Findings on the Preservation of Authentic Electronic Records

Final Report to the National Historical Publications and Records Commission (Grants # 99-073 and # 2001-005)

September 2002

US-InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

This report and other products of the US-InterPARES Project are available on the project's website: <http://is.gseis.ucla.edu/us-interpares/>.

The final report of the international research team, *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, is on the international project's website: <http://www.interpares.org>.

US-InterPARES Project Researchers

Philip B. Eppard, School of Information Science and Policy, University at Albany, State University of New York (Co-Director)
Anne J. Gilliland-Swetland, Department of Information Studies, University of California, Los Angeles (Co-Director)

Jason Baron, National Archives and Records Administration
Robert Chadduck, National Archives and Records Administration
Su-Shing Chen, Department of Computer Engineering and Computer Science, University of Missouri-Columbia
Michèle V. Cloonan, Department of Information Studies, University of California, Los Angeles
Fynnette Eaton, Smithsonian Institution Archives
Sharon Farb, University Libraries, University of California, Los Angeles
Mark Giguere, National Archives and Records Administration
Lisa Haralampus, National Archives and Records Administration
P.C. Hariharan, Department of Chemistry, The Johns Hopkins University
Reagan Moore, San Diego Supercomputer Center
Leon Stout, Department of Special Collections, Pennsylvania State University
Kenneth Thibodeau, National Archives and Records Administration
William Underwood, Georgia Tech Research Institute

Research Assistants

School of Information Science and Policy, University at Albany, State University of New York: Andrew Ashton, Kevin Glick, Rebecca Hatcher, Irene Kaplan, Michelle Powers, Mark Wolfe

Department of Information Studies, University of California, Los Angeles: Francesca Marini, Eun Park, Marisol Ramos, Shelby Sanett, Kalpana Shankar, Melissa Taitano, Ciaran Trace, Julio Vera

Project Administrators

Kevin Glick
Richard Sloma

Acknowledgments

We wish to thank all the InterPARES Project staff for its administrative support of the work of the American team and all other aspects of the InterPARES Project. Staff members include: Tahra Fung (University of British Columbia), Kevin Glick (University at Albany, State University of New York), Jean-Pascal Morghese (University of British Columbia), Richard Sloma (University at Albany, State University of New York) and Peter Van Garderen (University of British Columbia). We would also like to thank the student research assistants at the Netherlands Institute for Archival Education and Research; the State Archives of Rome; the University at Albany, State University of New York; the University of British Columbia; the University of California, Los Angeles; and the University of Glasgow.

We would like to thank all following institutions, organizations, and their staff who provided their assistance in conducting the case studies. We also acknowledge gratefully the following agencies and institutions for their support: Banca d'Italia; Beijing Municipal Archives; the Central State Archives of Italy; the National Archives of Canada; the National Research Council of Italy; the Netherlands Institute for Archival Education and Research; the Smithsonian Institution; the Social Sciences and Humanities Research Council of Canada; the State Archives Administration of China; the State Archives of Rome; the National Archives and Records Administration; the University at Albany, State University of New York; the University of British Columbia; the University of California, Los Angeles; and the University of Glasgow.

In particular, we would like to acknowledge the National Historical Publications and Records Commission, whose funding made this project possible.

Table of Contents

Executive Summary

Electronic records have become essential elements in the life of organizations, businesses, government agencies, and individuals. Everything that we used to do on paper is now done, at least in part, electronically. Yet the intellectual and physical integrity of these records is at risk because they are vulnerable to change that can compromise their trustworthiness. We need to take steps that will ensure that we preserve the authenticity of records, which is at risk whenever records are transmitted across space and over time.

This report addresses the question of the long-term preservation of authentic electronic records. It reports on the work of the United States research team participating in the InterPARES Project, an international research initiative composed of experts in archival and computer science, preservation, and law drawn from national archives and academic, cultural, and corporate institutions in North America, Europe, Asia, and Australia. The team from the United States is supported for the research reported here with funding from the National Historical Publications and Records Commission and the National Archives and Records Administration.

The project builds on an earlier research project at the University of British Columbia, in collaboration with the U.S. Department of Defense Records Management Task Force. This previous project formulated requirements for creating, handling, and maintaining electronic records in active systems. These requirements were subsequently incorporated in 1997 into a standard for use by the U.S. Department of Defense in procuring records management software (DoD 5015.2-STD, Design Criteria Standard for Electronic Records Management Software Applications). The scope and complexity of larger issues surrounding the *long-term* preservation of authentic electronic records, however, pointed to the need for a broader, interdisciplinary, international approach.

InterPARES divided its research into four domains: authenticity, appraisal, preservation, and strategy. Each domain was addressed by a task force with a specific set of research questions, with the strategy domain integrating the outcomes and products of the other three domains into an action and implementation framework. This report discusses the general findings of the project along with some specific work of the U.S. team that contextualizes the research in the American environment.

The goal of the Authenticity Task Force was to identify conceptual requirements for assessing and maintaining the authenticity of electronic records. It conducted extensive case studies of electronic record systems and evaluated the data collected against a Template for Analysis, which was a construct of all the elements that would constitute a record. This research resulted in the construction of a detailed profile of the complexity of contemporary electronic records and in the identification of the extent to which the records are embedded within the specific juridical-administrative, provenancial, procedural, documentary, and technological contexts in which they are created. The Task Force found that most contemporary records systems are a hybrid of electronic and paper records, that few explicit measures are employed to ensure the authenticity of electronic records, and that authenticity is generally assured through procedural means. The Task Force then developed two sets of conceptual requirements. The first set (Benchmark Requirements) includes requirements that support the presumption of the authenticity of electronic records before they are transferred to the preserver’s custody, while the second set (Baseline Requirements) includes requirements that support the production of authentic copies of electronic records after they have been transferred to the preserver’s custody.

The Appraisal Task Force considered whether electronic records should be appraised differently from traditional records, when they should be appraised, and who should do it. The primary product of the

research was the creation of a series of functional models depicting and decomposing the activities in the process named “Select Electronic Records.” The model identifies and clarifies the steps involved in the appraisal and disposition of electronic records. In addition to the traditional practice of basing appraisal decisions on judgments of continuing value, it recommends that appraisal of electronic records needs to take into account an assessment of the authenticity of the records and the feasibility of their preservation, specifically the feasibility of their preservation from a technological standpoint. It recommends early appraisal of electronic records and regular monitoring of appraisal decisions to ensure that changes to the records and their contexts over time have not negatively affected their identity or integrity or the ability to preserve them.

The goal of research in the preservation domain was to identify and develop the procedures and resources required for the implementation of the conceptual requirements and the criteria identified in the authenticity and appraisal domains. The Preservation Task Force surveyed existing digital preservation programs, conducted a study of storage media, and gathered information on methods of authentication. Researchers constructed a model of the preservation function, which provides a framework for preserving electronic records that archival institutions can use to satisfy the Authenticity Task Force's Baseline Requirements. Components of this framework include preservation strategies for specific bodies of records, preservation methods, preservation action plans, and terms and conditions for the transfer of electronic records. The model does not prescribe a particular computer system or recommend specific technological tools or methods for preservation. It can be used, however, for analyzing the problem of preserving electronic records and for guiding the choice and evaluation of various technological options and specific strategies for ensuring the continuing availability of authentic copies of records over time and over different generations of technology.

The Strategy Task Force devised an intellectual framework to support the development of policies, strategies, and standards facilitating the long-term preservation of authentic electronic records and a set of principles and criteria that should govern the development of any records preservation policy, strategy, or standard. The US-InterPARES research team prepared a detailed analysis of the stakeholder groups which should be concerned about the preservation of authentic electronic records and reviewed existing laws and regulations that are directly relevant to archivists who are working to preserve electronic records in the United States. As part of the effort to translate the research products into outcomes, the project has been developing workshops and model curricula that will facilitate dissemination of the research methods and findings of InterPARES.

Additional research questions emerged during the course of the work and many of these will be addressed in the second phase of InterPARES, which began in 2002. This new phase of the research is addressing problems involved in the creation, maintenance, and preservation of authentic records created in emerging experiential, dynamic, and interactive systems in the artistic, scientific, and governmental sectors.

I. Introduction

Electronic records, that is records that are created and maintained in electronic form, are an inescapable part of modern bureaucratic environments and, indeed, twenty-first century life. These records come in all shapes and sizes, and may be more or less like, or integrated with, traditional paper records. Authenticity, however, is a quality of any record that we often take for granted until it is challenged. That is when we realize that because of—or in spite of—preservation and other long-term management processes, the essential nature of the electronic record—its intellectual and physical integrity—can easily be changed or compromised in ways that throw its authenticity into doubt. In recent months alone, we have seen several instances—including the destruction of Enron electronic records, rejection by U.S. universities of academic transcripts supplied by overseas students, calls for audit trails for electronic and Internet-based voting—where the technological or procedural controls over records creation, maintenance, and disposition have come under scrutiny. It is, therefore, essential that we understand what it takes to assure the authenticity of records. An authentic record is one that is what it claims to be. It is genuine. It has not been counterfeited or tampered with, and it is free from corruption.

Our trust in such records often rests in controls we have put in place during their creation, transmission, and active use. For electronic records, these increasingly take the form of technological and procedural controls, such as electronic signatures, audit trails, and quality control measures. However, these are reliability guarantees that control the record only *in time* and not *over time*, and they are often only partial guarantees at that. Some records, such as articles of incorporation, contracts, property deeds, and registered patents need to endure and remain trustworthy for a long time, sometimes permanently. What then, should be our controls, guarantees, strategies, policies, and methods over time for establishing and maintaining the authenticity of a record that must be preserved in the long term for future uses? Are they substantially different for electronic records than those conventions that have developed over several hundred years for traditional paper records?

These are the essential questions driving the research of InterPARES (International Research on Permanent Authentic Records in Electronic Systems). One of the largest archival research collaborations ever undertaken, InterPARES is an international research initiative created to help solve the critical problem of preserving trustworthy electronic records for future use. InterPARES is an acronym for International Research on Permanent Authentic Records in Electronic Systems. The word is also a composite of the Latin *inter* (“between,” or “among”) and *par* (“equal”) joined together to mean “among equals.” Thus the name signifies that researchers from many countries and different disciplines have been working together as equal partners. InterPARES is made up of a group of experts in archival and computer science, preservation, and law drawn from national archives and academic, cultural, and corporate institutions in North America, Europe, Asia, and Australia. It is organized into national teams and a global industry team. The team from the United States is supported for the research reported herein with funding from the National Historical Publications and Records Commission and the National Archives and Records Administration. The co-directors of the United States team are Dr. Philip B. Eppard from the University at Albany, State University of New York, and Dr. Anne J. Gilliland-Swetland from the University of California, Los Angeles. The director of the InterPARES Project is Dr. Luciana Duranti from the University of British Columbia in Canada.

