Digital Materials*. York, England, December 2000.
Downloaded from: http://www.rlg.org/events/pres-2000/ (March 2001).

Conchon, Michèle. "Processing and Conservation of Computer Archives in France: the Constance Method." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Conway, Paul. "Digital Preservation: Theory and Reality." Paper presented at the *School for Scanning*. Chicago, 3 June 1999.

Conway, Paul. "Digital Technology: Products and Process." Paper presented at the *School for Scanning*. Chicago, 3 June 1999.

Doggen, Jacobus. "The Electronic Archives of the Publications Office." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Doorn, P.K. and H.D. Tjalsma. "Historical Data Archives: Preserving and Documenting Historical Data." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Duranti, Luciana. "The Future of Archival Scholarship." Unpublished paper presented at University College. Dublin, 2 October 1998.

Duranti, Luciana. "The Preservation of the Integrity of Electronic Records." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Exon, Maggie. "Long-Term Management Issues in the Preservation of Electronic Information." Paper presented at the*2nd National Preservation Office Conference: Multimedia Preservation - Capturing the Rainbow*. Brisbane, 28-30 November 1995. Downloaded from: http://www.nla.gov.au/niac/meetings/npo95me.html (August 1999).

Garcia-Molina, H. and Arturo Crespo, "Digital Library Repositories." Paper presented at the *NSF Workshop on Data Archiving & Information Preservation*. Washington, D.C., 5 March 1999.
Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

http://www-db.stanford.edu/archivalrep

Halem, M. et al. "Technology Assessment of High Capacity Data storage Systems: Can We Avoid A Data Survivability Crisis?" Paper presented at the *NSF Workshop on Data Archiving & Information Preservation*. Washington, D.C., 5 March 1999. Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

Hannigan, Ken. "A Summary on Electronic Records Management in the EU Member States: Relations Between Public Administrations and Archives Services." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Hedstrom, Margaret. "Digital Preservation: A Time Bomb for Digital Libraries." Paper presented at the*NSF Workshop on Data Archiving & Information Preservation*. Washington, D.C., 5 March 1999.
Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

Hedstrom, Margaret. "Digital Preservation: Matching Problems, Requirements and Solutions." Presented at the *NSF Workshop on Data Archiving & Information*

Preservation. Washington, D.C., 5 March 1999.
Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

Jones, Richard. "Painting the Rainbow in Australia: The Way Ahead." Paper presented at the *2nd National Preservation Office Conference: Multimedia Preservation - Capturing the Rainbow*. Brisbane, 28-30 November 1995.
Downloaded from: http://www.nla.gov.au/niac/meetings/npo95me.html (August 1999).

Kiuru, Jarmo, et al. "Digital Archiving of Legislative Documents: A Case Study of Finland." Presented at the *NSF Workshop on Data Archiving & Information Preservation*. Washington, D.C., 5 March 1999.
Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

Le Cerf, Phillippe., et al. "Standards for Electronic Document Management." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Lord, Philip, et al. "Strategies and Tactics for Managing Electronic Data Records: A View from the Pharmaceutical Industry." *Proceedings of the DLM-Forum on Electronic Records*. Brussels, 18-20 December 1996.

Mintzer, Fred. "Long Term Storage of Digital Media Objects for Digital Libraries." Presented at the *NSF Workshop on Data Archiving & Information Preservation*. Washington, D.C., 5 March 1999.
Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

Price-Wilkin, John. "Case Study: University of Michigan." Paper presented at the *School for Scanning*. Chicago, 3 June 1999.

Reilly, Bernard and Matthew Cook. "Case Study: CHS Works in Progress - Chicago Daily News and the Haymarket Project." Paper presented at the *School for Scanning*. Chicago, 3 June 1999.

Rothenberg, Jeff. "Using Emulation to Preserve Digital Information." Presented at the NSF Workshop on Data Archiving & Information Preservation. Washington, D.C., 5 March 1999.
Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

Smith, Abby. "Next Steps: Bringing It All Back Home." Paper presented at the School for Scanning. Chicago, 3 June 1999.

Tennant, Roy. "Hey, What About Access?: A Guide to Practical Decision-Making." Paper presented at the School for Scanning . Chicago, 3 June 1999.

Thibodeau, Ken. "Boundaries and Transformations: An Object-Oriented Strategy for the Preservation of Electronic Records. "Proceedings of the DLM-Forum on Electronic Records. Brussels, 18-20 December 1996.

Thibodeau, Ken. "Resolving the Inherent Tensions in Digital Preservation." Presented at the NSF Workshop on Data Archiving & Information Preservation. Washington, D.C., 5 March 1999.
Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop/thibpp.html (July 1999).

Tombs, K. "Governmental, Industry and User Perspectives of Achieving Standard Storage Mechanisms for Long-Term Archival Activities." Proceedings of the DLM-Forum on Electronic Records. Brussels, 18-20 December 1996.

Vissers, H. "The Dutch Touch: Overview of the Digital Longevity Programme in the Netherlands." Proceedings of the DLM-Forum on Electronic Records. Brussels, 18-20 December 1996.

Vogt-O'Connor, Diane. "Case Study: Selection in the National Park Service." Paper presented at the School for Scanning. Chicago, 3 June 1999.

Waters, D.J. "Toward the Long Term Maintenance of Digital Publications: The Development of Organizational Models". Presented at the NSF Workshop on Data Archiving & Information Preservation. Washington, D.C., 5 March 1999. Downloaded from: http://www.cecs.missouri.edu/NSFWorkshop (July 1999).

Wettengel, M. "Technical Infrastructure for the Archiving of Digital Databases: Requirements and Procedures." Proceedings of the DLM-Forum on Electronic Records. Brussels, 18-20 December 1996.

Glossaries

AIC Definitions of Conservation Terminology. *AIC News*. 21.5 (May 1996): 15.

Bellardo, Lewis J. and Lynn Lady Bellardo.*A Glossary for Archivists, Manuscript Curators, and Records Managers*. Chicago: The Society of American Archivists, 1992.

Electronic Imaging Glossary. Compiled by Mimi King. Chicago: American Library Association. June 1994.

Fox, Michael J. and Peter L. Witherspoon, ed. Susanne P. Warren, "Introduction to Archival Organization and Description." 1998. Los Angeles. *Getty Information Institute*. See glossary, pp. 49-51.

Genesis and Preservation of an Agency's Archival Fonds Glossary. University of British Columbia.
Downloaded from: http://www.slais.ubc.ca/users/duranti/gloss.htm

Glossary of Selected Preservation Terms. ALCTS Newsletter< 1(1990): 2, 14-15.

International Council of Archivists. *Glossary* (1999). Multilingual: English, Spanish, French and German. In press.

InterPARES Glossary. Downloaded from: http://www.interpares.org/reports.htm (June 2001).

Lanzi, Elisa. Vocabularies: Enhancing Access to Cultural Heritage Information. Los Angeles. *Getty Information Institute* 1998.

This work focuses on controlled vocabularies for cultural heritage information; it lists several multilingual glossaries.

Universal Preservation Format (UPF) Glossary.
Downloaded from: http://info.wgbh.org/upf/glossary.html

Palacios, Antonio Lozano. "Vocabulario Ingles-Espanol Para Los Estudios de Biblio-Documentacion, 3a edicion revisada y aumentada. *Universidad de Granada*: Granada, 1996.

Mendez, Eva M. "Algunas Precisiones Terminologicas en Torno a la Materia de Metadatos y Recuperacion de Informacion. *Dpto. Biblioteconomia y Documentacion*. Universidad Carlos III de Madrid, 1999.

Webography

Bearman, David. "Reality and Chimeras in the Preservation of Electronic Records." Published in *D-Lib Magazine*, April 1999, Volume 5, Number 4.
Downloaded from: http://www.dlib.org/dlib/april99/bearman/bearman-notes.html (September 1999).

"The Cedars Project: Working Definitions of Commonly Used Terms."
Downloaded from: http://www.leeds.ac.uk/cedars/documents/PSW01.htm (July 1999).

"Cedars 2: Emulation for Preservation."
Downloaded from: http://www.leeds.ac.uk/cedars/cedars2/index.htm (October 1999).

Chapman, Steve; Paul Conway and Anne R. Kenney. "Digital Imaging and Preservation Microfilm: The Future of the Hybrid Approach for the Preservation of Brittle Books."
Downloaded from: http://www.clir.org/programs/cpa/hybridintro.html#full (July 1999).

Day, Michael. "Metadata for Digital Preservation: An Update."
Downloaded from: http://www.ariadne.ac.uk/issue22/metadata/intro.html (March 2000).

Elkington, Nancy. "Joint RLG and NPO Conference on Guidelines for Digital Imaging." Published in *RLG DigiNews*, 15 October 1998, Volume 2, Number 5.
Downloaded from: http://www.rlg.org/preserv/diginews (June 1999).

Granger, Stewart. "Emulation as a Digital Preservation Strategy"
Downloaded from: http://www.dlib.org/october00/granger/10granger.html (May 2001).

Green, Ann; JoAnn Dionne, and Martin Dennis. "Preserving the Whole: A Two-Track Approach to Rescuing Social Science Data and Metadata." Washington, D.C.: The Digital Library Federation, 1999.
Downloaded from: http://www.clir.org/pubs/reports/pub83/pub83.pdf (September 1999).

Hedstrom, Margaret. "The Role of National Initiative in Digital Preservation." Published in *RLG DigiNews*, 15 October 1998, Volume 2, Number 5.
Downloaded from: http://www.rlg.org/preserv/diginews (June 1999).

Kahle, Brewster. "Preserving the Internet." Published In *Scientific American*, March 1997.
Downloaded from: http://www.sciam.com/0397issue/0397kahle.html (May 2001).

Kenney, Anne. "The Cornell Digital to Microfilm Conversion Project: Final Report to NEH." Published in *RLG DigiNews*, 15 April 1997, Volume 1, Number 2. Downloaded from: http://www.rlg.org/preserv/diginews (June 1999).

Lynch, Clifford. "Canonicalization: A Fundamental Tool to Facilitate Preservation and Management of Digital Information." Published in *D-Lib Magazine*, September 1999, Volume 5, Number 9. Downloaded from: http://www.dlib.org/dlib/september99/09lynch.html (18 September 1999).

Moore, Reagan, et al. "Collection-Based Persistent Digital Archives - Part 1" Downloaded from: http://www.dlib.org/march00/moore/03moore-pt1.html (March 2000).

Moore, Reagan, et al. "Collection-Based Persistent Digital Archives - Part 2" Downloaded from: http://www.dlib.org/April00/moore/04moore-pt2.html (May 2001).

Rieger, Olga. "Digital Archiving: Approaches for Statistical Files, Moving Images and Audio Recordings." Published in *RLG DigiNews*, 15 December 1998, Volume 2, Number 6. Downloaded from: http://www.rlg.org/preserv/diginews (June 1999).

Russell, Kelly. "The Cedars Project: Why Can't We Preserve Everything?" Downloaded from: http://www.leeds.ac.uk/cedars/documents/ABS01.htm (July 1999).

Russell, Kelly. "Digital Preservation: Ensuring Access to Digital Materials Into the Future." June 1999. Downloaded from: http://www.leeds.ac.uk/cedars/Chapter.htm (November 2000).

Russell, Kelly and Derek Sargeant. "The Cedars Project: Implementing a Model for Distributed Digital Archives." Published in *RLG DigiNews*, 15 June 1999, Volume 3, Number 3. Downloaded from: http://www.rlg.org/preserv/diginews (June 1999).

Society of American Archivists. "Statement on The Preservation of Digitized Reproductions." Downloaded from: http://www.archivists.org/statements/digitized.html (July 1999).

van der Werf-Davelaar, Titia. "Long-term Preservation of Electronic Publications: The NEDLIB Project." Published in *D-Lib Magazine*, September 1999, Volume 5, Number 9.
Downloaded from:
http://www.dlib.org/dlib/september99/vanderwerf/09vanderwerf.html (September 1999).

Bibliographies

Day, Michael. *Preservation of Electronic Information: A Bibliography*.
http://homes.ukoln.ac.uk/-lismd/preservation.html (Last visited: 27 October, 1999)

Schnare, Robert E., Jr., et al. *Bibliography of Preservation Literature, 1983-1996*. Lanham, MD and London: Scarecrow Press, 2001.

Society of American Archivists. *Selected Readings in Preservation*, 1995-1998.
http://dlis.gseis.ucla.edu/saapreserv/text/bib9596.htm and
http://dlis.gseis.ucla.edu/saapreserv/text/bib9798.htm

Non-English Publications and Translations

1. Spanish translations of relevant CPA/CLIR reports [See below]

2. Spanish translation of parts of the *NEDCC Manual*: *El manual de preservacion de bibliotecas y archivos del Northeast Document Conservation Center*. The Manual was translated at the National Library of Venezuela as number 7 in the series, *Conservaplan*. The fascicles of interest for this bibliography are:

Fasciculo 1 *Prioridades de preservacion* (Planning and prioritizing)

Fasciculo 5 *Cambios de formato* (Reformatting)

Conservaplan is part of a larger program for the dissemination of information on conservation and preservation in Spanish. These publications are part of a larger program at the National Library of Venezuela to support the activities of the IFLA-PAC (Preservation and Conservation) Regional Center for Latin America and the Carribbean.

3. Other Spanish preservation translations of the National Library of Venezuela in the*Conservaplan* series include the following relevant CLIR publications:

Number 6 *El cuidado de archivos fotograficos*, T.J. Collings (Archival care of still pictures)

Number 8 *El cuidajo y manejo de grabaciones sonoras*, Gilles St-Laurent. (The care and handling of recorded sound materials)

Number 9 *Del microfilm a la imagen digital Informe de la Yale University*, Donald Waters (From Microfilm to Digital Imagery)

Number 10 *Almacenamiento y manipulacion de cintas magneticas: guia para bibliotecas y archivos*, John W.C. Bogart (Magnetic tape storage and handling)

Number 11 *Redefiniendo la preservacion del cine. Un plan nacional: Recommendaciones de Director de la Library of Congress en consulta con el National Film Preservation Board*, (Redefining Film Preservation: A National Plan)

Number 12 *Guia del IPI para almacenamiento de peliculas de acetato*, James M. Reilly (IPI storage guide for acetate film)

Number 13 *Tecnologia de Preservacion y Acceso. La relacion entre la tecnologia de conversion digital y otros procesos de conversion de medios: Glosario estructurado de terminos tecnicos*, M. Stuart Lynn (Preservation and access technology: The relationship between digital and other media conversion processes)

*****This includes primarily publications from 1993-1999. Some important and relevant earlier works are also cited.

Rev. October 1, 2001

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Media

Draft Appendix

October, 2001

P C Hariharan

Media

Historical evidence for written records dates from about the middle of the third millennium BC. The writing is on media[1] like a rock face, cave wall, clay tablets, papyrus scrolls and metallic discs. Writing, which was at first logographic, went through various stages such as ideography, polyphonic syllabary, monophonic syllabary and the very condensed alphabetic systems used by the major European languages today. The choice of the medium on which the writing was done has played a significant part in the development of writing. Thus, the Egyptians used hieroglyphic symbols for monumental and epigraphic writing, but began to adopt the slightly different *hieratic* form of it on papyri where it coexisted with hieroglyphics. Later, *demotic* was derived from hieratic for more popular uses. In writing systems based on the Greek and Roman alphabet, monumental writing made minimal use of *uncials* and there was often no space between words; a soft surface, and a stylus one does not have to hammer on, are conducive to cursive writing.

Early scribes did not have a wide choice of media or writing instruments. Charcoal, pigments derived from mineral ores, awls and chisels have all been used on hard media. Cuneiform writing on clay tablets, and Egyptian hieroglyphic and hieratic writing on papyrus scrolls, permitted the use of a stylus made from reeds. These could be shaped and kept in writing trim by the scribe, and the knowledge and skill needed for their use was a cherished skill often as valuable as the knowledge of writing itself.