The InterPARES Project had its roots in an earlier project. In the late 1990s, a research project at the University of British Columbia, in collaboration with the U.S. Department of Defense Records Management Task Force, titled “The Preservation of the Integrity of Electronic Records” (UBC Project), formulated requirements for creating, handling, and maintaining electronic records in active systems. These requirements were subsequently incorporated in 1997 into a standard for use by the U.S. Department of Defense in procuring records management software (DoD 5015.2-STD, Design Criteria Standard for Electronic Records Management Software Applications). The scope and complexity of larger issues surrounding the long-term preservation of authentic electronic records pointed to the need for a broader, interdisciplinary, international approach. The InterPARES Project was officially launched on January 1, 1999. The first phase of the project (InterPARES 1) concluded in Spring 2002. The second phase (InterPARES 2) is now underway and is planned to conclude in 2007.

This publication reports on the findings of the first three years of research and development activities of the American team participating in InterPARES 1 (referred to hereafter as US-InterPARES), in particular the development of *Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records*, *Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records*, and sets of activity models that decompose the archival functions of appraising electronic records and preserving authentic copies of archival electronic records. The document also frames and presents several other strategic and educational products developed by US-InterPARES.

A considerable amount of recent research and development has addressed related areas, notably digital signature and encryption technologies; models and technologies for acquiring, archiving, and accessing digital objects; XML-based initiatives such as XBML, XBRL, and Legal XML; and legislative and regulatory developments on areas such as the legality and security of business-to-business transactions. InterPARES, however, has been the only major interdisciplinary project to emanate out of the archival community and to promote an evidence-based approach to the preservation of authentic electronic *records*. InterPARES has maintained a consistent focus on the needs of records and recordkeeping, a strong theoretical foundation in archival science, and a commitment to demonstrating the utility and relevance of archival science in the digital age. Perhaps the most important outcome of InterPARES is how much we, as researchers have learned about the complexity of the questions we have been asking. This knowledge has enabled us to establish and investigate a deep and sophisticated research agenda that we are continuing to pursue as a multi-community effort through the work of InterPARES 2.

InterPARES 1 divided its research into four domains: authenticity, appraisal, preservation, and strategy. Each domain was addressed by a task force with a specific set of research questions, with the strategy domain integrating the outcomes and products of the other three domains into an action and implementation framework. Section II of this report summarizes the work of the first three domains, while Section III, *Translating Research Outcomes into Practice*, discusses the action and implementation implications of the work of the first three domains. The concluding section indicates areas for further research that arose during the course of InterPARES, and the appendices provide examples of some of the additional products of the research. Additional products referred to in the text of the report are available on the US-InterPARES website <http://is.gseis.ucla.edu/us-interpares>. The final report of the international team, *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project*, which includes more detailed reports on the activities within each research domain, is available on the InterPARES website <http://www.interpares.org/book/index.htm>.

II. InterPARES Research Activities and Outcomes

A. What is required to prove the authenticity of electronic records?

The goal of the Authenticity Task Force was to identify conceptual requirements for assessing and maintaining the authenticity of electronic records according to five initial research questions:

- What are the elements that all electronic records share?
- What are the elements that allow us to differentiate between different types of electronic records?
- Of those elements, which will permit us to verify their authenticity over time?
- Are the elements for verifying authenticity over time the same as those that permit us to verify their authenticity in time, that is, at the point at which they are originally created and transmitted?
- Can the elements be removed from where they are currently found to a place where they can more easily be preserved and still maintain the same validity?

This section, therefore, discusses and presents the primary products that resulted from those activities, namely the *Template for Analysis*, the *Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records*, and the *Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records*.[1]

To achieve its goal, the Authenticity Task Force adopted two distinct, yet related analytical approaches. The first approach was theoretical and deductive, based on contemporary archival diplomatics. The second approach was inductive and empirical and employed selected case studies of extant electronic systems.[2] Between Spring 1999 and Spring 2001, four successive rounds of case studies were conducted by institutional and student researchers in government, university, and corporate agencies in the United States, Canada, Italy, the United Kingdom, the Netherlands, and China. The case studies included large-scale databases (such as patent and student registration systems), geographic information systems, and interactive web-based applications. The data gathered through the case studies was then used to test and extend a *Template for Analysis* that directly addressed the research questions established for the Authenticity Domain. Using a diplomatics-based framework, the *Template* enabled us to deconstruct and understand the relevance of elements of existing and future types of electronic records in terms of how they support a presumption of authenticity.[3] (The full text of the *Template* is presented in Appendix A.) The translation of the case study data into a form that could be analyzed diplomatically by the *Template* was achieved by coding the data for inter-related themes and concepts using a *Template Element Data Gathering Instrument (TEDGI)*.

[1] The complete report of the Authenticity Task Force is contained in *Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* available on the InterPARES Website at: <http://www.interpares.org/book/index.htm>.

[2] For a fuller discussion of the methods used by the Authenticity Task Force, see Heather MacNeil, "Providing Grounds for Trust: Developing Conceptual Requirements for the Long-Term Preservation of Authentic Electronic Records," *Archivaria* 50 (Fall 2000): 52-78; Anne J. Gilliland-Swetland, "Testing Our Truths: Delineating the Parameters of the Authentic Archival Electronic Record," *American Archivist* 65(Fall/Winter 2002), forthcoming; and Ciaran Trace, "Applying Content Analysis to Case Study Data: A Preliminary Report," June 2001. Available on the US-InterPARES website: <http://is.gseis.ucla.edu/us-interpares/>.

[3] The principles of diplomatics were first delineated in 1681 by Jean Mabillon, and have subsequently been developed and extended to address a growing body of document types. Diplomatics has been employed as a means for assessing authenticity for legal, philological, and historiographical as well as archival purposes.

The deductive and the inductive approaches resulted in the construction of a detailed profile of the complexity of contemporary electronic records and in the identification of the extent to which the records are embedded within the specific juridical-administrative, provenancial, procedural, documentary, and technological contexts in which they are created.[4] The Task Force found that most contemporary records systems are a hybrid of electronic and paper records, that few explicit measures are employed to ensure the authenticity of electronic records, and that authenticity is generally assured through procedural means.

The other primary outcome of the work of the Task Force has been the development of two sets of conceptual requirements [see Tables 1 and 2]. The first set includes requirements that support the presumption of the authenticity of electronic records before they are transferred to the preserver's custody, while the second set includes requirements that support the production of authentic copies of electronic records after they have been transferred to the preserver's custody. Taken together with the preservation model developed by the Preservation Task Force, these requirements also provide a basis for an analytical framework for evaluating digital technology, media, migration processes, and conversion methods that will continue to be developed through the work of InterPARES 2.

1. Conceptual Framework for the Requirements for Assessing and Maintaining the Authenticity of Electronic Records—the Benchmark and Baseline Requirements

Authenticity is defined as "the quality of being authentic, or entitled to acceptance."[5] *Authentic* means "worthy of acceptance or belief as conforming to or based on fact" and is synonymous with the terms *genuine* and *bona fide*. *Genuine* "implies actual character not counterfeited, imitated, or adulterated [and] connotes definite origin from a source." *Bona fide* "implies good faith and sincerity of intention."[6] From these definitions it follows that an *authentic record* is a record that is what it purports to be and is free from tampering or corruption.

In both archival theory and jurisprudence, records that the creator relies on in the usual and ordinary course of business are presumed authentic.[7] However, digital information technology creates significant risks that electronic records may be altered, either inadvertently or intentionally. Therefore, in the case of records maintained in electronic systems, the presumption of authenticity must be supported by evidence that a record is what it purports to be and has not been modified or corrupted in essential respects. To assess the authenticity of an electronic record, the preserver must be able to establish its *identity* and demonstrate its *integrity*.

The identity of a record refers to the distinguishing character of a record, that is, the attributes of a record that uniquely characterize it and distinguish it from other records. From an archival-diplomatic perspective, such attributes include: the names of the persons concurring in its formation (i.e., its author, addressee, writer, and originator); its date(s) of creation (i.e., the date it was made, received, and set aside) and its date(s) of transmission; an indication of the action or matter in which it participates; the expression of its archival bond, which links it to other records participating in the same action (e.g., a classification code or other unique identifier); as well as an indication of any attachment(s) since an

[4] See Anne J. Gilliland-Swetland and Philip B. Eppard, "Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project," *D-Lib Magazine*, 6 (July/August 2000). Available at: <http://www.dlib.org/dlib/july00/eppard/07eppard.html>.

[5] *Oxford English Dictionary*, 2nd ed., s.v. "authenticity."

[6] *Merriam-Webster Online Collegiate Dictionary*, s.v. "authentic."

[7] The creator is the physical or juridical person in whose archival fonds the record exists. The *fonds* is the whole of the records created (meaning made or received and set aside for action or reference) by a physical or juridical person in the course of carrying out its activities.

attachment is considered an integral part of a record.[8] The attributes that establish the identity of a record may be explicitly expressed in an element of the record, in metadata related to the record, or they may be implicit in its various contexts.[9] Those contexts include: its *documentary context*, that is, the archival aggregation to which a record belongs, and its internal structure; its *procedural context*, that is, the business process in the course of which the record is created; its *technological context*, that is, the characteristics of the technical components of an electronic computing system in which records are created; its *provenancial context*, that is, the creating body, its mandate, structure, and functions; and its *juridical-administrative* context, that is, the legal and organizational system in which the creating body belongs.

The *integrity* of a record refers to its wholeness and soundness: a record has integrity when it is complete and uncorrupted in all its essential respects. This does not mean that the record must be precisely the same as it was when first created in order for its integrity to exist and be demonstrated. Even in the paper world, records are subject to deterioration, alteration, and/or loss with the passage of time. In the electronic world, the fragility of the media, the obsolescence of technology, and the idiosyncrasies of systems likewise affect the integrity of records. When we refer to an electronic record, we consider it essentially complete and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered. This implies that its physical integrity, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and any required annotations and elements of documentary form remain the same.[10] The integrity of a record may be demonstrated by evidence found on the face of the record, in metadata related to the record, or in one or more of its various contexts.

The Benchmark Requirements are the conditions that serve as a basis for the preserver's assessment of the authenticity of the creator's electronic records. A presumption of authenticity will be based upon the number of requirements that have been met and the degree to which each has been met. The requirements are, therefore, cumulative; the higher the number of satisfied requirements, and the greater the degree to which an individual requirement has been satisfied, the stronger the presumption of authenticity. This is why these requirements are termed "benchmark" requirements. Satisfaction of these Benchmark Requirements will enable the preserver to infer a record's authenticity on the basis of the manner in which the records have been created, handled, and maintained by the creator. The assessment of the authenticity of the creator's records takes place as part of the appraisal function. This assessment should be verified when the records are transferred to the preserver's custody. (Both the appraisal and the preservation functions are delineated in more detail in the models developed within the Appraisal and Preservation Domains discussed later in this section).

[8] An attachment is a document that constitutes an integral part of the whole record, notwithstanding the fact that it exists as a linked, but physically separate entity.

[9] The use of the terms *attribute* and *element* in this report should not be confused with the way the terms are used in other contexts, such as the various Standard Generalized Mark-up Languages (SGML). In this report, a record attribute is a defining characteristic of a record or of a record element. A *record element* is a constituent part of the record's documentary form and may be either extrinsic or intrinsic. An attribute may manifest itself in one or more elements of a record's documentary form. For example, the name of the author of a record is an attribute, which may be expressed as a superscription or a signature, both of which are intrinsic elements of documentary form. For a more detailed explanation of the extrinsic and intrinsic elements of documentary form see the *Template for Analysis* in Appendix A. An attribute may also manifest itself in the form of an annotation to a record, in metadata linked to it, or in one or more of its various contexts.

[10] For example, for an electronic mail message, an authentic copy of a complete message may include only the text. Provided it clearly indicated the author, addressee, receivers, and date as well as the content, it would not need to appear in the same way in which it was seen by the author or addressee. In contrast, an authentic copy of a map would have to retain its original presentation features, including color and feature presentation. Provided these requirements were met, an authentic copy could be produced in GIF, JPEG, or GML format.

The establishment and implementation of the Baseline Requirements take place as part of the function of managing preservation. After the records have been presumed or verified authentic in the appraisal process, and have been transferred from the creator to the preserver, their authenticity needs to be maintained by the preserver. In order to do so, the preserver must carry forward the records in accordance with the Baseline Requirements that apply to the maintenance of records, producing copies according to procedures that also maintain authenticity.[11] The production of authentic copies is regulated by the Baseline Requirements. Unlike the Benchmark Requirements, all of the requirements included in the Baseline Requirements must be met before the preserver can attest to the authenticity of the electronic copies in its custody. This is why the requirements for the production of authentic electronic copies are termed "baseline" requirements. Satisfaction of these Baseline Requirements will enable the preserver to certify that copies of electronic records are authentic. Traditionally, the official preserver of the records has been the person entrusted with issuing authentic copies of such records. To fulfill that role, the preserver needed simply to attest that the copy conformed to the record being reproduced. With electronic records, and the accompanying difficulties related to preservation, the prudent path would be for the preserver to produce and maintain documentation relating to the manner in which it has maintained the records over time as well as the manner in which it has reproduced them. This documentation would help support its attestation of authenticity.

Within the Benchmark Requirements, Requirement A.1 identifies the core information about an electronic record—the immediate context of its creation and the manner in which it has been handled and maintained—that establishes the record's identity and lays a foundation for demonstrating its integrity [Table 1]. Requirements A.2–A.8 identify the kinds of procedural controls over the record's creation, handling, and maintenance that support a presumption of its integrity.

[11] It is understood that the records maintained by the preserver exist only as copies of the creator's records.

Table 1. Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records (Requirement Set A)

	To support a presumption of authenticity the preserver must obtain evidence that:
REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record	the value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records. ***A.1.a*** Identity of the record: ***A.1.a.i*** Names of the persons concurring in the formation of the record, that is: • name of author[12] • name of writer[13] (if different from the author) • name of originator[14] (if different from name of author or writer) • name of addressee[15] ***A.1.a.ii*** Name of action or matter ***A.1.a.iii*** Date(s) of creation and transmission, that is: • chronological date[16] • received date[17] • archival date[18] • transmission date(s)[19] ***A.1.a.iv*** Expression of archival bond[20] (e.g., classification code, file identifier) ***A.1.a.v*** Indication of attachments

[12] The name of the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.

[13] The name of the physical or juridical person having the authority and capacity to articulate the content of the record.

[14] The name of the physical or juridical person assigned the electronic address in which the record has been generated and/or sent.

[15] The name of the physical or juridical person(s) to whom the record is directed or for whom the record is intended.

[16] The date, and possibly the time, of compilation of a record included in the record by the author or the electronic system on the author's behalf.

[17] The date, and possibly the time, when a record is received by the addressee.

[18] The date, and possibly the time, when a record is officially incorporated into the creator's records.

[19] The date and time when a record leaves the space in which it was generated.

[20] The archival bond is the relationship that links each record, incrementally, to the previous and subsequent ones and to all those participate in the same activity. It is originary (i.e., it comes into existence when a record is made or received and set aside), necessary (i.e., it exists for every record), and determined (i.e., it is characterized by the purpose of the record).

A.1.b Integrity of the record:

- ***A.1.b.i*** Name of handling office[21]
- ***A.1.b.ii*** Name of office of primary responsibility[22] (if different from handling office)
- ***A.1.b.iii*** Indication of types of annotations added to the record[23]
- ***A.1.b.iv*** Indication of technical modifications;[24]

REQUIREMENT A.2: Access Privileges	the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records;

REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records	the creator has established and effectively implemented procedures to prevent, discover, and correct loss or corruption of records;

REQUIREMENT A.4: Protective Procedures: Media and Technology	the creator has established and effectively implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change;

REQUIREMENT A.5: Establishment of Documentary Forms	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;

REQUIREMENT A.6: Authentication of Records	if authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication;

REQUIREMENT A.7: Identification of Authoritative Record	if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative;

REQUIREMENT A.8: Removal and Transfer of Relevant Documentation	if there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and effectively implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

[21] The office (or officer) formally competent for carrying out the action to which the record relates or for the matter to which the record pertains.

[22] The office (or officer) given the formal competence for maintaining the authoritative record, that is, the record considered by the creator to be its official record.

[23] Annotations are additions made to a record after it has been completed. Therefore, they are not considered elements of the record's documentary form.

[24] Technical modifications are any changes in the digital components of the record as defined by the Preservation Task Force. Such modifications would include any changes in the way any elements of the record are digitally encoded and changes in the methods (software) applied to reproduce the record from the stored digital components; that is, any changes that might raise questions as to whether the reproduced record is the same as it would have been before the technical modification. The indication of modifications might refer to additional documentation external to the record that explains in more detail the nature of those modifications.

Commentary on the Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records

A.1 Expression of Record Attributes and Linkage to Record

The presumption of a record's authenticity is strengthened by knowledge of certain basic facts about it. The attributes identified in this requirement embody those facts. The requirement that the attributes be expressed explicitly and linked inextricably[25] to the record during its life, and carried forward with it over time and space, reflects the Task Force's belief that such expression and linkage provide a strong foundation on which to establish a record's identity and demonstrate its integrity. The case studies undertaken as part of the work of the Task Force revealed very little consistency in the way the attributes that specifically establish the identity of a record are captured and expressed from one electronic system to another. In certain systems, some attributes were explicitly mentioned on the face of the record; in others they could be found in a wide range of metadata linked to the record or they were simply implicit in one or more of the record's contexts. In many cases, certain attributes (e.g., the expression of the archival bond) were not captured at all. The Task Force's concern is that, in the absence of a precise and explicit statement of the basic facts concerning a record's identity and integrity, it will be necessary for the preserver to acquire enormous, and otherwise unnecessary, quantities of data and documentation simply to establish those facts.

The link between the record and the attributes listed in Requirement A.1 is viewed by the Task Force as a *conceptual* rather than a *physical* one, and the requirement could be satisfied in different ways, depending on the nature of the electronic system in which the record resides. For example, in electronic records management systems, this requirement is usually met through the creation of a record profile.[26] In other types of systems, the requirement could be fulfilled through a topic map. A topic map expresses the characteristics (i.e., *topics*) of subjects (e.g., records or record attributes) and the relationships between and among them.

When a record is exported from the live system, migrated in a system update, or transferred to the preserver, the attributes should be linked to the record and available to the user. When pulling together the data prior to export, the creator should also ensure that the data captured are the right data. For example, in the case of distribution lists, the creator must ensure that if the recipients specified on "List A" were changed at some point in the active life of records, the accurate "List A: Version 1" is exported with the records associated with the first version, and that the second version is sent forward with those records sent to recipients on "List A: Version 2."

A.2 Access Privileges

Defining access privileges means assigning responsibility for the creation, modification, annotation, relocation, and destruction of records on the basis of competence, which is the authority and capacity to carry out an administrative action. Implementing access privileges means conferring exclusive capability to exercise such responsibility. In electronic systems, access privileges are usually articulated in tables of user profiles. Effective implementation of access privileges involves the monitoring of access through an audit trail that records every interaction that an officer has with each record (with the possible exception

[25] For the purposes of this requirement, inextricable means incapable of being disentangled or untied, and link means a connecting structure.

[26] If the attribute values contained in the profile are also expressed independently as entries in a register of all records made or received by the creator, then, in addition to establishing the identity and supporting the inference of the integrity of the record, they would corroborate such identity and strengthen the inference of integrity.

of viewing the record). If the access privileges are not embedded within the electronic system but are based on an external security system (such as the exclusive assignment of keys to a location), the effective implementation of access privileges will involve monitoring the security system.

A.3 Protective Procedures: Loss and Corruption of Records

Procedures to protect records against loss or corruption include: prescribing regular back-up copies of records and their attributes; maintaining a system back-up that includes system programs, operating system files, etc.; maintaining an audit trail of additions and changes to records since the last periodic back-up; ensuring that, following any system failure, the back-up and recovery procedures will automatically guarantee that all complete updates (records and any control information such as indexes required to access the records) contained in the audit trail are reflected in the rebuilt files and also guarantee that any incomplete operation is backed up. The capability should be provided to rebuild forward from any back-up copy, using the back-up copy and all subsequent audit trails.

A.4 Protective Procedures: Media and Technology

Procedures to counteract media fragility and technological obsolescence include: planning upgrades to the organization's technology base; ensuring the ability to retrieve, access, and use stored records when components of the electronic system are changed; refreshing the records by regularly moving them from one storage medium to another; and migrating records from an obsolescent technology to a new technology.

A.5 Establishment of Documentary Forms

The documentary form of a record may be determined in connection to a specific administrative procedure, or in connection to a specific phase(s) within a procedure. The documentary form may be prescribed by business process and workflow control technology, where specific record forms identify each step in an administrative procedure. If a creator customizes a specific application, such as an electronic mail application, to carry certain fields, the customized form becomes, by default, the required documentary form. It is understood that the creator, acting either on the basis of its own needs or the requirements of the juridical system, not an individual officer, establishes the required documentary form(s) of records.

When the creator establishes the documentary form in connection to a procedure, or to specific phases of a procedure, it is understood that this includes the determination of the intrinsic and extrinsic elements of form[27] that will allow for the maintenance of the authenticity of the record. Because, generally speaking, that determination will vary from one form of a record to another, and from one creator to another, it is not possible to predetermine or generalize the relevance of specific intrinsic and extrinsic elements of documentary form in relation to authenticity.

A.6 Authentication of Records

In common usage, to authenticate means to prove, or serve to prove the authenticity of something. More specifically, the term implies establishing genuineness by adducing legal or official documents or expert opinion. For the purposes of the Benchmark Requirements, authentication is understood to be a declaration of a record's authenticity at a specific point in time by a juridical person entrusted with the

[27] The extrinsic and intrinsic elements of form are defined and explained in the Authenticity Task Force's *Template for Analysis* in Appendix A.

authority to make such declaration. It takes the form of an authoritative statement (which may be in the form of words or symbols) that is added to or inserted in the record attesting that the record is authentic.[28] The requirement may be met by linking the authentication of specific types of records to business procedures and assigning responsibility to a specific office or officer for authentication.

The authentication of copies differs from the validation of the process of reproduction of the digital components of the records. The latter process occurs every time the records of the creator are moved from one medium to another or migrated from one technology to another.