Historically, a characteristic of human writing, which distinguishes it from *electronic* recording, is that it involved the transfer of *mass*, either from the writing surface (as in etching and carving), or to it, in the form of ink, from a pen or brush. The writing is perceived as a difference in color, or as a difference in texture, of the written portions of the surface. Both in Rome and Greece, the wax tablet, which was a piece of board with a very shallow depression covering almost its entire area and coated with a thin layer of beeswax, was used for making non-permanent records, and this may be viewed as the predecessor to rewritable media. Edison experimented with recording sound on a wax-coated cylinder, but settled on tinfoil instead. Today, the most widely used distribution medium for audio and computer software, the Compact Disc (CD) and the CD-ROM (Read-Only Memory), are replicated using a technique which is based on the removal of mass from the media; information is recorded on a CD in the form of pits, just as they were on the 70 rpm records, and its successors, the LP records. However, the pit diameter, the track pitch and the pit depth are microscopic, and the fidelity with which millions of copies of such an intricate and complex pattern can be faithfully reproduced is a tribute to modern technology. In the not too distant future, it is likely that there will be further reductions of two or three orders of magnitude in the size of the features or domains, which store the information. The Rosetta HD-ROM (High Density Read-Only Memory) from Norsam Technologies of Santa Fe NM relies on milling using an ion beam, and can reproduce features of the order of 25 nm on a number of substrates, from plastic polymers to silicon to metal. With the exception of CD, DVD, CD-ROM, DVD-ROM and HD-ROM, the electronic recording methods use an *energy transfer* process for information transfer rather than the traditional mass transfer process.

Analog and digital recording in use today are based on either permanent magnetism (*ferromagnetism*), or the use of a laser (*optical recording*) to etch away material, or cause a chemical or physical change. Magnetic media are available as rigid disks, floppy disks, and flexible tape, of various widths, wound on an open reel, or enclosed in a cartridge

[1] Since the 1970's, the custom has been to use the plural *media*, to stand for both singular and plural.

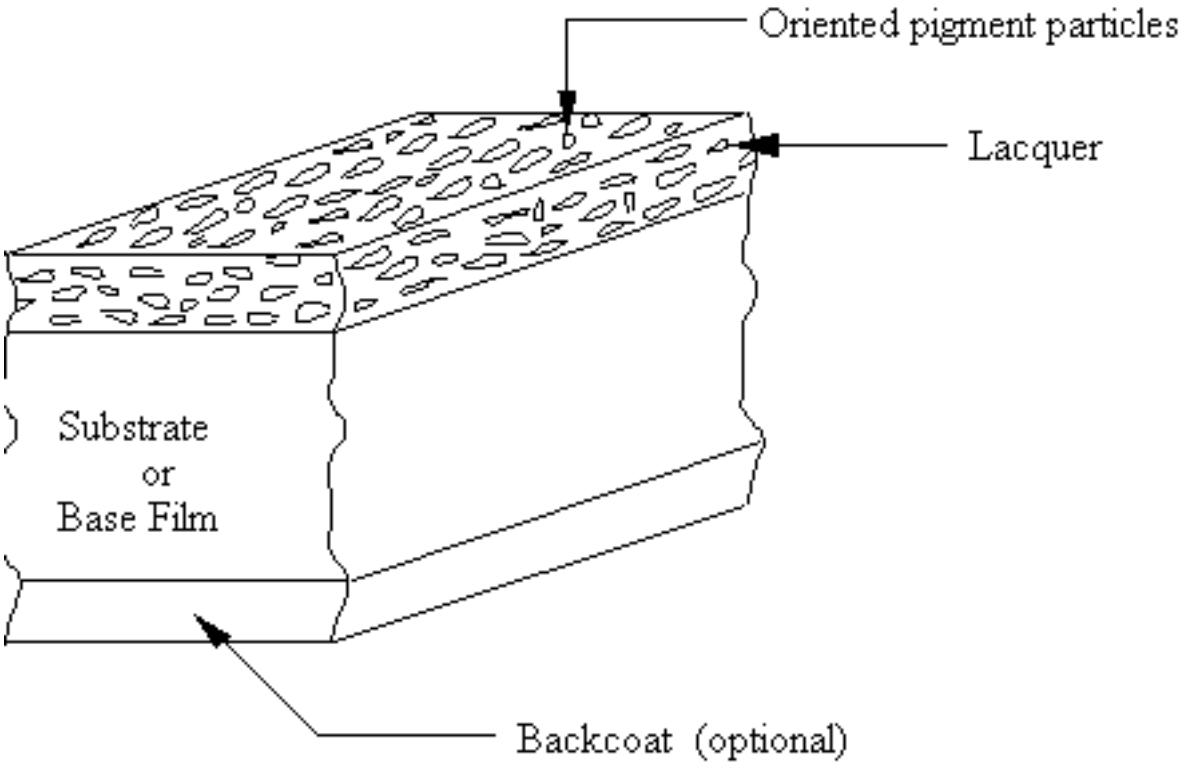


Figure 1-1: Cross-sectional view of magnetic tape

containing a single reel, or in a *cassette* containing two reels (supply and take-up reels). Optical media are available only in the circular disk format, but may enter the market in a tape format soon.

For users of media, a number of characteristics of the technology are important: linear density, areal density, volumetric density, transfer rate, reliability, error correction, stability of medium and longevity of technology. The mass transfer methods have proved, over historic times, the stability of both the technology and the media, which were used. On the other hand, in about a century over which magnetic recording has been employed, its susceptibility to decay has been noted by almost everyone who has used it. The shorter lifetime is not due just to media characteristics, but is also related to the short lifetimes of particular technologies.

While rock carvings have survived for millennia under exposed conditions (*e.g.*, the Behistun inscriptions of Darius I), and frangible material like papyrus was preserved in arid climates such as the Egyptian desert, twentieth century electronic media are less durable, and require storage under controlled conditions. The complex nature of the media is one reason. Table 1 compares two commonly used media of the current decade, magnetic tape and CD (Compact Disk).

	Magnetic tape	CD
Number of layers	2 or 3	3
Substrate	PET	Polycarbonate
Signal carrier	γ-Fe_2O_3 MP	Metallic backing on polycarbonate
Signal encoded in	Magnetic domains	Reflectivity change

Table 1-1.

Compared to metal, papyrus or clay tablets, magnetic tape and CD (Figs 1-1 and 1-2) are made up of layers with different characteristics. Each of them reacts differently to changes in the most basic environmental conditions: temperature and humidity. Manufacturers claim a hundred year life for CDs, but magnetic tape will last barely 15 years unless stored in carefully controlled environments. Polethylene terephthalate (PET), the substrate used in magnetic tapes, has been estimated to have a lifetime of a thousand years [Smith, Brown and Lowry 1982, 1983], but the complex soup which goes to make up the lacquer containing the magnetic pigment is susceptible to hydrolytic degradation of the binder. It is doubtful, however, if the devices required for playback will be available for longer than a decade after introduction of the technology.

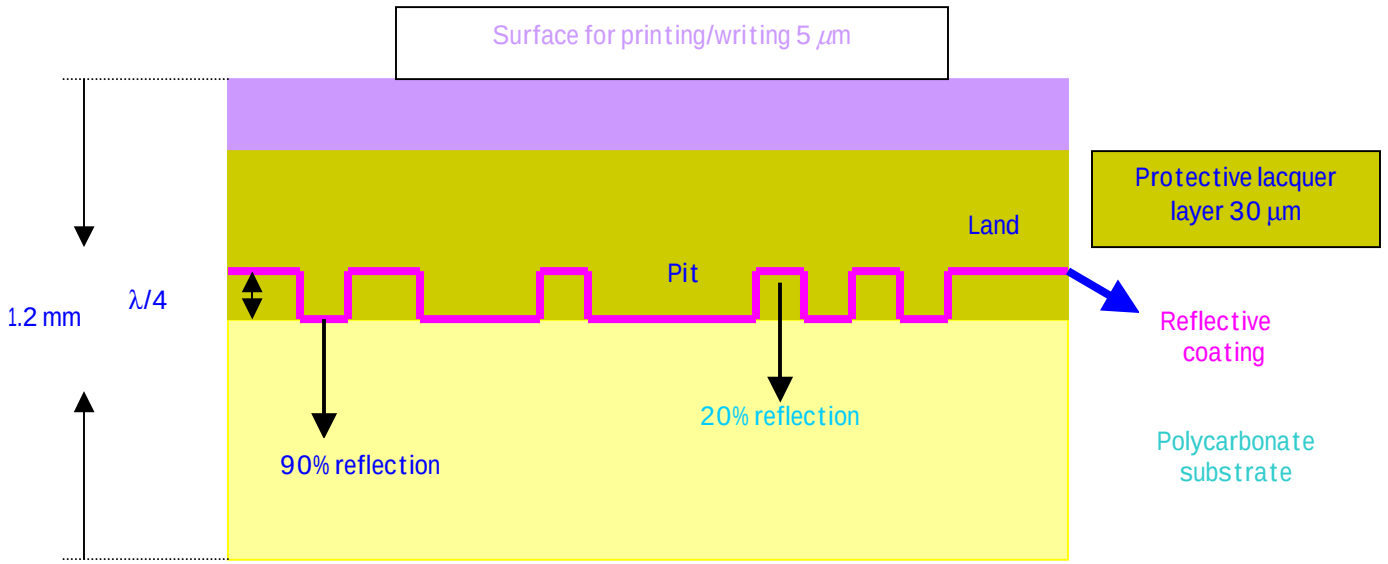


Figure 1-2. Cross-sectional view of Compact Disc

Non-human readable recording can be *analog* or *digital* (Fig #a). Analog recordings can be produced both on magnetic and on optical media. The response of an analog device is generally proportional to the input stimulus; in other words, the output is a *linear* function

of the input. The response of a digital device is of the on-off, abrupt kind, compared to the smooth, continuous response from an analog one; stated differently, a digital device responds in a *non-linear* fashion to its input.

Analog recordings lose their fidelity as they are copied, and the loss is more noticeable with the later generations. Washed out colors on a VHS tape, which has been copied from another VHS copy, exemplify this situation. Even the original recording can become corrupted as the recording does not have error correction, or even detection, built in. Digital recordings, on the other hand, can be copied from generation to generation with little loss of data integrity. This is possible because digital data, represented as *bytes* or *words*, can be subjected to arithmetic and logical operations to generate extra *CRC* (cyclic redundancy check) or parity bits or bytes and, when these are stored and retrieved with the original data, the operations can be reversed to determine if an error has occurred during readback and possibly even corrected. Error correcting codes (ECC) are designed to handle both *random* errors (also called *bit* errors) and *burst* errors (when a whole sequence of adjacent bits can either not be read, or the error extends over a sequence of bits). The CD-ROM ECC, for example, can handle data missing due to a 4 mm scratch on the disk surface. Appendix B gives an overview of the calculation and layout of the data and ECC bytes for the Ampex DD-2 tape system.

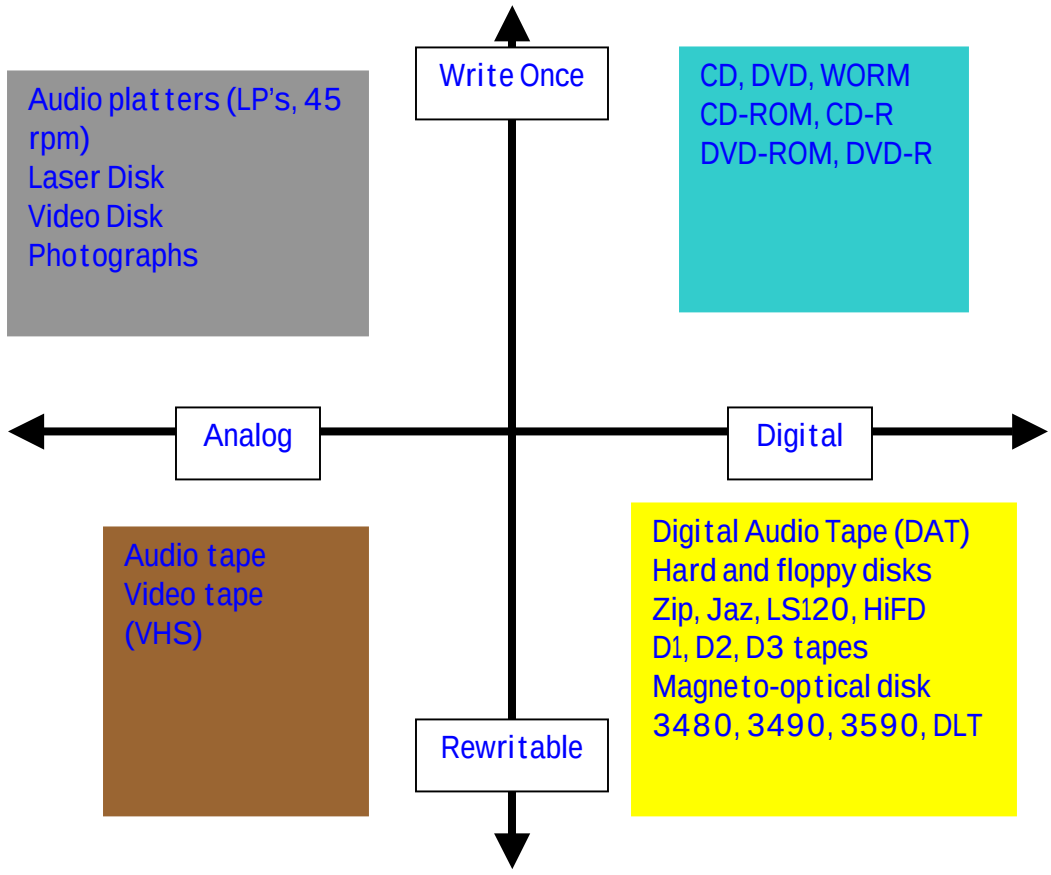


Figure 1-3. Two-dimensional diagram of data recording

Unlike printed books, recording technology available today can be used to store not just text drawings and paintings, but also sound, animation and moving pictures, separately and in combination. These recordings can be made on both analog as well as digital technology, but only the digital format is capable of producing clone-like copies of the original recording. The digital format introduces a number of processing layers.:

1. Size of word and order of bits within the word (*byte length, endian order*)
2. Selection of error correction code (ECC): number of levels of ECC, interleaving
3. Channel code (also called modulation code): this is designed to reduce the dc content (over a given sequence of bits, the number of 1's must equal the number of 0's), enable clocking (which tells the detecting system the starting and ending points of the receiving window), and reduce or eliminate the effects of *intersymbol interference* (ISS).

These requirements, making up the *physical format*, impose a heavy processing load on the recording system, and the processing itself is accomplished using software (in the form of *firmware* in a Read-Only Memory [ROM] in the recorder). This is the *digital overhead.* The application, which creates the records on the media, may impose a *logical* format, *e.g.*, a file structure. Together, the physical and logical formats add to the complexity and to the difficulty of preserving the data. *Recovery of the bit stream from a media is not sufficient for the recovery of the data committed to the media*.

A simple example will illustrate the complexity. CD-ROMs use an 8-bit byte as the data unit. 680,000,000 user bytes (~650 Mebibytes, see Appendix A on units and prefixes) can be recorded on a CD-ROM. We will denote this by *C*. The ECC overhead is 13%, and increases the number of bytes needed to be recorded to ~770,000,000 (1.13*C*). The channel code is Eight to Fourteen (EFM) and 3 additional merge bits are used; each 8 bit-byte is written as 14+3 = 17 bits on the CD. The total number of bits written to the CD is therefore $770{,}000{,}000 \times 17/8 \approx 1{,}636{,}630{,}000$. The numbers are approximate, and no attempt has been made to show how the bytes are ordered, nor the steps carried out to generate the ECC bytes. A somewhat longer code walk-through for the generation of the bit-stream for the Ampex DD-2 tape system is outlined in Appendix B.