A.7 Identification of Authoritative Record

An authoritative record is a record that is considered by the creator to be its official record and is usually subject to procedural controls that are not required for other copies. The identification of authoritative records corresponds to the designation of an office of primary responsibility as one of the components of a record retention schedule. The Office of Primary Responsibility is the office given the formal competence for maintaining the authoritative (that is, official) records belonging to a given class within an integrated classification scheme and retention schedule. The purpose of designating an Office of Primary Responsibility for each class of record is to reduce duplication and to designate accountability for records. It is understood that in certain circumstances there may be multiple authoritative copies of records, depending on the purpose for which the record is created.

A.8 Removal and Transfer of Relevant Documentation

This requirement implies that the creator needs to carry forward with the removed records all the information that is necessary to establish the identity and demonstrate the integrity of those records, as well as the information necessary to place the records in their relevant contexts.

[28] The meaning of authentication as it is used by the Authenticity Task Force in this report is broader than its meaning in public key infrastructure (PKI) applications. In such applications, authentication is restricted to proving identity and public key ownership over a communication network.

Table 2. Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records (Requirement Set B)

The preserver should be able to demonstrate that:

REQUIREMENT B.1: Controls over Records Transfer, Maintenance, and Reproduction	the procedures and system(s) used to transfer records to the archival institution or program; maintain them; and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that **B.1.a** Unbroken custody of the records is maintained; **B.1.b** Security and control procedures are implemented and monitored; and **B.1.c** The content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.

REQUIREMENT B.2: Documentation of Reproduction Process and its Effects	the activity of reproduction has been documented, and this documentation includes **B.2.a** The date of the records' reproduction and the name of the responsible person; **B.2.b** The relationship between the records acquired from the creator and the copies produced by the preserver; **B.2.c** The impact of the reproduction process on their form, content, accessibility and use; and **B.2.d** In those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user;

REQUIREMENT B.3: Archival Description	the archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created.

Commentary on the Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records

B.1 Controls over Records Transfer, Maintenance, and Reproduction

The controls over the transfer of electronic records to archival custody include establishing, implementing, and monitoring procedures for registering the records' transfer; verifying the authority for transfer; examining the records to determine whether they correspond to the records that are designated in the terms and conditions governing their transfer; and accessioning the records.

As part of the transfer process, the assessment of the authenticity of the creator's records, which has taken place as part of the appraisal process, should be verified. This includes verifying that the attributes relating to the records' identity and integrity have been carried forward with them (Requirement A.1), along with any relevant documentation (Requirement A.8).

The controls over the maintenance of electronic records once they have been transferred to archival custody are similar to several of the ones enumerated in the Benchmark Requirements. For example, the preserver should establish access privileges concerning the access, use, and reproduction of records (Requirement A.2); establish procedures to prevent, discover, and correct loss or corruption of records (Requirement A.3), as well as procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change (Requirement A.4). Once established, the privileges and procedures should be effectively implemented and regularly monitored. If authentication of the records is required, the preserver should establish specific rules regarding who is authorized to authenticate them and the means of authentication that will be used (Requirement A.6).

The controls over the reproduction of records include establishing, implementing, and monitoring reproduction procedures that are capable of ensuring that the content of the record is not changed in the course of reproduction.

B.2 Documentation of Reproduction Process and its Effects

Documenting the reproduction process and its effects is an essential means of demonstrating that the reproduction process is transparent (i.e., free from pretence or deceit). Such transparency is necessary to the effective fulfillment of the preserver's role as a trusted custodian of the records. Documenting the reproduction process and its effects is also important for the users of records since the history of reproduction is an essential part of the history of the record itself. Documentation of the process and its effects provides users of the records with a critical tool for assessing and interpreting the records.

B.3 Archival Description

Traditionally it has been a function of archival description to authenticate the records and perpetuate their administrative and documentary relationships. With electronic records, this function becomes critical. Once the records no longer exist except as authentic copies, the archival description is the primary source of information about the history of the record, that is, its various reproductions and the changes to the record that have resulted from them. While it is true that the documentation of each reproduction of the record copies[29] may be preserved, the archival description summarizes the history of all the reproductions, thereby obviating the need to preserve all the documentation for each and every reproduction. In this respect, the description constitutes a collective attestation of the authenticity of the records and their relationships in the context of the fonds to which the records belong. This is different from a certificate of authenticity, which attests to the authenticity of individual records. The importance of this collective attestation is that it authenticates and perpetuates the relationships between and among records within the same

2. Typing Electronic Records According to Their Authenticity Requirements

One additional Authenticity Task Force activity was the quest to develop a typology of attributes supporting the authenticity of different types of electronic records. While the Task Force was successful in

[29] Although, technically, every reproduction of a record that follows its acquisition by the preserver is an authentic copy, it is the only record that exists and, therefore, should normally be referred to as "the record" rather than as "the copy."

developing a conceptual framework for establishing the requirements for preserving authentic electronic records, it did not succeed in creating a single, comprehensive typology of authenticity requirements for electronic records. Instead, as the following discussion indicates, it identified possible perspectives from which a typology could be constructed and which merit further exploration.

The relationship between any typology and an analytical framework must necessarily be symbiotic, since a typology, unlike a taxonomy, attempts to delineate an extensible schema within which not only known, but also unknown objects or phenomena can be typed. A typology may be derived “top-down” based upon a theoretical analysis of the nature of objects. It may also emerge “bottom-up” from observation of objects. Optimally, a typology should be developed through some combination of both approaches. A typology identifies characteristics or properties that objects have in common, as well as those that make them distinct, thus enabling them to be identified and acted upon individually or as members of types or classes of objects. InterPARES’ theoretical underpinnings in diplomatics provided a strong top-down approach that enabled us to break down the different elements of records and their contexts, as well as their roles in establishing the authenticity of the records. The case studies conducted by InterPARES provided bottom-up insights into the nature of electronic records and record-keeping systems, as well as other information objects commonly collocated in today’s organizational environments. These two approaches together provided a rich set of inputs for the development of a typology and analytical framework. However, these inputs proved difficult to reconcile.

Several other issues have also worked to confound the development of a typology. Records are highly contingent objects, and any typology requires the choice of a single viewpoint from which to work. A primary issue to be resolved, therefore, was what should be modeled within a typology. Was it types of records? Types of record-keeping systems? Or was it different clusters of record elements or requirements for establishing authenticity with greater or lesser degrees of certainty? Another key problem was the dynamism of the electronic environment with a resulting lack of well-understood or canonical forms, even for the by-products of common recordkeeping activities. We were faced with two critical questions: Could a typology ever be sufficiently extensible, if not to anticipate, at least to be able to accommodate emerging record types? Can any typology meaningfully represent the many hybrid recordkeeping systems that today straddle both electronic and paper forms?

There are several different ways in which typologies could be constructed to help archivists identify and apply authenticity requirements to electronic records: by function associated with the record; by type of business system generating the record; and by type of information object. The first such typology would involve classification based on the functions or activities of records. An example of this approach would be the DAVID Project in Antwerp, where information systems were examined and grouped according to the business function they performed and the associated business processes.[30] A second typological approach would model system types. In such a case, one could either type by kind of system, e.g., registry, electronic mail, financial system; or one could type by the role the system plays within the organization, e.g., executive support system, strategic information system. System types are commonly delineated in both of these ways in business information systems, and computing information systems literature. In both of these approaches, however, the increasing deployment of multi-function and multi-user systems, as well as the embeddedness of those systems in a wider procedural and documentary context make them problematic for the basis of a typology that will assist archivists in establishing authenticity requirements for permanent records. A third approach is to develop a typology that relates to the information objects. An *information object* is commonly defined as “a digital item or group of items referred to as a unit, regardless of type or format, that a computer can address or manipulate as a single

[30] For information on the DAVID Project, see <http://www.antwerpen.be/david/eng/index.htm>.

object."[31] In this typology, however, information objects would include not only electronic records, but a diversity of other non-digital and hybrid digital/non-digital records and other types of information objects. After considerable discussion and analysis, the Task Force decided that a typology based upon individual creators and the acts/procedures/functions they carry out is likely to be the most effective starting point in any typification of electronic records, but this was beyond the scope of InterPARES 1 to complete.

B. How do we select electronic records for preservation?

The InterPARES Appraisal Task Force was charged with addressing the following research questions:

- What is the influence of digital technology on appraisal?
- What is the influence of retrievability, intelligibility, functionality, and research needs on appraisal?
- What are the influences of the medium and physical form of the record on appraisal?
- When in the course of their existence should electronic records be appraised?
- Should electronic records be appraised more than once in the course of their existence, and if so, when?
- Who should be responsible for appraising electronic records?
- What are the appraisal criteria and methods for authentic electronic records?

The Task Force engaged in three different sets of research activities in its efforts to answer these questions. First, it conducted a literature review of the existing writings on the appraisal of electronic records. Second, it examined policies, procedures, and appraisal reports from archival institutions that had experience in the appraisal of electronic records. Finally, the Task Force's most significant activity was the creation of a series of functional models depicting and decomposing the activities in the process that we named "Select Electronic Records." The "select" function was viewed as a high-level activity encompassing both appraisal decision-making and the disposition of records, either by destruction or by transfer to a preserver.

The Appraisal Task Force's literature review together with the complete IDEF0 model of the "Select Electronic Records" function are included as an appendix in the published final report of the InterPARES Project.[32] This report of the US-InterPARES contribution, therefore, will simply summarize the results of the research by reviewing how the Task Force answered the research questions stated above. It will then examine the potential impact of the key findings and recommendations of the Appraisal Task Force on current appraisal practice in the United States.

1. How the Appraisal Task Force Answered the Research Questions

What is the influence of digital technology on appraisal?

It seems clear that digital technology has a significant influence on appraisal practice as the Appraisal Task Force has delineated it. At the heart of this influence is the volatility of the digital environment and

[31] Murtha Baca, ed. "Glossary," *Introduction to Metadata: Pathways to Digital Information* (Los Angeles: Getty Research Institute, 2001). Available at: <http://www.getty.edu/research/institute/standards/intrometadata//4_glossary/index.html>.

[32] The complete report of the Appraisal Task Force, see the final report, *Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* available on the InterPARES Website at: <www.interpares.org/book/index.htm>.

the opportunities that exist for significant changes in the records after the appraisal decision has been made. This influence is manifested in the earliest stages of the records' life cycle, as the task force confirmed the widely held view that early appraisal of electronic records is essential in order that knowledge about the technological contexts of the records can be captured. Appraisal of electronic records requires a detailed understanding of the technological contexts in which the records were created and in which they exist at the time they are transferred to the archives. The nature of digital technology also imposes new burdens on the archivist to verify the authenticity of records (because of the potential for the easy corruption of electronic records) and in determining the feasibility of maintaining records over time (because of the technical requirements that would need to be met). Determining feasibility means that the appraisal archivists will need to know how record elements are manifested in the digital environment. They will need to identify the digital components of the records to be preserved. In short, they must be immersed in the technical details of the digital environment. In some case, it may be judged that *because of the nature of the digital environment* some records cannot be preserved, or at least not in authentic form.