2. Magnetic Recording

Valdemar Poulsen demonstrated magnetic recording in 1898 in Copenhagen, Denmark. His set-up is described in Jorgensen [Finn Jorgensen: The Complete Handbook of Magnetic Recording]. As with other electronic equipment, recording devices were mostly in studios and laboratories at first. After WW II, electronics increasingly expanded into the consumer entertainment segment, and round-reel recorders for music replay began to be available. In the early 50's, IBM and UNIVAC began selling digital computers in the US. Magnetic tapes were the ubiquitous interchange media in the days before networking. As computer peripherals, they have been around since 1952, and predate the disk drive by five years; thus they also served as the earliest secondary storage media, giving way, gradually, to the magnetic drum and, later, to the magnetic disk. Table 1 lists the various ½-inch tape drives that IBM, once the principal peripheral provider, has introduced.

Year	Product	Package Geometry	Pigment	Linear Density* Bpi	Transfer rate KB/s	Capacity MB	Track density tpi
1952	IBM726	Round	γ-Fe_2O_3	100	7.5	1.4	14
1953	IBM727	Round	γ-Fe_2O_3	200	15	5.8	14
1957	IBM729	Round	γ-Fe_2O_3	800	90	23	14
1965	IBM2401	Round	γ-Fe_2O_3	1600	180	46	18
1968	IBM2420	Round	γ-Fe_2O_3	1600	320	46	18
1973	IBM3420	Round	γ-Fe_2O_3	6250	1250	180	18
1984	IBM3480	Square	CrO_2	38000	3000	200	36
1991	IBM3490E	Square	CrO_2	76000	3000	800	72
1995	IBM3590	Square	MP-II	92000	9000	10000	256

*This is the user bit density; the number of flux reversals per inch (fci), which is not indicated here, will include the channel (modulation) code, ECC (error correction and control) and other overhead is higher.

Table 2-1. From Round Reel to Square – the evolution of tape technology from IBM

This market, however, is not restricted to just ½-inch tapes; other widths in use in the computer segment are: 4 mm (called DDS and derived from the Digital Audio Tape), ¼-inch (called Quarter Inch Cartridge or QIC, and, more recently, Travan), 8 mm (Exabyte, Mammoth, Sony Advanced Intelligent Tape, Ecrix), and 19 mm (ID-1, DD-2). A representative list appears in Table 2.

On any medium, data are written along *tracks.* On disks, the tracks can be either concentric circles, or one long spiral, either right-handed or left-handed, moving out from the center towards the outer edge of the disk. In the case of tapes, there are three methods in use currently for track layout. The simplest is *longitudinal* where the tracks define lines along the length of the tape, see Fig 2-1. This corresponds to the way books are printed in most west- European languages.

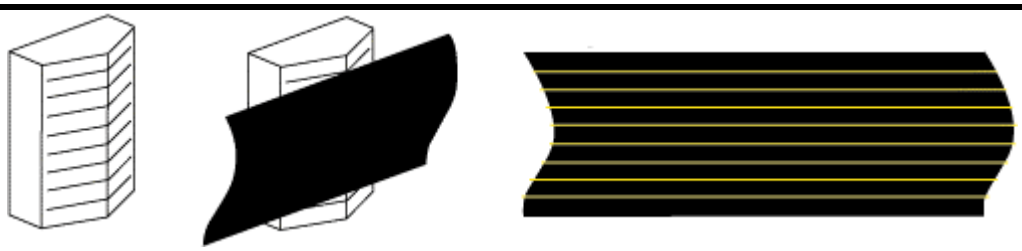


Figure 2-1. Longitudinal tracks

Type	Drive Manufacturer	Product Name	Capacity* (GB)	Transfer Rate (MB/s)	Recording Format
19mm	Ampex	DD-2	50(S)	15	Helical
19mm	Ampex	DD-2	150(M)	15	Helical
19mm	Ampex	DD-2	330(L)	15	Helical
19mm	SONY	ID-1	8.7	2 - 64	Helical
19mm	SONY	ID-1	43	2 - 64	Helical
19mm	SONY	ID-1	96	2 - 64	Helical
1/2 in*	IBM	3590	10	9	Serpentine
1/2 in*	STK	Redwood	50	11	Helical
1/2 in*	STK	Eagle 9840	20	10	Serpentine
1/2 in*	Quantum	DLT2000	15	1.25	Serpentine
1/2 in*	Quantum	DLT4000	20	1.5	Serpentine
1/2 in*	Quantum	DLT7000	35	5	Serpentine
1/2 in*	SONY	DTF-1	42	12	Helical

1/2 in*	SONY	DTF-2	100	24	Helical
1/2 in*	SONY	DIR-120	38(S)	15	Helical
1/2 in*	SONY	DIR-240	38(S)	30	Helical
1/2 in*	SONY	DIR-120	125(L)	15	Helical
1/2 in*	SONY	DIR-240	125(L)	30	Helical
1/2 in*	LMS	NCTP	18	10	longitudinal
1/2 in*	HP, IBM, SEAGATE	LTO			
8 mm	EXABYTE	Mammoth	20	3	Helical
8 mm	SONY	AIT-1	25	3	Helical
QIC^	TANDBERG, SEAGATE,HP	QIC-4	4	0.8	Serpentine
QIC^	TANDBERG,SEAGATE,HP	QIC-5010	16	1.5	Serpentine
QIC^	TANDBERG, SEAGATE,HP	QIC-5020	25	4.5	Serpentine
4 mm	HP, SONY, EXABYTE	DDS-1	2	0.183	Helical
4 mm	HP, SONY, EXABYTE	DDS-2	4	0.366	Helical
4 mm	HP,SONY, EXABYTE	DDS-3	12	1.5	Helical

*S, M, L – Small, Medium, Large denote cassette sizes

Table 2-2. The variety of tape formats

The video industry introduced the *helical scan tracking* format, Fig 2-2. Instead of the tracks running parallel to the long edges of the tape, they subtend an angle of ~5° to the edge. The chief advantage of this format is its ability to use most of the tape area for data recording, *i.e.*, an increased *areal efficiency*. DAT, VHS and Exabyte tapes are examples of technologies using helical scan recording. The recording/playback heads are mounted on a drum which rotates at about 3600 rpm while the tape itself moves forward at a few meters per minute. The relative velocity between the head and the tapes, however, can reach values as high as 50 km/hr (about 30 mph). It can be seen from the figure that the tape is wrapped around the drum at a shallow angle, and the wrap angle (the angle between the entrance and exit points of the tape wrap) is between 90° and 180° depending on the particular technology. Contact between tape and heads is extremely close on helical scan drives, and the high relative velocity can cause abrasive wear in both head and media.

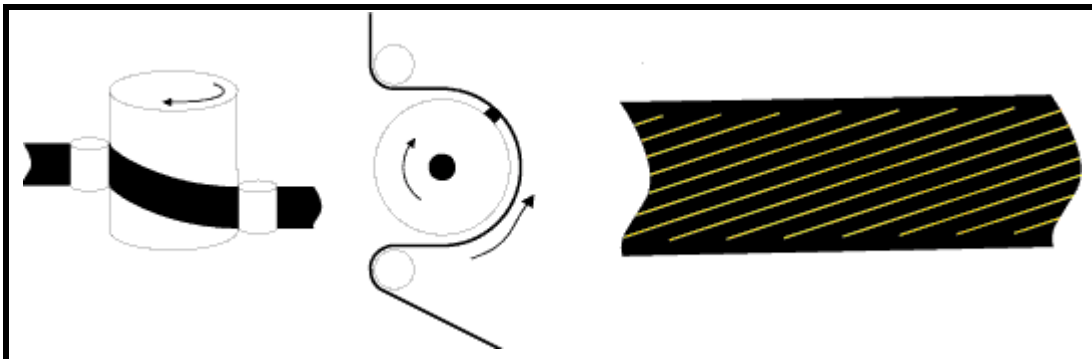


Figure 2-2. Helical scan recording

The third recording format is *serpentine* which is known as *boustrophedon* when applied to human writing. The serpentine format, illustrated in Fig 2-3, is very much like longitudinal except that the tracks change direction when they reach the end of the tape.

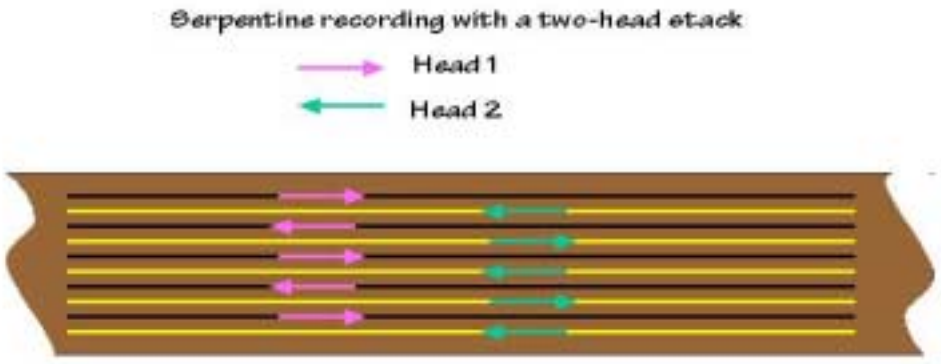


Figure 2-3. Serpentine Recording

There is one recorder, the DCSRi, from Ampex, which uses *transverse* recording. The DCSRi uses 1-inch wide tape, which passes over a canoe while in contact with a cylindrical

read/write head (see Fig 2-3). This results in a set of tracks being written which are perpendicular to the long edges of the tape.

Figure 2-4. Transverse Recording

Arctuate track recording has been tried out in the laboratory, but has never been brought to market in a product. Figure 2-5 shows an example of arctuate recording. The read/write heads are mounted on a circular disc rotating in a plane parallel to that of the moving tape. The rotational motion of the disc, combined with the linear movement of the tape, results in tracks in the shape of an arc being written on the tape.

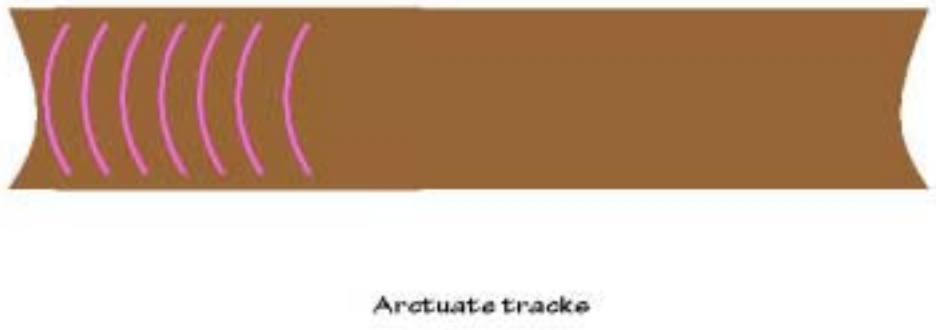


Figure 2-5. Arctuate Recording

Tape drive prices range from just above a hundred dollars for low-end QIC and Travan systems with EIDE interfaces to well over a quarter of a million dollars for high transfer rate 19 mm helical scan recorders. Media costs average from a few dollars for 3480-type cartridges to a couple of hundred dollars for high-capacity DD-2 cassettes.

Inspection of Table 2-1 shows that both linear bit density (bpi) and capacity (MB) have increased by a factor of 10^4 between 1952 and 1995, a period of 43 years. In the same time span, track density has increased only modestly, from 14 tpi to 256 tpi. Helical scan recording packs tracks more closely together, and the track density is almost an order of magnitude larger than in longitudinal recording. Transfer rates have lagged capacity and linear density increases, and has grown only by a factor of 10^3. Large DD-2 cassettes are already approaching a TB in capacity; if the drives can only write to them at 15 MB/sec, it takes about 66 sec to write a GB if data can be streamed to the drive at this rate in a sustained fashion. A full cassette will then require about 6600 sec to be filled, or read completely.

In the semiconductor industry, Moore's Law (named after Gordon Moore, one of the founders of Intel Corporation) states that the number of components on an integrated circuit (IC) doubles every 18 to 24 months. Since 1990, areal density of magnetic disks has been doubling every 18 months. More recently, magnetic tapes have been able to adapt to this pace too. In September 1999, IBM announced a laboratory demonstration of 35.5 Gb/in^2 for disks, and products at this density are likely to be available in the market within the next 3 years. This raises an interesting question: what is the ultimate areal density, which can be achieved in magnetic recording. The answer is not straightforward. Using current technology, it is very likely that the *superparamagnetic* limit will restrict densities to be below ~100 Gb/in^2. The industry is already using single-domain grains; as their size decreases, thermal effects can cause orientational instability in the domains and this leads to the loss of recorded data. With the exception of magneto-optical recording, all other forms of magnetic recording result in the creation of magnetic regions, which lie in the plane of the media. Work has been going on in the last two decades, mostly in Japan, on *perpendicular* recording. In this technique, the magnetic dipoles are oriented at right angles to the media surface, and it may permit the density to go beyond the superparamagnetic limit for in-plane magnetization. However, ring heads can no longer be used for writing; the two poles of the inductive write head need to be on opposite sides of the media being recorded. Alternatively, a magnetically soft material may be used as an underlay to complete the flux circuit for a modified ring head. Barium Ferrite ($BaO.6Fe_2O_3$), a green pigment often used in the lining on refrigerator doors to keep them closed, is an ideal material for perpendicular recording.

3. Optical recording

Optical recording is a term which is used for a number of technologies whose common characteristic is the use of a laser, principally for reading. It embraces both the

mass-transfer and energy-transfer methods. The compact disc (CD), introduced in 1982, can be mass-produced because, once a glass master has been made, copies can be reliably and rapidly manufactured by injection molding. On a CD, a spiral track, ~5 km long, holds data in the form of *pits* of various lengths, with pieces of intervening *land* of different lengths. The pit depth is $\lambda/4$, where λ is the wavelength of the laser used for reading (for *CD circa* 780 nm, which lies in the infra-red). The light reflected from the bottom of a pit is 180^{o} out of phase with the incident radiation, and interferes destructively with it. The net result is a substantial decrease in the intensity if light reflected from the bottom of a pit; about 90% of the laser light is reflected from a *land* area, whereas the reflectivity of a pit is only about 20%. DVD (Digital Versatile Disk) uses the same principle, although the wavelength of the laser light is smaller. Table 3-1 summarizes the principal characteristics of CD and DVD.

	CD	DVD
Disc diameter (mm)	80, 120	120
Number of sides	1	2
Thickness per side (mm)	1.2	0.6
Track pitch (μm)	1.6	0.74
Minimum pit length (μm)	0.834	0.4
Laser wavelength (nm)	780	640
Data capacity per layer (GB)	0.68	4.7
Number of layers	1	1, 2 or 4
Numerical aperture	0.6	0.38 to 0.45
Modulation code	8/14 (8/17 with merge bits)	8/16
User data rate, 1X (Mbps)	1.41	11.08
Error correction	CIRC	RS-PC
Error correction overhead %	34	13
Format overhead %	252	136
Scanning speed (m/s)	1.2 to 1.4	3.49 to 3.84
Rotation speed (1X) (rpm)	200 to 500	570 to 1600

Table 3-1. Principal characteristics of CD and DVD

Although the CD and DVD disks have the same diameter and thickness, the storage capacity of DVD, on a single layer, is almost seven times that of a CD. This is achieved through the use of laser with a shorter wavelength, a more efficient channel code and stronger error correction code. The diameter of the spot which can be produced using a source of wavelength λ is given by the expression

$$d = C\lambda/NA$$

where *NA* is the numerical aperture of the lens and *C* depends on the distribution of the energy of the beam at the focal plane. *C* is 0.6 for lenses used in CD and DVD drives.

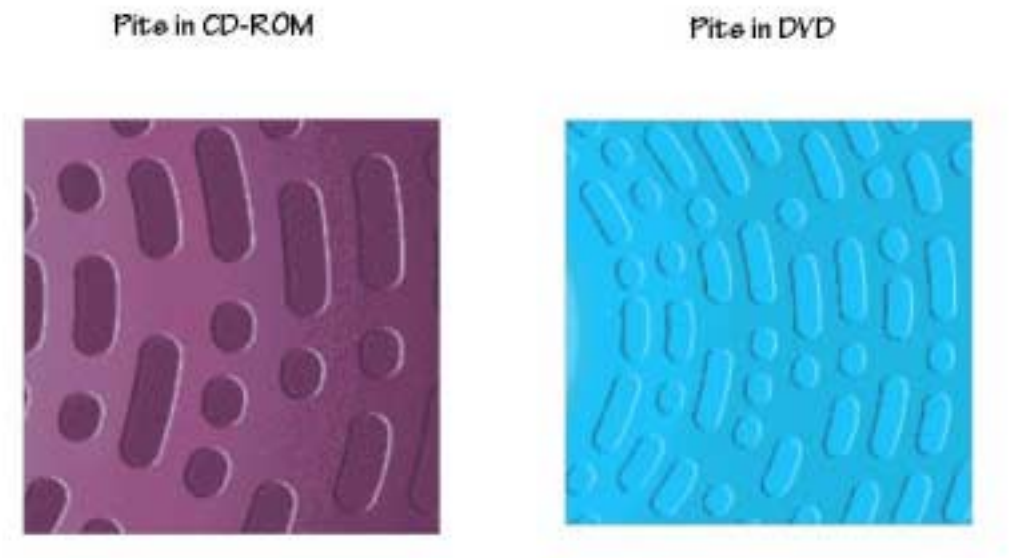


Figure 3-1. Pit size in CD and DVD

It is possible to have reflectivity differences created in other ways than by removal of mass. In CD-R (CD-Recordable), this is achieved by a laser-induced reaction in a polymer dye; the ones generally used are cyanine and phthalocyanine dyes. The *irreversible chemical* reaction, which gives the user a Write Once Read Many (WORM) media, has regions of low reflectivity where the dye has undergone a reaction. CD-RW (Rewritable) media use a coating of an alloy which can change from an amorphous to a crystalline phase when illuminated by a laser. The crystalline regions have higher reflectivity than the amorphous areas. The amorphous to crystalline, and the reverse, *phase transformations* are *reversible.* Manufacturers, however, caution that material fatigue constrains their use to a maximum of about 1000 write cycles. Since the laser power used for reading is below the threshold for initiating a phase transformation, a CD-RW disc can be read more times than it can be written on.

Calimetrics, of Alameda CA, has developed a *pit-depth modulation* technique which is based on an *M-ary* code and promises to increase both the storage capacity and transfer rate by a factor of 3 or more. The technique can be used with both CD and DVD technologies.

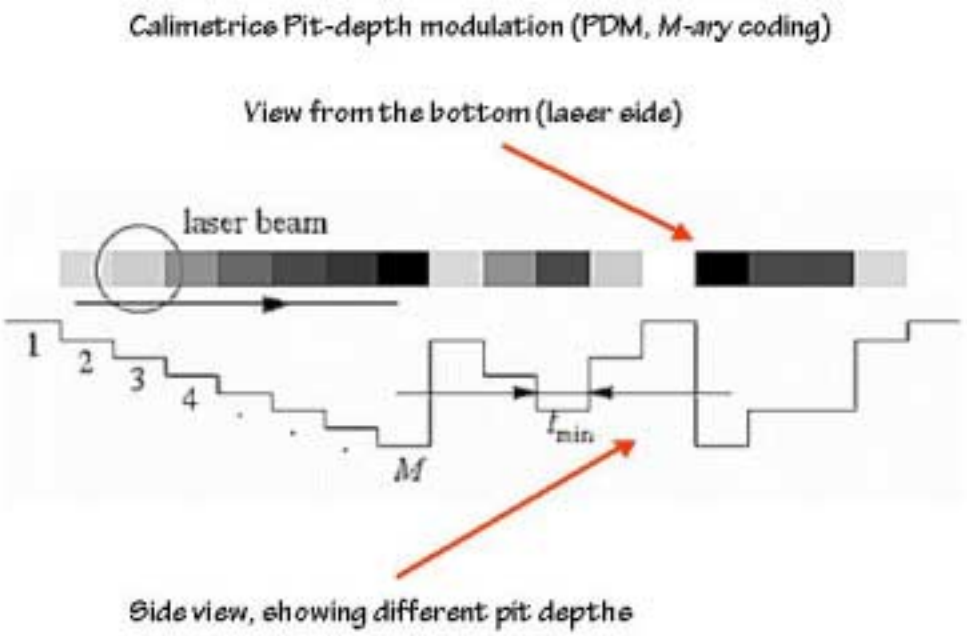


Figure 3-2. Pit depth modulation

Magnetic hard disks, and removable media like Jaz, Zip and magneto-optical disks, record data along a series of *concentric* circular tracks on the surface of the disk. Track density depends on the particular technology, but is about 20,000 tpi for hard disks. Hard disks rotate at a constant rate; it used to be 3600 rpm (revolutions per minute), but more recent products operate at 5400, 7200 or even 10,000 rpm. The length *l* of a track is related to its radius *r* by the expression

$$l = 2\pi r$$

Inner tracks, being shorter, hold less data than outer tracks. Hard disks, and most high-density removable disks, therefore use *Zoned Bit Recording* (ZBR), which subdivides the disk surface into zones of different linear densities. A 3.5" hard disk drive for a PC, for example, typically has about 15 zones. One consequence of ZBR is that the transfer rate to/from the disk is dependent on the location of the data. The highest transfer rates are achieved for the outer tracks, but the innermost tracks have the highest data densities. Vendors have been known to load and locate files on the outer tracks to boost perceived transfer rates in benchmark tests [I Bird, R Chambers, M Davis, A Kowalski, B Lukens, S Philpott and R Whitney, *Evaluating RAID in the Real World*, Sixth Goddard Conference on Mass Storage Systems and Technologies, GSFC/CP-1998-206850, pp 345-354]. CD and DVD, on the other hand, record along a *single, continuous, spiral* track, which starts from the inner region of the media and runs outward in a clockwise fashion. One advantage of starting the track at

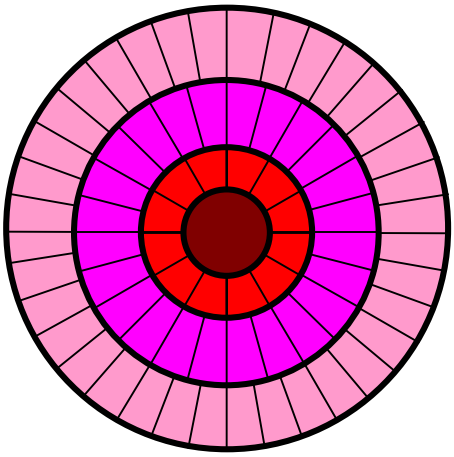


Figure 3-3. Zoned Bit Recording

the inner edge is that media of more than one diameter can be designed. Small CD disks, with a diameter of 80 mm, can be played on the same drives which are used for the 120 mm discs. Both CD and DVD drives use a technique called *constant linear velocity* (CLV) to read data from the tracks. The track passes under the read head at the same linear velocity irrespective of where on the disk surface the track segment is located; since outer tracks are longer than inner tracks, this means that the rotational speed of the CD is highest at the innermost track, and smallest at the outermost one and varies continuously from 200 to 500 rpm (for 1X drives) as the head moves in from the outer tracks. Rectangular CD disks, whose dimensions are close to those of a business card, with a capacity of 80 MB, are being used in place of traditional business cards by some organizations. The printed area contains the usual information, but these cards can be played on most multi-read CD-ROM drives to display multimedia images or play sound clips.

Recordable versions of CD exist in both WORM and rewritable formats; these are CD-R and CD-RW. Neither CD-R, nor CD-RW is designed for mass production. In CD-Rs, the signal carrier is a dye (cyanine or phthalocyanine, see Fig 3-1), which is polymerized by a high-power laser. Areas, which have undergone reaction, have a smaller reflectivity than the unwritten areas. However, the reflectivity values, and differences, are not the same as those for CD. Hence, CD-Rs and CD-RWs can be read only by those CD drives which have *multi-read* capability. Unlike the regular edges found for the pits on stamped CDs, the peripheries of the marks on both CD-R and CD-RW disks are more irregular. Another factor affecting the shape of the marks is the speed of the recorders; CD-R and CD-RW recorders are available at various recording speeds: 1X, 2X, 4X, 6X and 8X.

There are four versions of recordable DVD: DVD-R, DVD-RAM, DVD-RW and DVD+RW. The last three are also writable and are competing formats. DVD-RAM (1.0) divides the disk surface into 24 zones, and uses *Zoned Constant Angular Velocity* (ZCAV) for reading and writing. Within a zone, the angular velocity is fixed, but it changes when a new zone is accessed.

In 1989, seven years after the introduction of CD, the installed base exceeded 25 million drives in the US. By comparison, 70 million Sony playstations were sold worldwide in less than five years.

CD-ROM drives have become standard peripherals on all Personal Computers. Most software is sold on CD-ROMs, and there is a standard called *El Torito* which permits booting from a CD-ROM. The CD has become the analog of the phonograph record in the multimedia era; it has enjoyed a long and successful life in the entertainment market, and has managed to move into computer software distribution and file backup.

CD and CD-ROM media are among the most stable. The data is recorded on them as pits, and a thin layer of a reflective metallic layer (Al) is deposited on it before it is sealed in with another protective layer. The substrate, polycarbonate, is less susceptible to hydrolytic degradation than polymethylmethacrylate (PMMA) used in laser discs. *Laser rot*, oxidation of the reflective metallic layer, has been reported in some cases, but material control and modern manufacturing practices have reduced its incidence. In rewritable media, the signal carrier, a phase change alloy of Te, Sb, Ge and/or Sn, is in a finely-divided state and therefore susceptible to oxidation. The chances of this occurring are minimized by using a number of protective layers: ZnS/SiO_2 and UV-hardened plastic. It is more likely that the number of write cycles will be exceeded before the disc becomes unusable due to other causes. CD-R, and DVD-R, use a dye polymer. The dye is generally cyanine or Metalized phthalocyanine and there were early anecdotal reports of its being bleached by UV and blue lights. It is best to keep all CD disks away from direct light and heat. Manufacturers quote a lifetime of 100 years for CD and CD-ROM disks, and 30 years for CD-R.

Summary

Electronic recording is now just over a hundred years old, starting out as magnetic recording. Optical recording is only about 25 years old, and had to wait for the development and availability of lasers. Magnetic media are available in fixed and removable disks, and flexible tape loaded in cartridges and cassettes. Optical media, on the other hand, can be had today only as disks. While media, when stored under controlled conditions, can last for decades, the lifetime of a particular technology, especially in the magnetic recording world, is less than 10 years. Charts 4-1 and 4-2 summarize the growth and development of the two technologies.

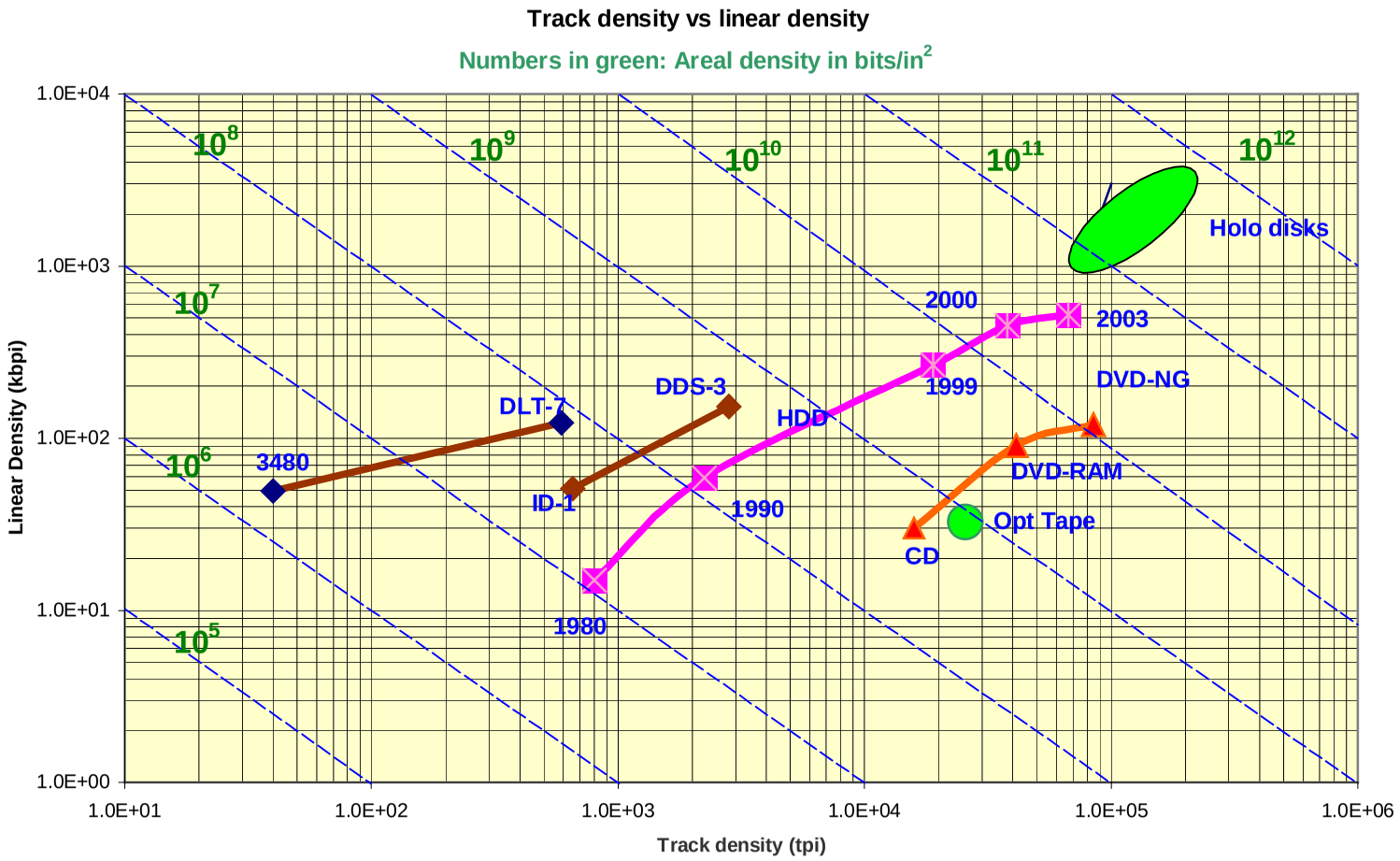
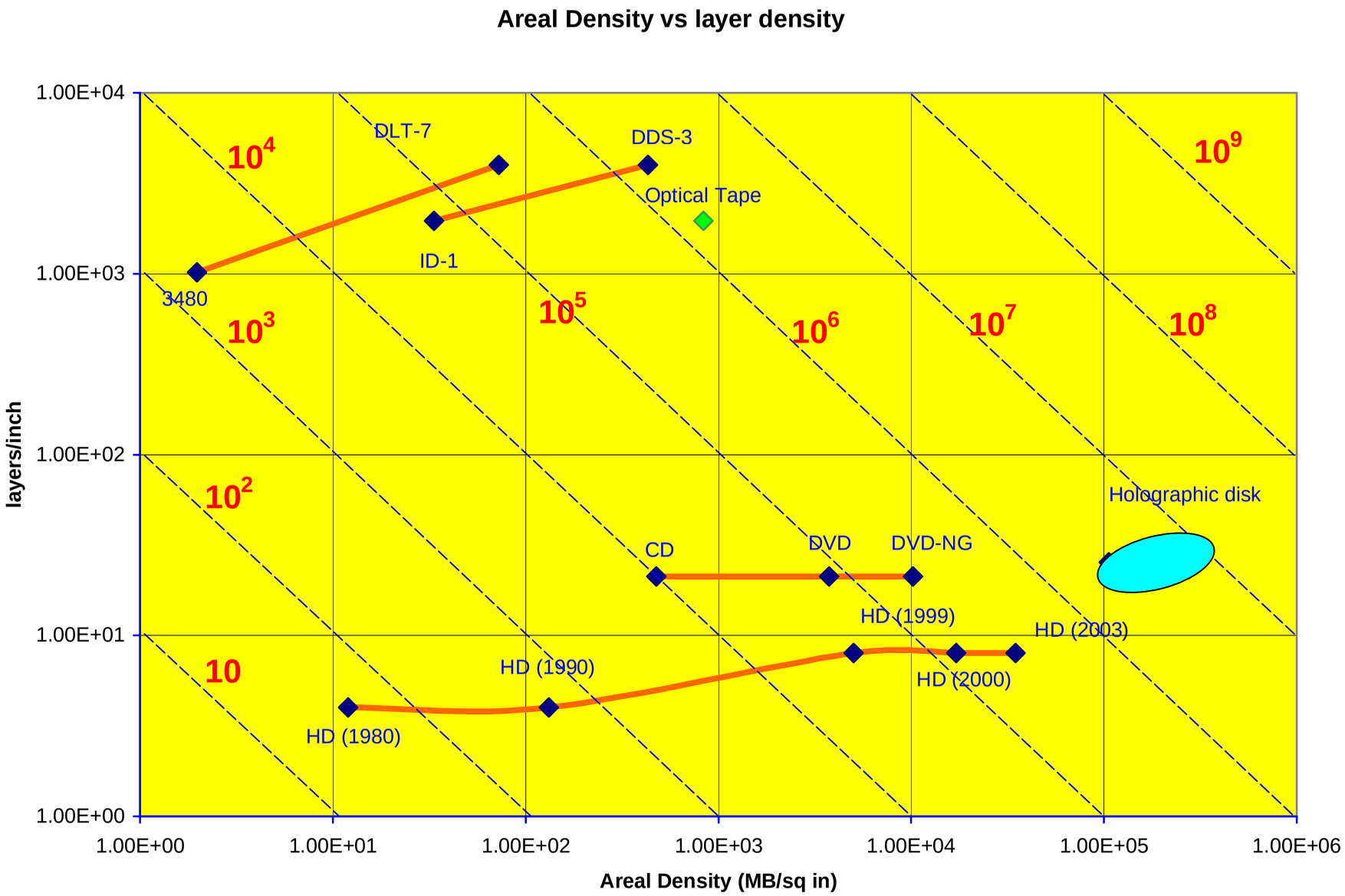