The volatile nature of electronic records is the reason the task force introduced the concept of monitoring the appraisal decision for any change to the records before they are transferred to the preserving body as a new element in the process of selecting electronic records for long-term retention. While administrative changes, for example, may have occasioned a review of appraisal decisions in the past, the potential for changes to the records in the digital environment are significant enough to merit regular checking of the validity of appraisal decisions. Minor changes might result in minor tweaking of the appraisal and disposition recommendations. More significant technological changes might require more significant revisions, while wholesale technological changes may require a whole new appraisal of the records.

What is the influence of retrievability, intelligibility, functionality, and research needs on appraisal?

The Appraisal Task Force did not investigate these questions directly, but its report acknowledges that the ability to read and retrieve or present records in a form that does not compromise their identity or integrity is an essential element for the preservation of authentic records. This capability, therefore, should be an essential part of an assessment of the feasibility of preserving electronic records. Some researchers have suggested that the preserver must also replicate the functionality of the system creating the records. We have assumed that replicating the functionality of the system is not necessary if the message the record was meant to communicate is preserved and its identity and integrity evident. However, in any given case, should the capability exist to replicate aspects of the functionality of the originating system, then the appraiser would naturally take that into account.

What are the influences of the medium and the physical form of the record on appraisal?

The InterPARES Authenticity Task Force determined that the medium of the record is in fact part of the technological context of the record and that not all aspects of physical form necessarily need to be reproduced in order to have authentic electronic records. Although clearly an understanding of the technological context is important in the appraisal process, the Task Force judged that these specific questions regarding medium and physical form were no longer appropriate in light of the findings of the Authenticity Task Force.

When in the course of their existence should electronic records be appraised?

The Task Force recognized that records must exist before they can be appraised, but also recognized that it is possible to build records retentions scheduling into the design of electronic record keeping systems.

Until records are actually created in the system and can be examined, however, questions about their authenticity and the feasibility of preserving them cannot adequately be made. Scheduling might be regarded as the first step in the appraisal process, when continuing value alone is judged. This initial step would need to be followed by an assessment of authenticity and determination of feasibility, most likely at the time that a transfer of records to the preserver is anticipated. The need to appraise early is important in enabling the archivist to gather information about the technological context of the records since assembling or reconstructing such information years later is often a difficult task.

Should electronic records be appraised more than once in the course of their existence, and, if so, when?

The ideal scenario, according to the Appraisal Task Force, is that an initial appraisal is made, preferably when records can be seen "live" in the system that generated them. The applicability of that appraisal would then be regularly monitored to take account of changes in the records and their contexts, with the last monitoring being at or near the time of transfer (disposition). Therefore it would be fair to say that electronic records must be appraised more than once in the sense that dynamic digital environment means that the assumptions and judgments of the appraisal must be validated before disposition action is taken. In short, the idea of monitoring is the Task Force's answer to questions that have been raised in the literature about the timing of the appraisal of electronic records.

Who should be responsible for appraising electronic records?

The Task Force worked on the assumption that appraisal is part of the primary responsibilities of the preserver of records; although we recognized that there are different ways in which this responsibility might actually be carried out in a given administrative setting. In appraising records, the preserver has the long-term interests of the records in mind, not just the interests of the creator. Key activities in the appraisal process, such as assessing documentation about the records and their authenticity, are carried out with the understanding that this information will benefit future users of the records, whether from inside or outside the creating body. Considering the feasibility of preservation would seem to require that the preserver is the one making the decision. Otherwise there might be a disconnect between the capabilities of the preserver and an appraisal decision that is not in line with those capabilities.

The model of the appraisal process certainly allows for ample input of the needs of the creating body. Appraising electronic records is a complex process, however, and it requires considerable professional expertise to perform. It seems only logical, therefore, that archivists, the professionals whose primary task is preservation, should have the responsibility of appraising electronic records.

What are the appraisal criteria and methods for authentic electronic records?

The model of the "select electronic records" function displays the Task Force's understanding of the methodology for appraising authentic electronic records. The requirements for assessing authenticity as part of assessing the value of electronic records, and the concepts developed for determining the record elements to be preserved and identifying the digital components to be preserved as part of determining the feasibility of preservation are the only criteria that, in our view, can be established to cover all situations. We did not try to establish criteria governing the assessment of continuing value because such assessments are so sensitive to the entire context in which appraisals are made.

2. Implications for Appraisal Practice in the United States

Despite the volume of literature devoted to the review of existing appraisal practices and to proposing new appraisal approaches, the appraisal of electronic records is not yet a well-developed activity in the American archival practice. The paucity of detailed appraisal reports that the Task Force was able to study is evidence of this fact. While electronic records are covered in records disposition schedules, for example, their peculiar characteristics as electronic records are generally not taken into account. From the appraisal standpoint, they are treated much like paper records and evaluated on the traditional criteria of evidential and informational value. It is hoped that the work of the Appraisal Task Force and the functional model of the selection process will clarify and perhaps demystify the process of appraising electronic records.

The recommendations of the Appraisal Task Force introduce some new criteria to the appraisal process that address the differences in digital records. Most notably, there is the question of factoring in the feasibility of preserving any given set of electronic records into the appraisal decision. Feasibility has certainly been considered in the past in making appraisal decisions, but it has usually been considered strictly from the standpoint of costs—whether the repository has the necessary financial and staff resources to maintain the records. With electronic records, the archivist must consider whether, in fact, it will be technically feasible to preserve the records over time, given current and prospective technical capabilities, as well as the costs involved.

A second crucial factor in the Appraisal Task Force's work is the suggestion that archivists will need to include an assessment of the authenticity of electronic records in making appraisal decisions. In the past, archivists have been able to assume that records created and maintained by their creator in the course of doing business were authentic. The fact that they were relied upon by their creator was evidence enough to regard them as authentic. The modern digital environment, however, poses new sets of problems. Records may no longer exist in the systems in which they were created, and thus they may have undergone changes that could affect their authenticity. While paper files could sit untouched for years in file cabinets with their authenticity intact, digital records cannot survive under such conditions of benign neglect. The appraisal archivist, therefore, should assess the authenticity of the records according to the Benchmark Requirements for Authenticity established by the InterPARES Authenticity Task Force.

A third factor in the Task Force's appraisal work that has implications for the American archival profession is the necessity of a deep knowledge of the technological context in which records are created, and specifically an understanding of the digital components of the records that relate to authenticity. This means that archivists who appraise electronic records must bring a strong technological background to their work. Understanding appraisal theory and practice alone will not suffice, and thus this recommendation will have significant implications for the content of graduate archival education programs.

Finally, the work of the Appraisal Task Force has introduced a new function to the appraisal process, the activity of monitoring the appraisal decision. This concept of monitoring the records is a new feature in the selection process for electronic records compared to that for paper records (and even first-generation electronic records), which are not subject to the volatility of the technological environment that affects electronic records in today's complex distributed environment. It is also a particularly crucial process for electronic records since the appraisal of electronic records is likely to be made at a time well before the records are actually transferred to the custody of a preserver.

C. How do we preserve authentic electronic records?

The goal of the research in this domain was 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.

The research questions for the Preservation Task Force (PTF) were:

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 Preservation Task Force addressed the preservation domain's research goal and research questions. The research design and methodology, research findings, research products and conclusions of the PTF are presented in the *Preservation Task Force Final Report*. The next section of this report describes the Task Force's approach to the research goals of the preservation domain, that 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." The two succeeding sections discuss the research questions and summarize the research results for this domain.

1. The PTF's Model of Preserving Electronic Records

This section describes the Preservation Task Force's analysis of the problem of preserving electronic records. The solution to this problem is a framework for preserving electronic records that archival institutions can use to satisfy the Authenticity Task Force's *Baseline Requirements Supporting the Production of Authentic Electronic Records*. Components of this framework include preservation strategies for specific bodies of records, preservation methods, preservation action plans, and terms and conditions for transfer of electronic records.

The preservation problem

The research objective was to develop a model of the process of preserving electronic records. The Integration Definition For Function Modeling (IDEF0) modeling notation and methodology was used to represent the problem and to represent the results of the analysis of the problem.[33] At the most abstract level, the IDEF0 diagram in figure 1 represents the problem of *Preserving Electronic Records*.[34]

[33] FIPS 183, Integration Definition for Function Modeling (IDEF0), US Department of Commerce, National Institute of Standards and Technology, Dec. 21, 1993. Available at: <http://www.itl.nist.gov/fipspubs/>.

[34] The explanation in this section is for version 5.x of the Preservation Model.

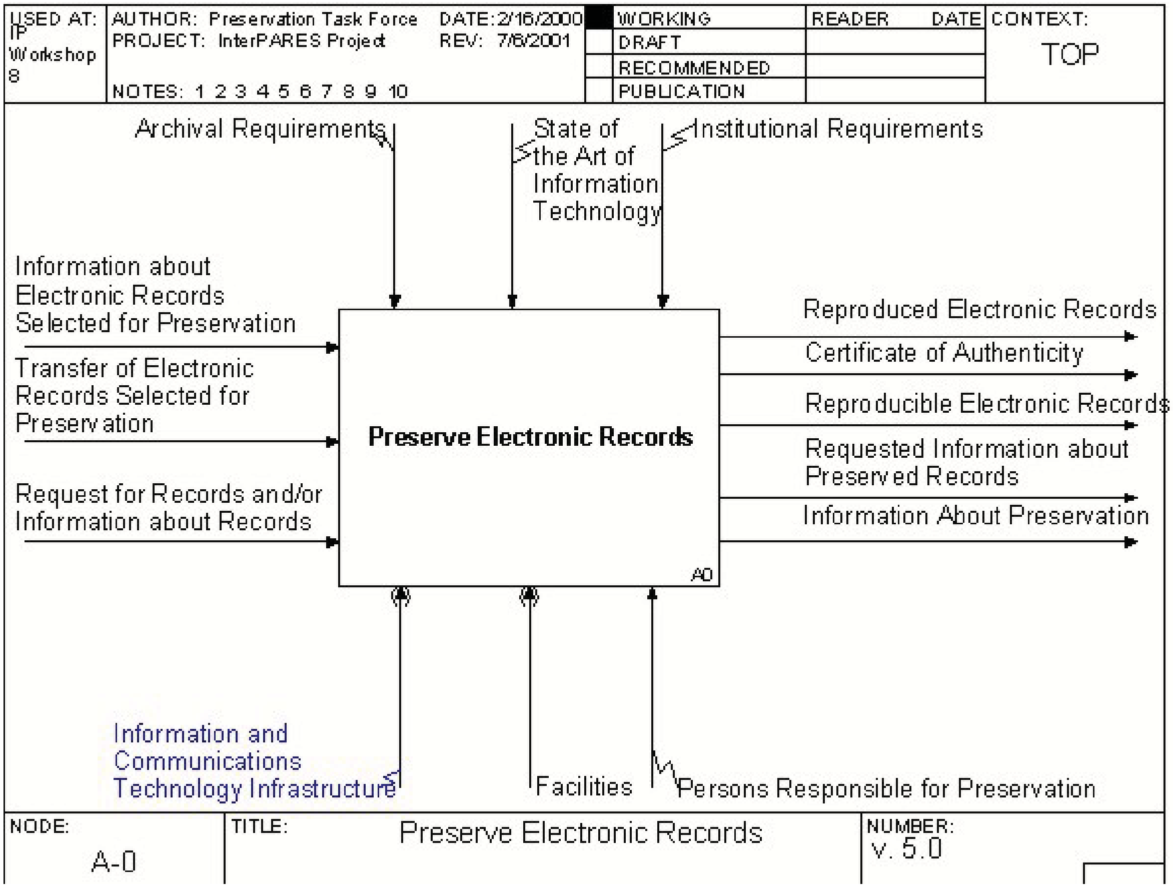

USED AT: IP Workshop 8
AUTHOR: Preservation Task Force
PROJECT: InterPARES Project
DATE: 2/16/2000
REV: 7/6/2001
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
TOP
Archival Requirements
State of the Art of Information Technology
Institutional Requirements
Information about Electronic Records Selected for Preservation
Transfer of Electronic Records Selected for Preservation
Request for Records and/or Information about Records
Preserve Electronic Records
A0
Reproduced Electronic Records
Certificate of Authenticity
Reproducible Electronic Records
Requested Information about Preserved Records
Information About Preservation
Information and Communications Technology Infrastructure
Facilities
Persons Responsible for Preservation
NODE:
A-0
TITLE:
Preserve Electronic Records
NUMBER:
v. 5.0

Fig. 1. IDEF0 Representation of the Preservation Problem.