Figure 4-4. Growth of linear and track densities for various recording technologies


Areal Density vs layer density
1.00E+04
1.00E+03
1.00E+02
1.00E+01
1.00E+00
layers/inch
1.00E+00
1.00E+01
1.00E+02
1.00E+03
1.00E+04
1.00E+05
1.00E+06
Areal Density (MB/sq in)
10^4
10^3
10^2
10
10^5
10^6
10^7
10^8
10^9
DLT-7
DDS-3
Optical Tape
ID-1
3480
CD
DVD
DVD-NG
Holographic disk
HD (1999)
HD (2003)
HD (1990)
HD (2000)
HD (1980)

Appendix A. Prefixes for magnitudes

Prefixes for magnitudes are defined by the International Conference on Weights and Measures ***ICBP***, and published by the ***BIPM*** (Bureau International des Poids et Mesures).
These prefixes all refer to *powers of 10*. In the computer industry, without further explanation, these prefixes are also *sometimes* used as *powers of 2*. The prefixes are usually four letters, and end in *–o* for the negative powers (but see below for exceptions).

yocto	y	10^{-24}
zepto	z	10^{-21}
atto	a	10^{-18}
femto	f	10^{-15}
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
mill*i*	m	10^{-3}
cent*i*	c	10^{-2}
dec*i*	d	10^{-1}
deca	da	10^{1}
hecto	h	10^{2}
kil*o*	k	10^{3}
mega	M	10^{6}
giga	G	10^{9}
tera	T	10^{12}
peta	P	10^{15}
exa	E	10^{18}
zetta	Z	10^{21}
yotta	Y	10^{24}

More recently, it has been proposed that, for positive powers based on the radix 2, the prefixes should be modified so that the last two letters, or the last letter if the prefix has only three letters, are changed to *bi*; thus a me*bi*byte is 2^{20} = 1048756 bytes, but a megabyte is 10^{6} = 1000000 bytes.
Always watch the units; they can be tricky and treacherous, *e.g.*, ream.
1 Ream of paper = 480 sheets, usually
= 472 sheets for handmade and drawing paper
= 500 sheets for book paper and for flat-plate newsprint
= 516 sheets in a perfect ream

Appendix B. Code walk-through for DD-2

1. Accept a set of 1,199,824 bytes (source bytes). After encoding for error-correction, section identification, etc., this will grow to a size of 1,584,000 bytes and will be written out as a physical block on 32 adjacent helical-scan (HS) tracks.
2. Subdivide this set into 8 blocks of 149,978 bytes each.
3. Add a 2-byte checksum to each block, increasing its size to 149,980 bytes. The input byte stream has now grown to 1,199,840 bytes.
4. Combine the eight blocks into four data streams of 299,960 bytes each.
5. Augment each stream with 8 bytes of error detection information, increasing its size to 299,968 bytes and that of the original input stream to 1,199,872 bytes.
6. Encapsulate the data in the C3 (outermost) error correction code. Each group of 86 bytes is replaced by a 96-byte C3 codeword. This code, denoted (96,86,5) is capable of correcting 5 bytes in an incorrect codeword when the location of the errors is not known. If the positions of the erroneous bits are known, upto 10 bytes can be corrected.
7. The C3-corrected stream has 13952 C3 codewords; the original data stream, at this stage, has increased to 13952 X 96 = 1,339,392 bytes. Shuffle the bytes among the 32 tracks on which they will be ultimately written. This is accomplished as follows:

byte 1 → track 1	byte 33 → track 1
byte 2 → track 2	byte 34 → track 2
.	.
.	.
.	.
byte 32 → track 32	byte 64 → track 32

assigning 41856 (=1,339,392/32) bytes to each track.
8. Insert 192 bytes of logical format information in each track whose size consequently increases to 42048 bytes (total 1,345,536 bytes).
9. Encapsulate the data stream from step 8 in the middle, C2, error correction code. This (106,96,5) code replaces each group of 96 bytes with a 106-byte C2 codeword; each of the 32 tracks, at this stage, has assigned to it, 438 C2 codewords (438X106 = 46428 bytes). The total number of bytes, in the 32 tracks, is 1,485,696.
10. Interleave the C2 codewords to reduce the effect of burst errors. The data is arranged in 438 columns, and is read out row-wise to fill 32 rows, each of which will occupy a helical scan track.
11. At the end of each group of 219 bytes, insert a 1-byte identifier. This is essentially the ordinal number of the group on the helical scan track. There are 212 groups on each track. The track now holds, with the addition of the 212 1-byte identifiers, 46,640 bytes. The total number, for all 32 tracks, is 1,492,480 bytes.
12. Encapsulate the C2-encoded data stream in the inner, C1, code. This (228,220,4) code replaces each group of 220 bytes(the group from step 11) by a 228-byte C1 codeword.
13. Each C1 codeword is augmented by 3 bytes of synchronizing information, leading to a 231-byte sync block. The number of bytes assigned to each track at this stage is 212×231 = 48336.
14. The data is now ready to be written out on tape. Each track starts with a 174-byte preamble which is followed by a 18-byte track sync block, the 212 sync blocks from step 13, and is terminated by a 174-byte postamble. Each helical scan track now has 49500 user bytes.

15. At the outset, the user data stream of 1,199,824 bytes, when spread out over 32 helical scan tracks, averages to only 37494.5 bytes per track. However, after encapsulation in error-correcting codes, each track has 48336 bytes. The overhead for error correction is, therefore, 48336 - 37495 = 10841 bytes. This amounts to an overhead of 28.9%.
16. The choice of Miller squared code for the channel facilitates the identification of error locations. This code has a lower rate, 0.5, than the 8/9 code used in ID-1, but the lower rate provides additional error-detecting capability. The 49,500 bytes in the data, when converted to the Miller-squared modulation code, expand to 99,000 bytes.
17. The total overhead in the DD-2 system is 128.9%. One thousand bits of user data are committed to tape as 2289 bits.

InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

Survey of Preservation Practices and Plans

Draft Appendix

June 1, 2001

Michèle V. Cloonan and Shelby Sanett

Preservation Strategies for Electronic Records, Round 1 (2000-2001)
Where We Are Now: Obliquity and Squint?
A Report to the National Historical Publications and Records Commission
Michèle V. Cloonan and Shelby Sanett
University of California, Los Angeles
June 1, 2001

Introduction

Information has never before been as fugitive as it is today. Whereas records were once written on media which could last hundreds—or even thousands—of years, electronic records are in danger of disappearing, becoming physically unusable or legally inadmissible, almost immediately. There are many causes for the short life span of digital and electronic records: media deterioration, technological obsolescence, a paucity of standards and guidelines, and the failure of many managers to plan for the maintenance and preservation of electronic records. Besser refers to this last cause as "the custodial problem."[i]

Records created or maintained in electronic form, either analog or digital, are herein referred to as *electronic records*. Electronic records are the focus of the InterPARES Project. It is impossible to avoid using the term *digital preservation*, however, because that is the term most frequently used in the preservation literature. (See Appendix 4, Preservation Bibliography. Nearly every author cited in this report uses it.) We recognize that it is not always the most precise term for archivists because it refers to the preservation of reformatted items, born-digital electronic materials, and born-again digital materials. We will use the term *digital preservation* to reflect the general preservation discourse found both in the literature and in our interviews, as well as *preservation of electronic records*, when appropriate. We have provided a glossary below of terms used in this study. Recently, the term *long-term retention* seems to be gaining currency.[ii] As this term is more general than either *digital preservation* or *preservation of electronic records*, we will also use it where appropriate. We believe that the flux surrounding terminology reflects an evolving new field.

It is the responsibility of archivists to ensure the authenticity, reliability, and long-term accessibility of permanent electronic records.[iii] Traditionally archivists have done so by gathering documents, establishing provenance, and maintaining and demonstrating an unbroken chain of custody in an evidence-based approach to managing records.[iv] Is it possible to ensure authenticity and reliability of records regardless of their formats? If so, which technologies and techniques are archivists using to preserve electronic records at the time of their creation as well as throughout their life cycle? And how are archivists defining the term *preservation*?

Purpose

The purpose of this study is to identify and describe the strategies and techniques that are in use or in development, in institutions and on research projects, to preserve electronic records. Such knowledge should inform archivists as they formulate strategies, policies, and standards for preservation.

Preservation of electronic records is a young field. Indeed, some people believe that the very concept of electronic record preservation is an oxymoron. The term *preservation* implies permanence, yet such media are inherently unstable. Add to the ongoing problem of technological obsolescence such challenges as copyright and fair use; organizational and custodial issues; different stakeholders and interests; substantial financial requirements and legal admissibility requirements; and prospects for preservation of electronic records seem grim, indeed.

It is possible to develop models, protocols, and standards that factor in such problems as media deterioration and technological obsolescence, and work is already being done in these areas. For example, some institutions already safeguard their digital files. Others maintain metadata separately from their master files of electronic records. The implementation of more standards may satisfy the basic requirement for preservation: to maintain authentic and reliable records for as long as they are needed.

Yet as Gilliland-Swetland has pointed out,

> Counterintuitively, perhaps, it is during the preservation of digital materials that evidential value is often most at risk of being compromised. Digital preservation techniques have moved beyond a concern for the longevity of digital media to a concern for the preservation of the information stored in those media during recurrent migration to new software and hardware. In the process, many of the intrinsic characteristics of information objects can disappear—data structures can be modified and presentation of the object on a computer screen can be altered.[v]

Since the field of electronic record preservation is fledgling, with many concepts and issues still to be fully comprehended and resolved, we decided to conduct this study in three rounds, from 2000 through 2003. Round 1 (2000-2001) surveyed 13 institutions, programs, and projects. This round also provides the baseline for rounds 2 and 3. In round 2 (2001-2002) we are administering a revised version of the questionnaire, interviewing key informants, and conducting case studies for two of the respondents surveyed in round 1. The third round (2002-2003) will focus on an even smaller number of programs and projects. By the end of the three rounds, we expect to present a sharper picture of preservation strategies in practice than we currently have. These strategies include preservation techniques (e.g., refreshing, migration, emulation); selection for preservation; staffing configurations; cost modeling; access to preserved records; and policymaking.

Although other studies have explored individual aspects of electronic and digital preservation—such as standards, intellectual property, or specific techniques such as emulation—we are examining the continuum of activities that constitute the broad range and long-term goals of preservation of electronic records. By documenting the variety of approaches that are being taken, we hope not only to shed light on current practices, but by the end of round 3, to offer informed consideration on where preservation might be headed.

When we began our research, no study had yet provided a comparison of preservation techniques for electronic records across institutions and projects. The closest research to ours is a study that was carried out by Hedstrom and Montgomery (1998) to examine long-term retention needs and

requirements in the Research Libraries Group (RLG) member institutions.[vi] Another related work is a book edited by Kenney and Rieger on digital imaging for libraries and archives.[vii] One section of the book, "Digital Preservation Strategies," gives a succinct overview of techniques such as refreshing, migration, emulation, etc. We designed this study to examine specific long-term retention strategies in more detail.

We undertook this research as part of our work on the InterPARES Preservation Task Force in order to explore one of the five research questions that the Task Force was charged with examining: What methods, procedures, and rules of long-term preservation are in use or being developed? Our three-round study seeks to ascertain the strategies that are either currently in use or are in development for preserving electronic records. We also plan to track the evaluative studies of these techniques. (For a complete list of the research questions, see Appendix 5.)

Research Questions

1. 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?
    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?

In addition to question 1, formulated by the InterPARES Preservation Task Force, we asked three additional questions as a result of responses to the questionnaire:

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

3. Will current strategies for preserving electronic records ensure longevity and authenticity?

4. How are costs for the preservation of electronic records derived? Have effective cost models been developed?

Definitions

We begin this section with the two broadest terms: preservation and digital preservation. From there, we define digital preservation techniques. These relationships are hierarchical, with

preservation at the top, followed by digital preservation and digital preservation techniques. The preservation of electronic records falls under digital preservation.

Preservation

Digital Preservation

Electronic Records Preservation

[Digital Preservation Techniques]
Refreshing
Migration
Knowledge-Based Persistent Object Preservation
Emulation
Bundling
Universal Preservation Format

[Other Techniques]
Robotics (a capture, not preservation technique)
Digital Archaeology ("*post hoc* rescue")

Preservation: the processes and activities which stabilize and protect cultural heritage materials so that they will be permanent and durable, or as long-lasting as it is possible to make them (authors).

Digital Preservation: the processes and activities which stabilize and protect reformatted and "born digital" authentic electronic materials in forms which are retrievable, readable, and usable over time (authors).

Digital Preservation Techniques

Refreshing: periodically moving records from one storage medium to another. It is a preventive measure and, because of rapid media obsolescence, it will be a necessary strategy for some years to come (authors).

Migration: the process of moving records from one hardware and/or software platform to another (authors).

Collection-Based Persistent Object Preservation (aka, **Knowledge-Based Persistent Object Preservation):** " . . .the retrieval or instantiation of the collection onto new technology. . . . [t]he processes used to ingest a collection, transform it into an infrastructure independent form, and store the collection in an archive comprise the persistent storage steps of a persistent archive. The processes used to recreate the collection on new technology, optimize the database, and recreate the user interface comprise the retrieval steps of a persistent archive. The two phases form a cycle that can be used for migrating data collections onto new infrastructure as technology evolves" (Moore, et al., April 2000, http://www.dlib.org/dlib/april00/moore/04moore-pt.2).[viii]

Emulation: an applications software approach that recreates the technical environment required to view earlier programs. Such software can theoretically mimic every type of application ever written and be run on current computers (authors).[ix]

Technology Preservation: "preserving the technical environment that runs the system, including software and hardware such as operating systems, original application software, media drives, etc. While technology preservation means preserving the technical environment rather than reengineering it, as emulation does, many of the same issues apply to both" (Kenney/Rieger).

Bundling: "…taking objects such as Word documents and, by using software, creating bundles of documents on an independent platform. This is another form of emulation" (PRO interview; permission not yet received).

Universal Preservation Format (UPF): "Deals with multiple formats. It takes a snapshot of the information and gives you everything you need to view it. It is difficult to capture the whole content, e.g., interactivity, without the original equipment. Closer to emulation than migration" (WGBH interview; permission not yet received).

Digital Archaeology**: " …accessing digital materials where the media has become damaged (through disaster or age) or where the hardware or software is either no longer available or unknown. The authors refer to this as "*post hoc* rescue" (Ross/Gow).**[x]

Robotics: the use of robots to download electronic documents. Downloading in and of itself does not preserve the records (authors).

Terms

Record: a document made or received and set aside in the course of a practical activity (InterPARES Glossary).

Authentic record: a record that is what it purports to be and is free from tampering or corruption (InterPARES Glossary).

Digital record: a record that now exists in electronic form though it may or may not have been created in electronic form. For example, a *digital record* may have been created on paper and digitized later. Subsequent digitization may remove or deplete its "recordness" (authors, Gilliland-Swetland).

Electronic record: **A record that is created (made or received and set aside) in electronic form (InterPARES Glossary).**

Program: **an ongoing set of services, around a common goal or activity, usually located within a single institution (authors).**

Project: **a specific undertaking or research endeavor, usually with special funding. Projects may take place within single, institutional programs, or at more than one site. Projects usually take place within a finite period (authors).**

Research Methodology

To answer our overarching questions, we decided to use 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. We chose to collect data using a questionnaire which we developed with feedback from other members of the InterPARES Preservation Task Force.

We identified 15 sites that we knew were developing one or more of the following techniques: refreshing, migration, emulation, collection-based persistent object preservation, bundling, and robotics. We ultimately interviewed representatives from 13 of the 15 sites that we selected (Appendix 1, "List of Sites").

Such a small population did not warrant a quantitative research design. Further, since we knew that we would be asking broad, sometimes open-ended questions, we decided on a qualitative design. Since it was not our goal to make statistical inferences, but to learn about processes and methodologies, we explored one research method and one research strategy: the case study and the survey.

We ruled out the case study method for round 1 because we felt that it was too early in the development of long-term retention strategies to study individual programs in depth. (This is a methodology we plan to use in rounds 2 and 3.) Rather, using a questionnaire for round 1, we hoped to establish a benchmark of current practices from which we could collect a general or comprehensive view. The questionnaire would provide baseline data about the current (year 2000) state of long-term retention practices that we could draw on in rounds 2 and 3 of the research.

With respect to procedures followed during the data-collection process, the participants were sent a consent letter which explained that if they volunteered to participate in this study, they agreed to read over the attached survey instrument (See Appendix 2, Questionnaire) and participate in a telephone interview based on this instrument. We followed-up on this letter by telephone or e-mail to arrange a suitable time for the interview. The telephone interview was not taped. In a few instances we were able to conduct the interviews in person. To the best of our knowledge, there were no foreseeable risks, discomforts, or inconveniences to anyone who participated in this study. In terms of the information obtained from this study, we agreed not to disseminate proprietary information, quote any of the interviewees, or disclose individual or institutional identities without express written permission.

Volunteers had the option to withdraw from the study at any time without consequences of any kind. They could also refuse to answer any questions they did not want to answer and still remain in the study. As investigators, we also maintained the option of withdrawing participants from the research if circumstances arose which warranted doing so.

Boundaries and Limitations of the Study

The data gathered and analyzed in round 1 of this research allow us to draw only tentative conclusions about current preservation techniques, and in a limited number of venues (6 archives, 6 programs and research projects, and 1 library). Although our aim was to identify as many different preservation techniques as possible—without regard to how many institutions and projects were experimenting with new techniques—there is the possibility that we missed learning about important new projects. Also, preservation techniques such as bundling and Universal Preservation Format (UPF) are still only in the earliest stages of development, so we did not learn as much about them as we had hoped. However, the study has the advantage of identifying and describing cutting-edge approaches to the preservation of electronic records and the researchers will continue to monitor new developments.

Another limitation is that InterPARES sponsored this research, and many of the participants in our study were affiliated with the project. Therefore, the needs and perspectives of the InterPARES Preservation Task Force drove the study. Since InterPARES focuses on electronic records, we focused on archives and archival projects. We selected one library, but the practices of that library demonstrated it to be a disconfirming case and we will not include it in round 2.

On the other hand, InterPARES is an international project with great visibility. Our affiliation with the project may have resulted in entrées to research projects that we might not have otherwise been able to gain access to.

The responses from our participants were probably not as comprehensive as they might have been had we taped them. We chose not to tape the interviewees thinking that it would cause the participants, as representatives of their institutions, programs, or projects, to be constrained in their responses. During the interviews we both took notes. After the interviews were completed, and upon further review of our notes and procedures, we felt that our notes represented more of an interpretive record of the interview than a record of quotations. The participants did not appear to be constrained in their responses, which led us to consider the possibility of taping subsequent interviews. We informally surveyed some of the participants and learned that they do not object to having their interviews taped for the second round of the survey. We feel that using the telephone during the interviews also introduced a barrier between the respondents and us. This was especially apparent in interviews where respondents spoke English as a second or third language. In the next round of the preservation survey, we have decided to ask for electronic responses to the survey instrument, and to tape the follow-up telephone interviews. When possible, we will conduct in-person interviews.

A final potential limitation of this study relates to concepts and terminology, a limitation pointed out by Hedstrom and Montgomery in their RLG study.[xi] We also surveyed practitioners and

researchers in both the United States and abroad. Like their study, we found that differences in terminology and concepts may reflect different cultural perspectives, or, since our interviewees included archivists, librarians, computer scientists, and engineers, professional differences.

Results

The questionnaire was divided into 14 sections:

A. Information about the Institution/Project and Respondent
B. Program and Policy
C. Specifics of Preservation Technique/Method/Strategy
D. Selection for Preservation
E. Cooperation
F. Staffing
G. Technical Questions
H. Costs
I. Preserving Records
J. Description/Documentation of Preservation Processes
K. Access to Preserved Records
L. Charges
M. Reproduction and Copyright
N. Preservation Policies

Not all sections or questions pertained to each project or program. However, we chose to be comprehensive in order to learn about as many aspects of each program as possible. We will revise the questionnaire in the 2001 iteration of the survey. In this iteration we tried to achieve breadth and scope; in the next version we will focus on some of the above sections in greater depth. (See Appendix 3, Tabulated Responses.)

> We have followed the procedures for Human Subjects research established by the University of California. Respondents were guaranteed anonymity unless they provided express written consent. Although we have requested such consents, at the time of this writing we have not yet heard from all of the respondents. All quotations used here will be designated as respondent 1, 2, etc., and references to projects as institution 1, 2, etc.

Section A: Information about the Institution/Project and Respondent

This section is summarized in Appendix 1.

Section B: Program and Policy

The questions were:

3. Please describe your institution's program or activities related to preserving digital objects over long periods of time.

3.1. When did your institution's program or activities begin?

3.2. Describe any institutional issues which impact upon the program.

3.2.1. Describe in broad terms what methods or techniques you are exploring or using for digital preservation.

3.3. Describe the digital materials your institution is preserving.

3.3.1. Do you consider any of these materials to be records?

3.4. Do you make any special provisions for preserving records, as opposed to other types of digital materials? If so, what?

3.5. Has the program or activity reached the point of either testing or evaluating any of the methods or techniques you are using? If so, what are the results to date?

3.6. Have you identified any problems, difficulties or threats to the integrity of the digital materials resulting from the use of these methods or techniques? If so, please describe.

3.7. How do you use the word "preservation" at your institution? In other words, what definition does your institution associate with the term "preservation"?

In this section we tried to ascertain the range of preservation activities for electronic records. We started by asking each respondent to give an overview of his or her program. When asked, "When did your institution's program or activities begin?" respondents traced electronic preservation programs back to the inception of *any* preservation activity in their institutions. Four respondents reported that their programs started in the 1970s, 1 in the 1980s, and 5 in the 1990s. Of the 4 who reported activity since the 1970s, 2 archives reported that they had been preserving digital objects since the 1970s. Three responded "not applicable" because they represented special projects. Each respondent then described the development of preservation programs over time.

Respondents were asked to describe institutional issues that impact their programs. Issues included: the problems associated with outsourcing; inadequate staffing; storage; prioritization of records to be preserved; inadequate resources, including funding; and increasing legal mandates for preservation.

We also asked what methods or techniques they were exploring or using for digital preservation. The response: Migration=4; Emulation=2; Knowledge-Based Persistent Object Preservation=3; Bundling=1; Refreshing=1; Digital Archaeology=1; Preservation copying=1; Physical Preservation=2; and Robotics=1. It is important to note that some institutions and projects are using more than one strategy, and that Bundling is in the exploratory stages only.

The next question asked respondents to describe the items that they are preserving. Respondent 1 reported preserving only born-electronic records. Others preserved a variety of born-electronic records and other digital artifacts including spreadsheets, databases, computer games, geomatics, serials, etc.

One question in this section caused confusion: "Do you consider any of these materials to be records?" Six answered "yes," one answered "no" and 6 answered "not applicable." Not all of the respondents distinguished between records, documents, and items. Our follow-up question, "Do you make any special provisions for preserving records, as opposed to other types of digital materials? Only respondent 11 said "yes." Five said "no" and 7 responded "not applicable," perhaps due to the same confusion demonstrated in the preceding question.

We also asked if the program or project was testing or evaluating any of the methods or techniques currently being used. Only three respondents are engaged in testing or evaluation.

The next question asked respondents to identify problems, difficulties, or threats to the integrity of digital materials with any of the above-mentioned techniques. Three respondents identified preserving the integrity of the original digital object as a threat. Respondent 4 discussed the issue of "acceptable loss" due to migration. Two other respondents identified the problem of changing standards. Finally, respondent 12 mentioned the issue of migration technology obsolescence.

From this section of the questionnaire, it became apparent that the thirteen institutions and projects were better versed in identifying problems than in developing solutions.

The final question asked respondents to define preservation. Their responses are described in detail below.

Section C: Specifics of Preservation Technique/Method/Strategy

4. What preservation technique does your program use?

4.1. How was this method selected?

4.1.1 Is it a hybrid, e.g. a combination of two or more preservation techniques/methods, such as microfilming and scanning? ❑ YES ❑ NO

4.2. If you are using a hybrid model, how did it evolve?

4.3. From other methods you have tried before or you are aware of that other repositories are using, how is this method different from other methods?

4.4. In selecting the preservation method or strategy, have you considered what its effect might be upon the intellectual integrity (e.g., authenticity and reliability) of the digital material? ❑ YES ❑ NO

4.4.1. If yes, are you able to prove/demonstrate that the intellectual integrity of the digital material has not been compromised through the preservation process? Please explain.

4.5. Is there evaluative data on the efficacy of this preservation method/model? Please describe.

This section asked respondents to elaborate on the preservation techniques that they identified in Section B. As reported above, the most common strategy is migration. We asked respondents in this section to tell us how they selected the method. Respondent 1 identified migration as the standard method of moving from one platform to another. Respondent 7 responded that her institution believes it to be the best method.

As a follow up question, we asked if any of the institutions or projects used a hybrid approach that combined two or more techniques. Four respondents said "yes," four said "not applicable," one said "maybe," and three said "no." Of the four yes answers, three indicated microfilming and scanning and one said scanning and physical treatment. No one reported combining such techniques as robotics and migration, for example.

We also asked whether the respondents had considered what effect their chosen techniques might have on the intellectual integrity of the digital materials. Nine respondents indicated "yes;" four, "no." Respondent 7 criticized the work of another institution that enhanced original photos, citing issues of veracity. Respondent 8 worried that emulation may not work for some classes of materials.

The final question in this section asked whether there is evaluative data on the efficacy of the preservation method/model used by each respondent. Only two respondents had formal evaluation procedures. Respondent 1 reported that audit trails are kept, and that his institution is moving to a document management program (ERM) to support audit trail development. Respondent 11 uses Knowledge-Based Persistent Object Preservation which conforms to the OAIS model.

Section D: Selection for Preservation

5. Which of these criteria guide selection of materials for preservation? (Check all relevant)

- ❑ Historical/cultural value
- ❑ Legal requirement to preserve
- ❑ Retard deterioration
- ❑ Increase access
- ❑ Save space
- ❑ Research into preservation processes
- ❑ Commercial use
- ❑ Other reasons (please specify)

5.1. Where did the materials you selected come from?

- ❑ Parent institution
- ❑ Collaborating institutions
- ❑ Government agencies
- ❑ Other organizations or associations
- ❑ Other (specify)

There were 13 responses; of these there was one 'not applicable.' *Historical/cultural value* and *Legal requirement to preserve* were the two criteria guiding selection of materials for

preservation for 9 of the respondents. The next two criteria were *research into preservation* and *retarding deterioration* (5). *Saving space* and *Other* ranked third. No one cited commercial reasons as criteria for selection. In the *Other* category, criteria includes: materials publishers wanted them to have; institutional requirements; asset management considerations; sampling Internet sites; and supporting the curriculum. The 'not applicable" response is from a respondent which preserves materials for clients.

The materials which were selected to be preserved, came from: government agencies (6); parent institutions (5); other, e.g. private, commercial publishers, politician's private papers, private individuals, commercial entities (3); and other organizations or associations, e.g. corporations (private sector) (2).

Section E: Cooperation

6. Did you cooperate with other organizations to develop your program?
❑ YES ❑ NO

6.1. If so, which? (Check all relevant)

❑ Archives ❑ Libraries ❑ Public companies ❑ Museums

6.2. Is your cooperation ❑ National ❑ International ❑ Local
❑ Shared facilities ❑ By institutional type
[churches, labor unions, etc.]

6.3. How is the work distributed?

❑ Equally ❑ Work distributed in a different way?

6.3.1. Please describe.

6.4. If your program is collaborative, how did it evolve?

6.4.1. Please describe the strengths and weaknesses of the collaboration.

Of the 13 respondents, 12 cooperate with other organizations to develop their preservation program, while one does not. The types of collaborating organizations include archives (9), libraries (9), public companies (5), museums (3); and in the 'other' category, government agencies, other programs and projects, for profit and non-profit educational institutions; and universities. Cooperation occurred on an international level (11), national level (9), local (2) and with shared facilities (1). Work was distributed both equally (5); and in a different way (6).

Collaborations tended to evolve rather than to be planned. Participants viewed the strengths of collaboration as being shared responsibilities and costs; shared information and resources; and opportunities to develop a consensual outlook. Weaknesses of the collaborative process include resource costs, people being far apart and influenced by local interests and resources; span of control issues; and research and development going off on tangents which were non-productive.

Section F: Staffing

7. Who is involved with the program and in what capacity (ies)?

7.1. Describe their duties.

The majority of staff is part time, work under the supervision of a person who usually holds at least one Master's degree; and divide their time with other projects and departments. Ph.D.s accounted for only two of the positions, and both worked on research projects. Educational backgrounds and skills include computer scientists, archivists, librarians, people with management and history backgrounds, and people who learned their technical skills on the job. One of the Ph.D.s is in History and Preservation; that person directs a facility.