Given *Information about Electronic Records Selected for Preservation* and *Transfers of Electronic Records*, the goal is to preserve these electronic records so that given a *Request for Records and/or Information about Records*, those records can be reproduced, and information about those records and preservation actions on those records can be provided.

The box in the center of this diagram represents the general activity *Preserve Electronic Records*. The labeled arrows entering the box from the left represent the inputs to the activity. The activity converts the inputs to the outputs, which are shown as labeled arrows leaving the right side of the activity box. The labeled arrows entering the top of the activity box represent controls that regulate the activity, for example, institutional policies govern the preservation of electronic records. Labels on arrows entering the activity box from below represent the physical resources required to perform this activity. They include information and communication technology, people, and facilities.

This problem can then be decomposed into the four subproblems that are shown in Figure 2.

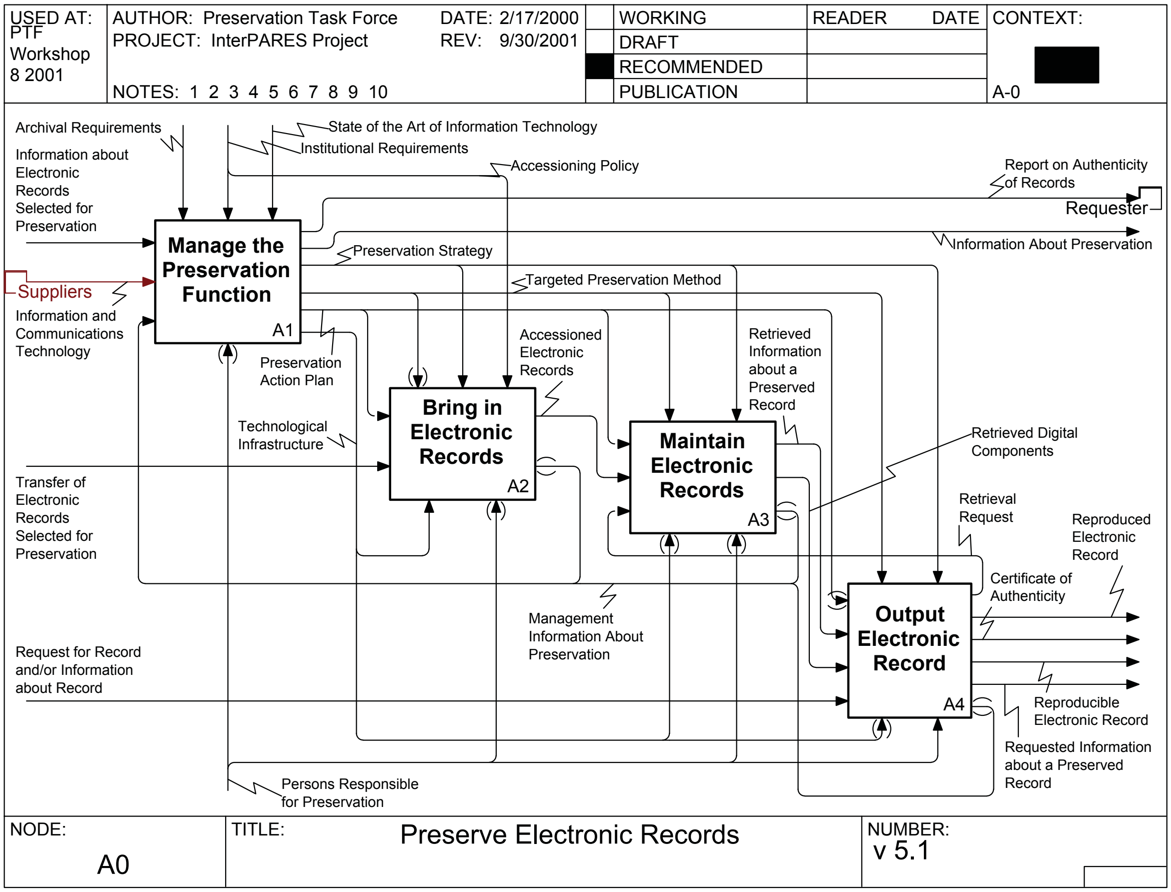

USED AT: PTF Workshop 8 2001
AUTHOR: Preservation Task Force
PROJECT: InterPARES Project
DATE: 2/17/2000
REV: 9/30/2001
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A-0
Archival Requirements
State of the Art of Information Technology
Institutional Requirements
Accessioning Policy
Information about Electronic Records Selected for Preservation
Report on Authenticity of Records
Requester
Manage the Preservation Function
A1
Preservation Strategy
Information About Preservation
Suppliers
Targeted Preservation Method
Information and Communications Technology
Accessioned Electronic Records
Retrieved Information about a Preserved Record
Preservation Action Plan
Bring in Electronic Records
A2
Technological Infrastructure
Maintain Electronic Records
A3
Retrieved Digital Components
Transfer of Electronic Records Selected for Preservation
Retrieval Request
Reproduced Electronic Record
Certificate of Authenticity
Management Information About Preservation
Output Electronic Record
A4
Request for Record and/or Information about Record
Reproducible Electronic Record
Requested Information about a Preserved Record
Persons Responsible for Preservation
NODE:
A0
TITLE:
Preserve Electronic Records
NUMBER:
v 5.1

Fig. 2. Decomposition of the Preservation Problem.

We can view this diagram by working backward from Box A4. The goals of reproducing records and providing information about these records can be achieved if we have the digital components of the records, methods for reproducing the records, and information about the records. The subgoals of digital components of the records and information about the records can be achieved, if we have maintained information about the accessioned digital records, maintained the digital components and have executed planned actions and methods for preserving the records [Box A3]. The subgoal of having accessioned electronic records can be achieved, if we can bring in the transferred electronic records and confirm that they comply with the terms or conditions for transfer or that preservation action plans and methods can be executed to bring them into compliance with the preservation strategy [Box A2]. The subgoals of having preservation action plans and preservation methods can be achieved, if we have information about the electronic records selected for preservation and preservation decisions are made based on archival requirements, the state of the art of information technology, and institutional requirements [Box A1].

Each of the four subproblems shown in figure 2 was analyzed and decomposed into sub-subproblems. For instance, problem A3, Maintain Electronic Records, was decomposed into the three subproblems shown in figure 3.

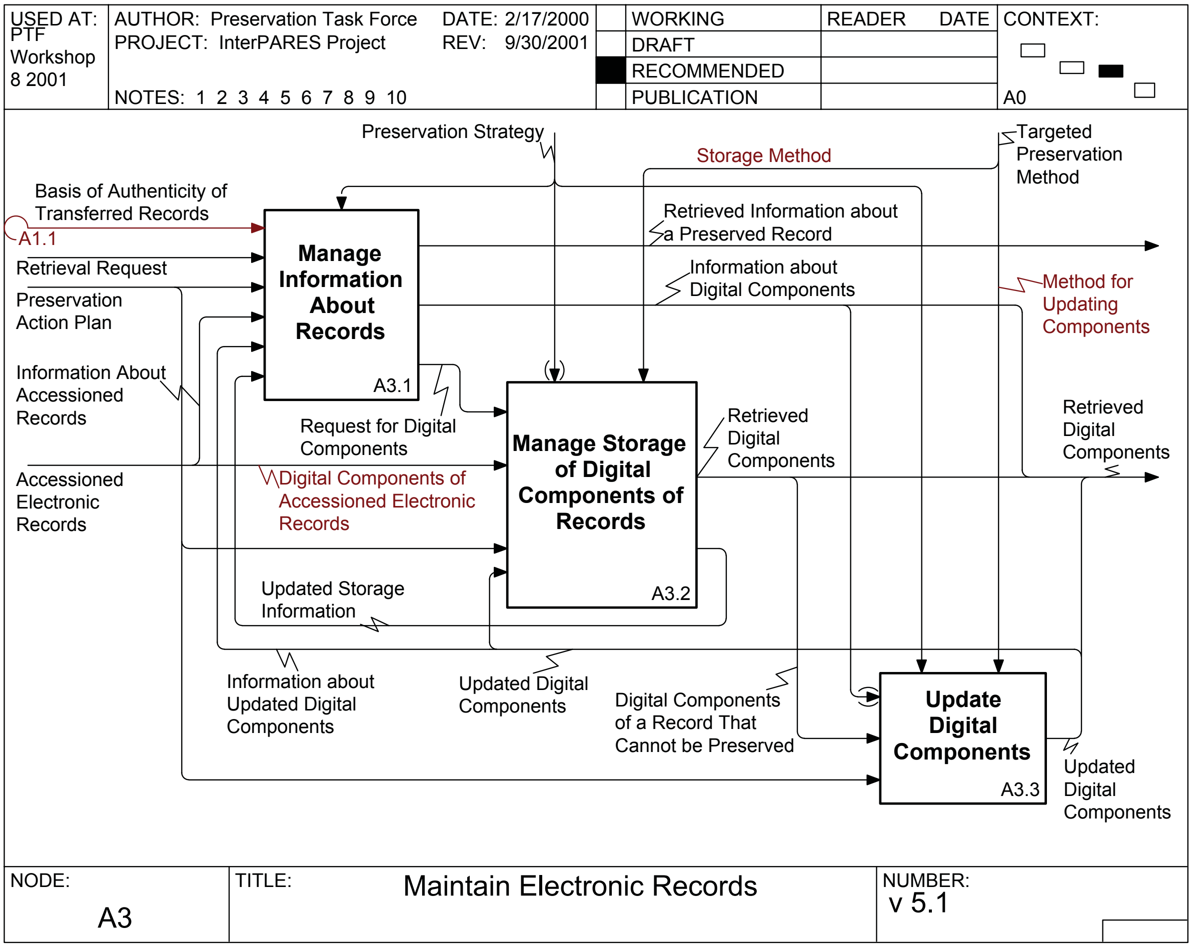

USED AT: PTF Workshop 8 2001
AUTHOR: Preservation Task Force
PROJECT: InterPARES Project
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 2/17/2000
REV: 9/30/2001
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A0
Preservation Strategy
Storage Method
Targeted Preservation Method
Basis of Authenticity of Transferred Records
A1.1
Retrieval Request
Preservation Action Plan
Information About Accessioned Records
Accessioned Electronic Records
Manage Information About Records
A3.1
Retrieved Information about a Preserved Record
Information about Digital Components
Method for Updating Components
Request for Digital Components
Digital Components of Accessioned Electronic Records
Manage Storage of Digital Components of Records
A3.2
Retrieved Digital Components
Retrieved Digital Components
Updated Storage Information
Information about Updated Digital Components
Updated Digital Components
Digital Components of a Record That Cannot be Preserved
Update Digital Components
A3.3
Updated Digital Components
NODE:
A3
TITLE:
Maintain Electronic Records
NUMBER:
v 5.1

Fig. 3. Decomposition of the Problem of Maintaining Electronic Records.