Section G: Technical Questions

8. Is preservation carried out by ❑ the institution (in-house)
❑ commercial vendor/contractor

8.1. Describe any pre-preservation preparation of records.

8.2. What do you consider to be the strong points of your institution's preservation methods or techniques?

8.2.1. What do you consider to be the weak points of your institution's preservation methods or techniques?

8.3. What quality control methods are applied to the preservation process or activity?

8.4. How are you storing the electronic records that have been preserved?

Ten respondents carry out preservation in-house, while (1) uses a commercial vendor. One respondent uses both in-house and commercial resources, while another respondent has disks on a shelf and is not taking pro-active steps to preserve them at this time.

Pre-preservation preparation of records includes documenting provenance; checking to make sure that records have not been tampered with; inspecting them; physical preparation as needed; putting the records into a standard format; accessioning; and arranging and describing the records prior to copying them.

The questionnaire asked respondents to discuss weak points in their institution's preservation methods or techniques. These include database problems; and some respondents working with GIS and CAD materials, found that tabular displays don't work. For electronic documents, respondents cited compound records, website material, attachments and nested materials, as being problems. Implementing finding aids for a million collections is a problem in one archive. Another archivist said that his archive has not yet done the right thing with respect to textual

documents. According to him, there are not enough resources, e.g., money and people; and the current staff does not have sufficient expertise for the work. Lastly, an archivist of a large national archive stated that the archival profession has not yet articulated its needs regarding system requirements; that the profession "is not used to system thinking in work, and never had an opportunity to do it before." During round 2, we will explore the scope of these issues in more detail to determine whether they are ongoing, and whether resolutions have been investigated or applied.

Quality control methods applied to the preservation process or activity include the use of a quality control manual for a large European archive, with more information in development. One respondent stated that his facility was currently using *ad hoc* methods not worthy of sharing. Another respondent cited quality control of metadata as being the most difficult technical challenge. One respondent stated that quality control methods would be a partial outcome of their project. An archivist in a European archive stated that there was no really organized quality control; they rely on the professionalism of the individual, who often has no formal training. Other responses include: the use of standard information technology techniques for verifying the quality or success of the copying; and peer review research to promote publishing and discussion.

Permanent storage of electronic records includes archival storage at partner sites; containers; server/redundant servers; and underground storage. One site is not storing backup copies of records offsite.

Section H: Costs

9. What do you estimate are the costs to preserve the records? (Please include staff, equipment, space, energy and other related costs)

9.1. If applicable, please describe the categories of your preservation costs.

9.2. What are the sources of funding for the program, and how are they allocated?

See below for discussion.

Section I: Preserving Records

10. How are preserved records organized?

10.1. How is provenance respected?

10.2. Are there any restrictions on access to the records? If so, how are they enforced?

The majority of respondents, who answered the questions, stated that in their institutions, they organized records according to the standards currently in use, e.g., the same as for paper records; or according to the archival arrangement and description standards adopted by their institution.

Five respondents stated that the question did not apply, while one institution responded that an organizational scheme would be an outcome of their project.

Regarding respecting provenance, four respondents stated that the question does not apply, while the balance of the respondents cited the development of metadata standards, describing records according to provenance and type of record, and conformance with the institution's general practice as examples.

Six respondents stated that the question did not apply. However, all of the remaining respondents cited examples of the restrictions that their institution imposed on records. Respondent one stated that there was a 30 year closure [from general public access] unless the record was in the public domain; and that the record can in fact, be closed for up to 100 years. Respondent four also cited the 30 year closure, with the addition that the records were still subject to subpoena. Other respondents stated that the access to records is negotiated individually with each donor, while respondent eleven explained that the type of restriction on the record would determine accessibility.

Section J: Description/Documentation of Preservation Processes

11. Describe record-keeping for the preserved material.

11.1. Are preserved materials described according to a recognized standard? ❑ YES ❑ NO

11.1.1. If yes, which one?

11.2. Are the records for preserved materials and the original ❑ the same
❑ independent of each other

11.3. How is metadata used to describe preserved materials?

Respondents generally described record keeping for the preserved material as detailed audit trails, catalogs, on databases, as part of corporate descriptive systems, or as part of the metadata which accompanied the preserved objects. The question did not apply to six respondents, because they are not yet using metadata. We expect to see more widespread use of preservation metadata in round 3. (RLG and OCLC collaborate on the international Preservation Metadata Working Group. The group is trying to establish approaches for preservation metadata that will work in a variety of settings for a variety of materials.)

Nine institutions use recognized standards to describe preserved materials. These standards include ISAD-G, EAD, MARC, modified LCSH and Dublin Core.

We asked whether the records for preserved materials and the original were the same or independent of each other and most of our respondents did not understand the question. Eight respondents said the question did not apply, while two respondents stated that the record(s) were both the same and independent. Two respondents stated that the record was the same, while

respondent ten stated that the institution does not recognize that preserved materials are copies; both copies are the records.

According to the respondents, five institutions are addressing metadata use. The question did not apply to four other institutions. Respondent six stated that the use of metadata is extensive, e.g., content description, preservation history, but that it was not always completed. We take that to mean that the metadata record was not always produced in its entirety, according to institutional standards. Respondent ten stated that metadata is being used in a variety of ways.

Section K: Access to Preserved Records

12. Are the preserved records available ❑ only on-site ❑ only within the institution ❑ through a website ❑ other (specify)

12.1. If available through a website, please give the URL

_____

12.2. Is the archival workstation equipped with ❑ access control mechanisms ❑ billing software

12.3. Is the archival workstation connected to ❑ internal institutional servers ❑ an Intranet ❑ the Internet

The majority of institutions make records available through a website (7). One institution makes hard copy available, while one institution makes copyright protected material available to the academic community only. We did not ask whether the records were full text, or all or some series. The question did not apply to four institutions.

Question 12.2. did not apply to twelve institutions, while the remaining respondent stated that the institution had dedicated machines which consisted of five workstations, some online, and some connected to secure servers.

The last question in this section did not apply to any of the respondents.

Section L: Charges

13. Do users have to pay to use the preserved material? ❑ YES ❑ NO

13.1. If yes, ❑ on-site ❑ outside the Institution ❑ when accessed through the website ❑ other (specify)

13.2. If charges are made, how are these calculated? ❑ single charge ❑ by time ❑ by volume of material ❑ by intended use (commercial/academic/students) ❑ customized service (please describe)

❑ other (give details)

13.3. If charges are made, how are they collected? ❑ invoice ❑ cash at point of use ❑ credit card ❑ electronic accounting ❑ other (give details)

Seven institutions do not charge users to use preserved materials. Four institutions charge a fee for copies, two institutions charge users a fee to use preserved materials— when the material is accessed through the website. Charges are calculated by time (1), by volume of material (2), by intended use (1), e.g., commercial, academic, students, or as an institutional charge (1). One institution collects charges by credit card; three send invoices; and one is in the process of determining the procedure. One institution collects charges both by credit card and by invoice. Several respondents did not know how the charges were collected in their institutions.

Section M: Reproduction and Copyright

14. Do you preserve material in copyright? ❑ YES ❑ NO

14.1. If yes, is this done ❑ under legal provisions for your institution ❑ with the owner's agreement ❑ by paying the owner a fee ❑ under license ❑ without formalities

14.2. Does the institution own the copyright for the electronic form of the records? ❑ YES ❑ NO

14.2.1. If no, who does?

14.3. Are users allowed to do any of the following?
❑ make printouts ❑ download to a PC ❑ download to a local network (LAN) ❑ download to a general network (WAN)

14.4. Are any electronic management systems used to control copying? ❑ YES ❑ NO

14.4.1. If yes, which ones?

Ten institutions preserve material in copyright, while one institution does not. The question did not apply to two institutions. Of the institutions which preserve material in copyright, (8) stated that this was done under legal provisions for their institution, with the owner's agreement (10) and without formalities (1).

Six institutions do not own the copyright for the electronic form of the records, while the question did not apply to seven respondents. Ownership of the copyright resides in the creator, owner, or government, depending on the institution.

Users are permitted to download material to a PC (2); download to a local network (LAN) (2); or download to a general network (WAN) (2). One respondent said that the institution does not permit users to download material. The question did not apply to nine respondents.

With respect to electronic management systems in use to control copying, one respondent stated that that was being researched, while four respondents stated that their institutions do not have electronic management systems of that type in use. The question did not apply to eight respondents.

Section N: Preservation Policies

15. Do you have a general preservation policy that includes records in electronic form?
❑ YES ❑ NO

15.1. If not, do you have a policy for reformatting, refreshing, migrating, emulating, or bundling data to newer technological platforms? ❑ YES ❑ NO

15.2. Please describe any policies you might have that relate to preservation of electronic records.

Five respondents do not have a general preservation policy in place that includes electronic records, while three respondents stated that their institutions did have such a policy in place. Two respondents indicated that their institutions had such a policy in development or being researched, while three respondents said that the question did not apply.

Regarding question 15.1., one institution has a policy in place that includes various types of preservation techniques, while four institutions do not. Respondent 10 stated that the institution has no formal policy, but that the established practice is to re-format to standard non-proprietary formats. Two respondents stated that their institutions did not have policies finalized as yet.

See below for further discussion.

Three questions in particular yielded stimulating and--we believe--pertinent observations: definitions of preservation (in the *Program and Policy* section), categories of costs (*Costs*), and preservation policies (*Preservation Policies*).

Definitions of Preservation

In the section on *Program and Policy*, we asked: "How do you use the word 'preservation' at your institution?" In other words, what definition does your institution associate with the term 'preservation'?" We included this question because we felt that such definitions might have a bearing on approaches that projects take to developing long-term retention strategies. For example, depending on your view of preservation you might select one approach over another. Since we interviewed archivists and librarians from six countries representing a dozen projects or

institutions, we anticipated getting a range of responses. As we continue this study into the next phase of this research, we will try to determine whether or not the definitions of preservation continue to evolve.

Eleven respondents defined 'preservation'[xii]: The following key phrases emerged:

Respondent 1: "Preservation for paper records is a regime which tries to slow entropy and avoid degradation. For digital records, it is to preserve the document to perpetuity. Digital Preservation includes issues of authenticity."

Respondent 2: "Preservation means ensuring the object is accessible over the long-term. Access and preservation are separate."

Respondent 4: "Preservation covers all activities directed towards ensuring the ongoing accessibility to the information content of the records. Hence, we consider the ambient conditions in our repositories as a preservation issue, along with the specifications of the media on which recorded information is stored. Migration of digital objects is thus a preservation strategy."

Respondent 5: "The ability to discover, access, and present electronic records through arbitrary changes of technology. We can preserve things forever."

Respondent 7: "Forward migration or prospective preservation to whatever new technologies exist. [We are beyond] thinking about 'x' number of years of preservation."

Respondent 8: "Enabling long-term access to materials."

Respondent 9: "Ability to present the record unchanged repeatedly."

Respondent 11: "Everything you have to do to guarantee you can deliver records [and] respecting the sanctity of the original order."

Respondent 12: [The technical and managerial processes that protect the integrity and longevity of materials – regardless of genre.]

Respondent 13: "Making collections useful to scholars in the future."

The respondents' key phrases fall into three components of preservation: 1. preservation processes; 2. length of time for retention; and, 3. preservation outcomes.

Overall, the responses demonstrate a shift taking place from defining preservation as a once-and-forever approach for paper-based materials, to an all-the-time approach for digital materials. (Paper-based materials also require all-the-time care, mainly through environmental controls and proper storage. However, digital materials require constant refreshing, migrating, etc., a much more pro-active and costly endeavor.)

To contextualize the respondents' definitions, we offer published definitions of preservation culled from the archives and the library fields.

The Paper-Based Perspective

SAA, 1974
A. The basic responsibility to provide adequate facilities for the protection, care, and maintenance of archives, records, and manuscripts. B. Specific measures, individual and collective, undertaken for the repair, maintenance, restoration, or protection of documents.[xiii]

Ratcliffe Report, 1984
Strictly, all the steps taken to protect materials, that is including conservation and restoration, but often used in reference to the treatment of materials on first entering the library; it is preventive rather than remedial.[xiv]

Transition to Digital

IFLA, 1986
Includes all managerial and financial considerations including storage and accommodation provisions, staffing levels, policies, techniques and methods involved in preserving library and archive materials and the information contained in them.[xv]

Feather, Matthews, and Eden, 1996
The managerial, financial and technical issues involved in preserving library materials in all formats--and/or their information content--so as to maximize life.[xvi]

Digital

SAA, 1997
Preservation of digital information is not so much about protecting physical objects as about specifying the creation and maintenance of intangible electronic files whose intellectual integrity is their primary characteristic. Preservation goes beyond saving such media as optical disks or magnetic tape; the access system itself must be preserved.[xvii]

ICA, 1997

An electronic record is preserved if and only if it continues to exist in a form that allows it to be retrieved, and, once retrieved, provides reliable and authentic evidence of the activity which produced the record.[xviii]

Kenney and Rieger, 2000
Digital Preservation means retaining digital image collections in a usable and interpretable form for the long term. While "long-term" suggests an indefinite future, David Bearman interprets it more usefully as 'retention for a period of continuing value.'[xix]

Archival and library definitions have shifted from the physical care and protection of materials to retaining them in retrievable form for an indefinite amount of time. In the paper-based information world, librarians and archivists sought to preserve books and documents for 500 years or more. As is apparent from both the study respondents, and the professional literature, professionals now think about maximizing "useful life" or preserving digital documents "forever" through emulation or forward migration, but without the emphasis on a specific number of years.

Further analysis of our data indicates that archivists and librarians view preservation through different lenses. This reflects a fundamental difference in the archival and library professions. Librarians tend to be custodians of printed materials that are not unique. Librarianship carries custodial responsibilities, but--with the exception of special collections--missing or damaged items can usually be replaced. Therefore, librarians often view their materials in terms of immediate utility. In the archival arena, when a record is gone, it is really gone and cannot be replaced--whether that is due to an accident or a disposal schedule. Archivists have responsibility for one-of-a-kind records, which are lodged in a repository. In current practice, the repository and the object cannot be divorced. This relationship differs from libraries and printed materials. In archives, long-term accessibility to the records may be mandated by legal warrant and business processes, and more broadly, by societal memory. The impact of electronic records may have an effect on the requirements that the repository and the object remain together in archives. In the digital environment *both* librarians and archivists have responsibility for documents and records that are born digital. These digital assets are susceptible to obsolescence and incompatibility.[xx] Therefore, the integrity and authenticity of digital objects is of mutual concern to both professions. As librarians and archivists work closely on long-term retention strategies, the definition of preservation may shift to accommodate both professional perspectives.

Costs

In the section on Costs, we asked: "What do you estimate are the costs to preserve the records?" Responses included staff, equipment, space, energy and other related costs.

In essence, we were asking, what is it going to cost the institution to preserve, maintain, and provide access to electronic records? We thought this was an important question because for many institutions and projects, knowing what the bottom line is, is THE major factor which

influences decision-making, and determines goals and objectives, as well as the strategies to meet them. Knowing about costs helps institutions lobby with parent institutions or funding agencies. What is the role of costs in situations where because of legal requirements, archives do not have a choice about what they preserve or even how they preserve records? The majority of the managers we interviewed are gathering financial data now and plan to report costs as part of their projects' results. Only a few projects are far enough along to have developed cost figures. The interviewees ranged from large national archives, to projects developing testbeds. The costs for electronic record preservation ranged from $10,000 to $2.6 million per year. Cost categories include staff, consultants, facilities, equipment, storage system monitoring, staff access and research and development.