Problem A3.1, *Manage Information about Electronic Records*, can be solved through a data model of the information about the accessioned and updated digital records and a database management system (DBMS) that uses the data model to support storage, update, and retrieval of information. Problem A3.2, *Manage Storage of Digital Components of Records* can be solved with an archival storage system that supports storage and retrieval of the digital components of accessioned and updated electronic records and refreshment of storage media. Problem A3.3, *Update Digital Components*, has as its goal that records be reproducible from their digital components. However, the obsolescence of the file formats of the digital components due to new computer hardware, operating system and application software places the records at risk of not being reproducible. This problem was decomposed into the three alternative subproblems shown in figure 4.

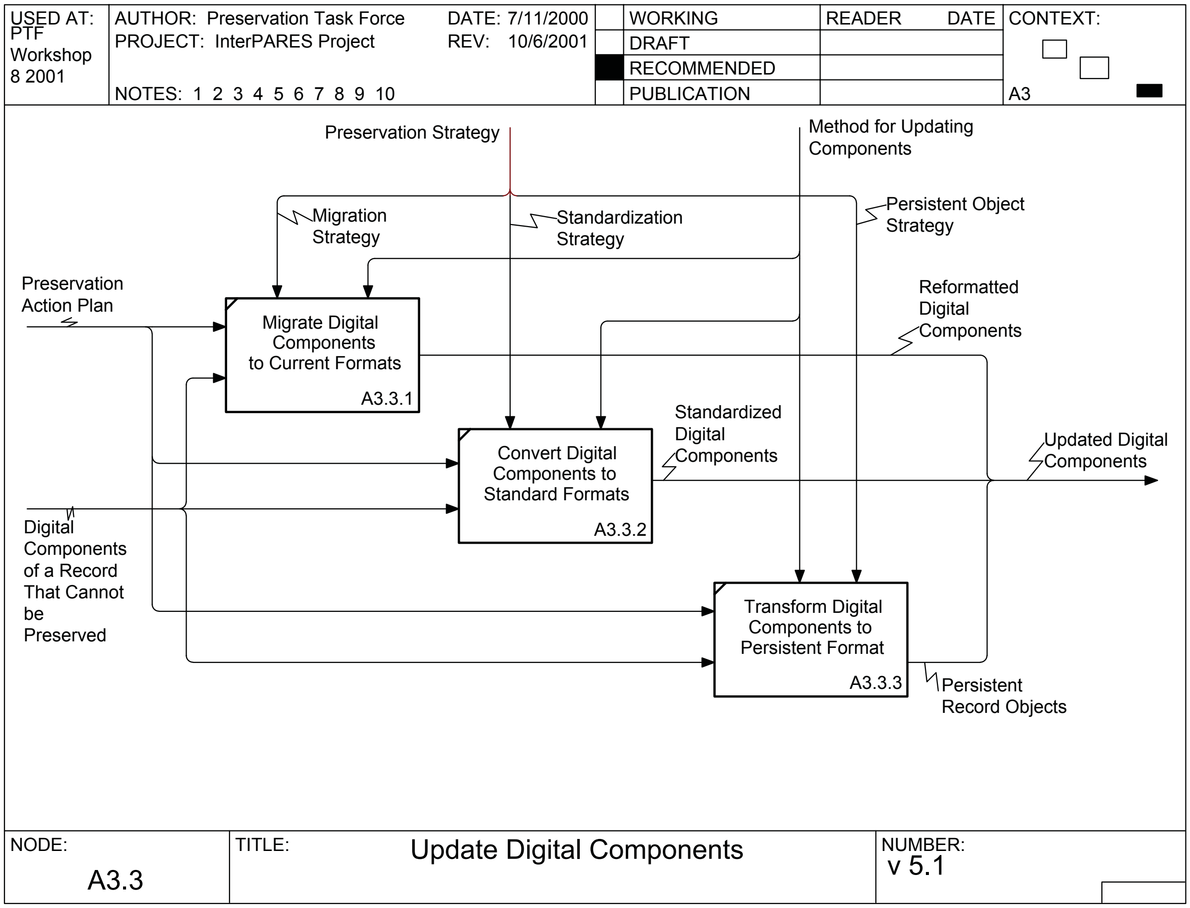

USED AT: PTF Workshop 8 2001
AUTHOR: Preservation Task Force
PROJECT: InterPARES Project
DATE: 7/11/2000
REV: 10/6/2001
NOTES: 1 2 3 4 5 6 7 8 9 10
WORKING
DRAFT
RECOMMENDED
PUBLICATION
READER
DATE
CONTEXT:
A3
Preservation Strategy
Method for Updating Components
Migration Strategy
Standardization Strategy
Persistent Object Strategy
Preservation Action Plan
Migrate Digital Components to Current Formats
A3.3.1
Reformatted Digital Components
Convert Digital Components to Standard Formats
A3.3.2
Standardized Digital Components
Updated Digital Components
Digital Components of a Record That Cannot be Preserved
Transform Digital Components to Persistent Format
A3.3.3
Persistent Record Objects
NODE:
A3.3
TITLE:
Update Digital Components
NUMBER:
v 5.1

Fig. 4. Decomposition of the Problem of Updating Digital Components.

Activity A3.3.1 solves problem A3.3 by using a preservation method that migrates digital components in obsolete formats to current file formats. Activity A3.3.2 solves problem A3.3 by using a preservation method that converts digital components represented in proprietary or obsolete file formats to standard file formats. Activity 3.3.3 solves problem A3.3 by using a preservation method that transforms digital components in a proprietary, obsolete, or standard format into software and hardware independent representations such as XML and XSL-FO.[35]

The process of decomposition is continued until all subproblems have a solution in terms of actions that can be performed by persons or by computer programs. All of the diagrams for version 6.0 of the Preservation Model are included in this report in Appendix D.[36]

[35] For a description of persistent objects and persistent archives, see Arcot Rajasekar, Richard Marciano, and Reagan Moore, "Collection-based Persistent Archives." Available at: <http://www.sdsc.edu/NARA/Publications/OTHER/Persistent/Persistent.html>.

[36] For the complete preservation model, including the activity and arrow definitions, see the final report, *Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* available on the InterPARES Website at: <www.interpares.org/book/index.htm>.

Validating the preservation model with walkthroughs

Can the PTF model of the activities for preserving authentic electronic records be applied in the real world? In other words, is it a valid model of the activities necessary to preserve authentic electronic records?

One of the merits of the InterPARES preservation model is its comprehensiveness and generality. It provides a generic preservation strategy (or framework) that can be used by archival institutions to develop their own preservation strategies depending on their institutional requirements and the specific bodies of records they must preserve. If specific management decisions were included in the model itself, it would compromise its generality. However, it needs to be demonstrated that this general preservation model works in specific cases. This would contribute to the belief on the part of archival institutions that the preservation framework actually works, and would provide them with an example of how to go about applying the framework in their own archival institutions.

A walkthrough of an activity model is one way of reviewing the model in order to validate it or falsify it. Falsification amounts to identifying activities, controls, inputs or outputs of the model that are not coherent or need further specification. If a model is falsified, it should be revised and another attempt should be made to validate it. During InterPARES I, Case Study 26, the New York State Workers' Compensation Board (WCB) Electronic Case Folder System (ECSF), was used in a walkthrough of the Preserve Electronic Records model.[37] The remainder of this section discusses some results of that walkthrough.

The key components of the preservation framework are preservation strategies for specific bodies of records, preservation methods, preservation action plans, and terms and conditions for transfer. An archival institution's preservation strategies for specific bodies of records depend on institutional requirements, the types of electronic records selected for preservation, and choices of preservation methods.

A preservation strategy is a set of preservation action plans. The following is an example of a preservation strategy for a body of records from the case study:

1. Convert data schema that are not represented in SQL to SQL.
2. Convert documents that are not in Multi-page TIFF format to Multi-page TIFF format.

A preservation action plan is a plan specifying preservation actions to be taken in accessioning records or in preserving maintained records. Figure 5 shows an example of a preservation action plan for converting documents that are not in TIFF format to TIFF format.

1. Retrieve digital components for "Claims for Benefits" in Electronic Case Folder System (ECFS) that are ASCII files.
2. Convert the ASCII text files to TIFF multi-page format using TIFFmaker.
3. Store the digital components converted to TIFF format back to archival Storage.
4. Store in the database "on this date the digital components of Claims for Benefits in the ECFS that were in ASCII format were converted to TIFF multi-page format.

Fig. 5. Example of a Preservation Action Plan.

[37] "Walkthrough Applying the 'Preserve Electronic Records' Model," Appendix to *Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project.*

The first instruction triggers an action in the activity Manage Information About Records to retrieve digital components for a specific series of records and a specific class of records that are in ASCII format. The second instruction uses a specific preservation method, a software program called TIFFmaker, to convert those digital components in ASCII format to TIFF multi-page format. The third instruction stores the converted digital components back into the Archival Storage System. The fourth instruction stores a record of the fact that the digital components were converted to TIFF multi-page format. Additional instructions can trigger actions that verify that the converted digital components are reproducible, whether the form and content of the record are preserved, and whether the records remain accessible and usable.

Preservation action plans are implemented with preservation methods. Preservation methods are software. Figure 6 shows some examples of preservation methods.

Preservation Method Description	Examples of Corresponding Software
Check integrity of transferred records	Hash functions (MD5, SHA-1)
Package digital components for storage	tar, untar, JAR
Refresh storage media	tape copy
Maintain information about records and digital components	data base management system (Oracle, Sybase)
Archival Storage	IBM High Performance Storage System, DLT tapes
Reproduce records	TIFF and PDF viewers, X86 emulator
Update components	TIFFmaker, word2pdf, word2XML

Fig. 6. Examples of Preservation Methods.