Most of the projects are currently funded through initial allocations, and some of these figures reflect the impact of early research and development costs, which could also account for the wide range of costs. In fact, as one respondent said, the costs for his project might be reduced by as much as half during the following year. This question will be followed up as part of the second phase of the research interviews. It will be interesting to see what the forecast figures for preservation, storage and staffing actually turn out to be, especially when the initial costs of research and development are reduced over time.

At the time of these interviews, none of the respondents had yet gathered enough information to determine the categories of preservation costs or cost modeling protocols.

Sources of funding include various government agencies, EU (European Union), NSF (National Science Foundation), NPACI (National Partnership for Advanced Computational Infrastructure), NEH (National Endowment for the Humanities), NHPRC (National Historical Publications and Records Commission), NARA (National Archives and Records Administration), and JISC (Joint Information System Committee). As always, the question remains as to what extent the source(s) of funding have shaped the research agenda and from there, the future.

The follow-up study will gather data on the further development of a preservation cost model. So far, cost modeling for digital projects has received scant attention. The present focus appears to be on budgeting for digital conversions rather than preserving authentic electronic records. In addition, there is scant literature in the area of cost models for born digital electronic records. Two exceptions are studies by Hendley and by Russell and Weinberger. Hendley, in his report on the *Comparison of Methods & Costs of Digital Preservation*[xxi], provides a "Table of Digital Preservation Cost Elements" which was compiled by Neil Beagrie, Daniel Greenstein, and the Arts and Humanities Data Service. In it, the cost elements involved in developing and preserving digital collections are keyed to the life cycle stages of a digital resource.[xxii] In their article, *Cost Elements of Digital Preservation*, Kelly Russell and Ellis Weinberger posit that the ongoing costs of digital preservation span a more extended timeframe than traditional preservation and will therefore require resource commitments of a different nature. Different strategies may necessitate different costing time frames and schedules. Russell and Weinberger state that current cost models have yet to reflect this more complex environment. They further state that, "The creation of a digital object is the true starting point for digital preservation."[xxiii]

To estimate a budget for image acquisition, Anne Kenney and Oya Rieger refer to the "RLG Worksheet for Estimating Digital Reformatting Costs" in their book, *Moving Theory into Practice: Digital Imaging for Libraries and Archives.*[xxiv] The Worksheet, in combination with an assessment of costs derived by Cornell's Department of Preservation, identified costs for image acquisition in six cost categories. These costs include personnel, equipment, cataloging, supplies, contingency and overhead/indirects.

To facilitate the development of a preservation cost model, a number of categories may be adapted from traditional cost models. These categories might include: costs for providing access to the materials; costs related to long-term creation and maintenance of digital materials, production of metadata, personnel, equipment, cataloging, supplies, contingency (e.g. emergency/unforeseen events), overhead, administration and, research and development.

One respondent provided information about plans to form a consortium of institutions to form a National Preservation Center. This idea should be explored not only because of its potential for cost-effectiveness of preservation, but also for the opportunities to enrich the library, archival and museum professions, which may occur as a result of providing a forum for communication across institutional settings and domains.

In a speech for directors of the Association of Research Libraries, Clifford Lynch stated,

> The fundamentally hard things about managing bits into the future mostly aren't technical; they're economic and organizational. Bits need care and feeding. They don't do well with benign neglect. This means that we need to come up with financial models to keep these bits cared for and healthy as they are migrated into the future. We don't lose a lot of bits to technical failures in a well-managed environment, but we lose a lot due to financial or organizational failures to maintain that well-managed, caring environment on a continual basis.[xxv]

We include this quote to emphasize that technical processes cannot be separated from economic issues. The library and archival professions have not fully grappled with the economic influences on preservation decisions. It is necessary to identify concepts and approaches for evaluating the full economic impact of long-term retention. Institutional, national, and multi-national policies must be put in place to assure preservation in perpetuity.

Preservation Policy

We concluded the survey with the following three questions about policy:

1. Do you have a general preservation policy that includes records in electronic form?

2. If not, do you have a policy for reformatting, refreshing, migrating, emulating, or bundling data to newer technological platforms?

3. Please describe any policies you might have that relate to preservation of electronic records.

Only three of the projects/institutions indicated having policies in place; two others are revising existing policies to include electronic records; and one is currently developing a policy that includes multiple media. Two of the research projects indicated that policy development would be an outcome of their research.

During Round 2 of the preservation survey, we will be interviewing key informants/experts who may shed more light on policy issues. We will try to ascertain whether or not international concern about the longevity of digital information is being followed up in policymaking arenas. We suspect that policy is lagging far behind the development of standards, because the development of good public policy requires the appropriate political climate as well as the cooperation of numerous stakeholders. Further, there must be a legal environment that enables the preservation of digital information. Yet laws may vary. For example, the Berne Convention and US copyright law have significant differences between them. These types of discrepancies may impede the development of consistent, rational public policy.

Conclusion

At present, the interviews indicate three broad themes. First, that the perception of what preservation is goes beyond library and archival practice to the media being preserved. Because electronic material is inherently ephemeral, and the timeframe involved to preserve and provide access to this material extends to perpetuity, we expect that traditional definitions of preservation may not apply. Indeed, 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. One respondent, who is fully funded by his government, put it succinctly when he said, "We haven't yet been asked to measure costs! We don't need to justify costs. Fixed costs are unknown." Another respondent said, "The result will be cost determinations." And a third answered that costs "should be a result of the current testbed project [that they are engaged in]." We believe that the problems posed by preserving authentic electronic records permanently (or as long as possible), requires the development of a cost model, which will be unique and not a hybrid of existing digital conversion cost models. We agree with Hendley, Russell, and Weinberger that preservation begins at the creation of the electronic material. A cost model for preserving authentic electronic records will need to reflect this perspective, which differs from the traditional preservation point of view. Costs, however, cannot single-handedly be used to justify *not* preserving otherwise valuable records.

We found that staff and equipment costs are the most consistent hard figures available so far, and of course, those will vary over time, which will ultimately connect with developing forecasting strategies. Many of the projects are nascent, and we suspect that for them, answering the survey questions was essentially a theoretical exercise. As the institutions and projects progress, we expect to be able to gather hard data during rounds 2 and 3 of the survey. By the conclusion of round 3, we expect to have a substantial amount of information about institutions and projects that will have been active for at least 3 years. From this, we hope to develop a cost model for preserving authentic electronic records, which can be applied to archives and libraries, and perhaps to other communities of practice.

Last, the lack of preservation policies in place is a distinct gap in the research design of many of the projects and possibly reflects a lack of commitment among the stakeholders in institutions. What is the reasoning behind developing policy as an end result of a project, instead of concurrently with its progress? We suspect that meeting the technological challenges of preserving electronic records is more of a priority within these institutions than developing policy and wonder whether, as a result, the overall progress in this new arena will be more uneven than is necessary. Several institutions that responded to our survey have had active programs for a long time and we note that often policy evolved, rather than being strategically planned. It is practically impossible to set policy 100% at the outset of a project—especially one in such a complex area as the preservation of electronic records. Policy will naturally evolve rather rapidly at the outset of a program when the practitioners encounter new, possibly unanticipated features of the program that require policy decisions. As the program matures, and even while it is still developing, policy will concomitantly need to be re-thought or newly conceived. In fact, policy must also drive technological development. When the program is in "full swing," policy will have reached a point at which it is now well thought out, though still subject to modification, as the program requires.

In the subsequent phases of our survey, we hope to explore not only the "why" behind the positioning of policy development within the institution, but also the development of its content. We want to explore the role of the stakeholders and the influence of the legal and political environments which provide the context in which policy is formed.

We note that one project we included in round 1, has discontinued its research because funding ran out. This particular project was unique in that it was exploring the preservation of multimedia material. The gap in potential knowledge that could have been disseminated is a loss. But it is also a reminder that the adoption of any new technology depends on politics, funding, and timing.

Round 2 of the preservation survey will focus on expanding our knowledge in several areas. These include staffing and personnel – where are future specialists in preserving electronic records going to come from? Another expanded area in the survey is cost activities. Because of the nascence of some aspects of the programs we studied, such as charges, access, reproduction and copyright, we were able to gather very little substantive information. As a result, these sections will probably drop out of round 2. However, we will re-evaluate their inclusion for

round 3. As well, we intend to explore in more depth, why certain questions did not apply to some respondents.

No matter which preservation method is chosen, cost will become a factor in making a management decision regarding preservation of electronic records. We have also expanded the area of our survey that asks for information on description/documentation of preservation processes, as well as the section on preservation policy. In addition, we have developed a second survey instrument that we will use to interview key informants/experts, whom we define as "individuals who provide useful insights in the fields of preservation and/or archives." These experts may work in a variety of settings including – but not limited to – universities, government archives, foundations/granting agencies, industry, professional organizations, or think tanks, or who serve as consultants. Most have extensive national and international experience that enables them to provide the long view of preservation as well as placing them at the forefront of their professions.

* * * * * * * * *

We began our study by asking four research questions:

1. 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?
    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?

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

3. Will current strategies for preserving electronic records ensure longevity and authenticity?

4. How are costs for the preservation of electronic records derived? Have effective cost models been developed?

Our survey provided us with only partial answers to Questions 1 and 3. We identified a number of preservation techniques that are currently in use—including migration, emulation, and robotics—but not one of these techniques could be considered to meet the conceptual requirements for authenticity. Until these methods are further developed and standardized, they cannot be relied upon to ensure the long-term preservation of electronic records. As for question 1b, "Which methods of long-term preservation need to be developed?," they all do. It is still too early in the development of all these techniques to fully evaluate them. Regarding question 1c, "Which of these methods are required or subject to standards, regulations, and guidelines in specific industry or institutional settings?," there is no simple answer. The projects represented in our survey are developing standards and guidelines. Some of the institutions we interviewed

are waiting to see the results of these projects before committing to a particular strategy. We hope to be able to answer question 1 more fully in subsequent rounds of this research.

Question 2 yielded richer results. It is clear that professionals are revising their definitions of preservation from a once-and-forever approach for paper-based materials to an all-the-time approach for digital materials. Preservation must now accommodate both media and access systems. Finally, while we once tended to think about preserving materials for a particular period of time—for example, permanent/durable paper was expected to last for five hundred years—we now think about retaining digital media for a period of continuing value.

Meaningful answers to Question 4, regarding costs for the preservation of electronic records, must also wait until rounds 2 and 3. Our survey revealed that in the rush to develop the technological processes necessary to preserve authentic electronic records, cost issues have often been pushed aside. This is in part because ample government and foundation funding is allowing some institutions to defer cost modeling. Many respondents reported that they are beginning to study the cost implications, and we hope to gather more information in the next round.

As a result of the information we will gather over the next two years about evolving preservation practices, we expect to strengthen the foundation underlying the development of the preservation function model, particularly those aspects which concern preservation, storage, and access to authentic electronic records over time. We also hope to provide insights which will contextualize the work of projects and institutions around the world, and which will ultimately provide a pool of knowledge that will benefit us all.

* * * * * * * * *

When John Steinbeck completed *Of Mice and Men,* he described it to his publisher as an experiment, adding, "don't publish it if you don't like it." So unsure was the author of his work that he did not even want to read the proofs. Christopher Morley, describing for *Book of the Month Club News* the publication of Steinbeck's book, wrote that "[I]n just such casual ways, in this our world of obliquity and squint, do masterpieces happen."[xxvi] "Obliquity and squint" beautifully captures the notion of looking at something without fully understanding or seeing it. It is as apt a description of electronic preservation as it is a description of the chance publication of an enduring work of literature. In this period of incunable electronic information, it is difficult to understand all the potential or all of the pitfalls of our newest cultural heritage. Until we do, however, electronic preservation itself will be seen as oblique.

[i] Howard Besser, "Digital Longevity," in *Handbook for Digital Projects: A Management Tool for Preservation and Access.* (Andover, MA: Northeast Document Conservation Center, 2000): 156.

[ii] See, for example, *Research Libraries Group News* 52 (Spring 2001): 3, 8-9.

For various definitions of *authenticity* and *authenticity of digital information*, see Nancy Brodie, "Authenticity, Preservation and Access in Digital Collections," *The New Review of Academic Librarianship* 6 (2000):225-238.

Anne Gilliland-Swetland, *Enduring Paradigm, New Opportunities: The Value of the Archival Perspective in the Digital Environment.* (Washington, D.C.: Council on Library and Information Resources, February 2000): Gilliland-Swetland, 11-12.

Gilliland-Swetland, 13.

Margaret Hedstrom and Sheon Montgomery, *Digital Preservation Needs and Requirements in RLG Member Institutions*. (Mountain View, CA: Research Libraries Group, 1998). http://www.rlg.org/preserv/digpres.html

Available: In Anne R. Kenney and Oya Y. Rieger's *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Mountain View, CA: Research Libraries Group. 2000), the contributors cover particular topics (e.g. "What Users Want from Digital Image Collections," "Benchmarking for Conversion," "What About Copyright?"), but the book doesn't compare practices across institutions.

Reagan Moore, et al., "Collection-Based Persistent Digital Archives – Part 2," *D-Lib Magazine* 6.4 (April 2000): 1.

See also, Stewart Granger, "Emulation as a Digital Preservation Strategy," *D-Lib Magazine* 6.10 (October 2000): 1-12We were not able to establish contacts at two of the sites (the National Library of Sweden and the Department of Computer Science at Stanford University).

Ross, Seamus and Ann Gow. *Digital Archeology: Rescuing Neglected and Damaged Data Resources, A K\JISC/NPO Study.* (Glasgow: Humanities Advanced Technology and Information Institute (HATII), University of Glasgow, February 1999): iv. http://www.hatii.arts.gla.ac.uk

Hedstrom and Montgomery, 5.

These have been slightly copyedited.

Frank Evans, et al. "A Basic Glossary for Archivists, Manuscript Curators, and Records Managers." *The American Archivist* 37.3(July 1974): 415-433.

F.W. Ratcliffe with D. Patterson. *Preservation Policies and Conservation in British Libraries: Report of the Cambridge University Library Conservation Project.* Library and Information Research Report, 25. (London: The British Library, 1984).

J.M. Dureau and D.W.G. Clements, *Principles for the Preservation and Conservation of Library Materials.* IFLA Professional Reports 8. (The Hague: IFLA, 1986).

A definition that Feather, Matthews, and Eden came up with based on broadly accepted definitions. See John Feather, et al. *Preservation Management: Policies and Practices in British Libraries.* (Aldershot, England: Gower Publishing, 1996), p. 5.

[17] *Statement on the Preservation of Digitized Reproductions.* (Chicago: Society of American Archivists, 1997).

[18] International Council on Archives, *Guide for Managing Electronic Records From an Archival Perspective. ICA 8* (Paris: ICA, February 1997): 32.

[19] Anne R. Kenney and Oya Y. Rieger, *Moving Theory into Practice: Digital Imaging for Libraries and Archives.* (Mountain View, CA: Research Libraries Group, 2000), p. 135.

[20] Jeff Rothenberg is now advocating use of the term "inherently digital" to characterize executable, dynamic, and interactive materials. See his *Using Emulation to Preserve Digital Documents.* The Hague: Koninklijke Bibliotheek, July 2000.

[21] Hendley, Tony. *Comparison of Methods & Costs of Digital Preservation.* (1998) p. 96. http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html (downloaded October 2000)

[22] Hendley, p. 65.

[23] Russell, Kelly and Ellis Weinberger. *Cost Elements of Digital Preservation.* http://www.leeds.ac.uk/cedars/documents/CIW01r.html [downloaded August 2000].

Kenney and Rieger, p. 166.

[25] Lynch, Clifford. "Strategic Issues: Technology, Trends and Solutions." In Preserving Digital Information, Vantage Point series, pp. 3-4. EBSCO Subscription Services, 2000.

[26] Christopher Morley as quoted in, Al Silverman, ed., *The Book of the Month: Sixty Years of Books in American Life.* (Boston: Little, Brown, 1986): 49.