Preservation methods include generic software for integrity checks, for packaging or archiving many files as one, for copying storage media, for data base management, and for archival storage. They also include specific preservation methods for reproducing records or for converting proprietary formats to standard formats, for example.

The *Terms and Conditions of Transfer* is a specification governing the transfer to the preserver of a body of electronic records selected for preservation. Figure 7 shows the kinds of information in the Terms and Conditions of Transfer with sample values from the case study.

Record creator's name: New York State Workers' Compensation Board
Transfer agent's name: John Doe, Records Manager
Identification of records
Title: Electronic Case Folder System
Description: Series of case files for adjudicating benefits of disabled workers.
Document Types: Claims for Benefits, Employer's reports of accidents
and illness, Correspondence, Medical Reports, Insurance Carrier's Reports
File Format: Multi-page TIFF
Volume: 300,000 open cases
Data structure: Relational Database Schema
Scheduled Transfer Date: To be determined
Medium or channel of transfer: DLT Tape
Technical Conditions for Transfer: MD5 hash code of all files for integrity check,
All documents converted to TIFF Multipage format,
Metadata schema represented in SQL
Information needed to support a presumption of authenticity

Fig. 7. Elements of the Terms and Conditions of Transfer.

The last item in the Terms and Conditions identifies the kinds of evidence that the record creator must provide to support a presumption of authenticity for the transferred records. Figure 8 shows examples of the kinds of information that would be required of the records from the Electronic Case Folder System to determine whether they could be presumed authentic.

Benchmark Authenticity Requirement	**Information identified at the time of appraisal that would be needed to support a presumption of authenticity**
A.1.a Identity of the record	
A.1.a.i Name of author Name of addressee A.1.a.ii Name of action or matter A.1.a.iii Chronological date	Does the metadata associate author's name, addressee, name of action or matter, and chronological date with each document?
A.1.a.iv Expression of Archival Bond	Contents of Case File ordered by document number which is created by FileNet Image Import system when document is imported into the ECFS.
A.1.a.v Indication of attachments	Does metadata indicate whether there are attachments to documents?
A.1.b Integrity of the record	
A.1.b.i Name of Handling Office	Name of Office that uses FileNet to convert documents to TIFF images
A.1.b.ii Name of OPR	NY WCB
A.1.b.iii Indication of types of annotations	Can Document images be annotated?
A.1.b.iv Indication of technical modifications	Paper documents were scanned into document images in TIFF 6 format and maintained on WORM disks.
A.2 Access Privileges	How is access to ECFS Controlled?
A.3 Protective Privileges: Loss and Corruption of Records	Is a backup (or archival copy) of the contents of the WORM disks maintained?
A.4 Protective Privileges: Media and Technology	WORM Disks are guaranteed for over 100 years.
A.5 Establishment of Documentary Forms	Are documentary forms for each of the following types of documents defined? Claims for Benefits, Employers' Reports of Accidents and Illness, Correspondence, Medical Reports, Insurance Carrier's Reports
A.6 Authentication of Records	Is authentication of documents or document images required at any time during the adjudication process, e.g., hearings?
A.7 Identification of Authoritative Record	Are the document images in a case file the authoritative record?
A.8 Removal and Transfer of Relevant Documentation	Is the metadata for case files and documents transferred from the ECFS?

Fig. 8. Information Needed to Assess a Presumption of Authenticity.

The Baseline Requirements

During the walkthrough of the Preservation Model, the PTF examined whether the preservation model satisfies the Authenticity Task Force's Baseline Requirements for Supporting the Production of Authentic

Copies of Electronic Records. It was concluded that Requirement 1 for Controls over Records Transfer, Maintenance and Reproduction is satisfied by:

- activity A1.3.2 for creating Terms and Conditions for Transfer;
- activity A2.2 that compares the transfer with the terms and conditions for transfer;
- activity 2.2.3 that takes the actions needed to preserve the records by carrying out Preservation Action Plans that use preservation methods to bring digital components into compliance with the preservation strategy; and
- activity A4 that reproduces the record from maintained digital components.

Requirement 1a, unbroken custody of the records, is satisfied by activity A1, which plans for preservation of selected records, activity A2, which brings in transferred records, and activity A3, which maintains the Electronic Records. Requirement 1b, security and control procedures are implemented and monitored, is satisfied in part by access control and access privileges of a DBMS and an Archival Storage System. Requirement 1c, that the content of the record remains unchanged after reproduction, is satisfied by activity A1.2.3, selecting methods to apply to preservation objects that preserve their content, and activity A2.3.2, verifying that transferred objects can be preserved and reproduced.

Requirement 2 for documentation of the update/reproduction process and its effects is satisfied by activity A1.2.3, selecting a reproduction method to apply to a class of preservation objects and by activity A1.4, evaluation of preservation.

Requirement 3, that the archival description for a body of records include information about changes to the records since they were first created, is satisfied by activity A3.3, Update Digital Components, and specifically by Preservation Action Plans that document the updates to digital components.

Each of the Baseline Requirements for Supporting the Production of Authentic Copies of Electronic Records is satisfied by some set of activities of the Preservation Framework.

2. Preservation Task Force Research Questions

Answers to each of the five general research questions are described in this section.

1. *What methods, procedures and rules of long-term preservation are in use or being developed?*

In order to answer this question, a survey was conducted of the activities of thirteen institutions and projects in the United States and abroad that employ or are exploring strategies to preserve authentic electronic records. These strategies included the following preservation techniques: migration,[38] emulation,[39] knowledge-based persistent object preservation,[40] bundling,[41] refreshing,[42] digital

[38] The process of moving records from one hardware and/or platform to another.

[39] 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.

[40] "Persistent archives can be characterized by two phases, the archiving of the collection, and the retrieval or instantiation of the collection onto new technology. The 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 create 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." Reagan Moore, et al., "Collection-Based Persistent Digital Archives—Part 2," *D-Lib Magazine* 6 (April 2000).

archaeology,[43] preservation copying, physical preservation, and robotics.[44] A full report on the results of the survey has been published in the *American Archivist*.[45] The survey also considered questions of selection for preservation, staffing configurations, cost modeling, access to preserved records, and policymaking. The survey detected three broad themes. First, traditional definitions of preservation used in the library and archival world seem to be shifting because of the inherently ephemeral nature of electronic materials. 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. Finally, 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.

1A. Which of these meet the conceptual requirements for authenticity identified in Domain I?

None of the methods or techniques identified in this survey can be considered to meet the conceptual requirements for authenticity identified by the Authenticity Task Force and described earlier in this report. Until these methods are further developed and standardized, they cannot be relied upon to ensure the long-term preservation of authentic electronic records.

1B. Which methods of long-term preservation need to be developed?

Migration, emulation, persistent object preservation, and bundling need further research and development. It is still too early in the development of these techniques to fully evaluate them. In addition, the methods of conversion to standard formats and multivalent documents need to be further investigated.

1C. Which of these methods are required or subject to standards, regulation, and guidelines in specific industry or institutional settings?

The Authenticity Task Force mapped the Baseline Requirements against the provisions of DoD 5015.2 Records Management Standard.[46] All of the methods being investigated would be subject to that records management standard. The projects represented in the survey are developing standards and guidelines, and some of the institutions interviewed are waiting to see the results of these projects before committing to a particular strategy. The authors hope to be able to answer question one more fully in subsequent rounds of this research.

[41] Bundling involves taking objects such as Word documents, by using software, creating bundles of documents on an independent platform. It can be seen as another form of emulation.

[42] Periodically moving records from one storage medium to another. It is a preventative measure and, because of rapid media obsolescence, it will be a necessary strategy for some years to come.

[43] " …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." Seamus Ross and Ann Gow refer to this as "*post hoc* rescue" in their *Digital Archeology: Rescuing Neglected and Damaged Data Resources* AK/JISC/NPO Study. (Glasgow: Humanities Advanced Technology and Information Institute [HATII]), University of Glasgow, February 1999, iv.

[44] The use of robots to download electronic documents. Downloading in and of itself does not preserve the records, but only captures them.

[45] Michèle V. Cloonan and Shelby Sanett, "preservation Strategies for Electronic Records: Where We Are Now—Obliquity and Squint?" *American Archivist* 65(Spring/Summer 2002): 70-106.

[46] "Documents Mapping InterPARES Authenticity Requirements against Existing Standards. Available at <http:www.interpares.org/reports.htm>. In this report, see Appendix 14.2: "Mapping of InterPARES Authenticity Requirements Against Provisions of DoD 5015.2 Records Management Standard."

2. What are the procedural methods of authentication for preserved electronic records?

Procedural methods of authentication are those procedures in an archival system that are not included in the Benchmark or Baseline Requirements.

A. In what way can archival description be a method of authentication for electronic records?

Requirement A.1—Expression of Record Attributes and Linkage to Record—of the Benchmark Requirements for Assessing the Authenticity of Electronic Records states:

> The value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records.

The "expression of record attributes and linkage" is a part of archival description of the record. It includes: the names of persons concurring in the formation of the record, the name of action or matter, the dates of creation and transmission, the expression of archival bond, and indication of attachments. It also includes the name of the handling office and name of the office of primary responsibility (if different from handling office), indication of types of annotations added to the record and an indication of technical modifications. Preservation model activity A1.1.6, "Determine basis of authenticity," uses this information and saves it as part of the archival description in an archival database.

Requirement B.3, Archival Description, of the Baseline Requirements states:

> The archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created.

Preservation model activities A2.3.2, "Verify That the Records in the Transfer can be preserved and Reproduced," A2.3.2, "Take Action Needed to Preserve Record," and A3.3.2, "Refresh Storage," are activities in which information about changes the electronic records have undergone since they were first created is saved as a part of the archival description of the records.

Retrieved information about a preserved record (including archival description) is used in activity A4.4, "Present Record," to determine whether a certificate of authenticity can be issued. However, since archival description is only one requirement of the Benchmark Requirements and one requirement of the Baseline Requirements, it is only part of a method for authenticating copies of electronic records.

B. In what way can appraisal and acquisition/accession reports be constructed to allow for the authentication of electronic records?

When the appraisal activity selects electronic records for preservation, the appraisal report or terms and conditions of transfer should indicate to the creator the kinds of information that will be needed in order to assess the authenticity of transferred records. If those kinds of information are not available for the electronic records selected for preservation, the appraisal report or terms and conditions of transfer should also indicate the kinds of information that will be necessary to verify the authenticity of the transferred records.

In the preservation model, the basis of authenticity of transferred records produced by activity A1.1.6, "Determine Basis of Authenticity," and the information about accessioned electronic records produced by activity A2.5, "Accession Records," include information about the transferred records that the preserver has used to assess the presumption of authenticity that can be accorded these records.

C. *What are the procedures for certifying electronic records when they cross technical boundaries (e.g., refreshing, copying, migrating) to preserve their authenticity?*

When electronic files representing the binary encoding of a record are refreshed, that is, copied to a new media, a cyclic redundancy check, as a part of the copying procedure will guarantee the data integrity of the copied files. Similarly, when a copy is made of the original file representing a record, a cyclic redundancy check can be used to guarantee the data integrity of the copy. To guarantee that the visible record reproduced from the file has the documentary form of the original, it is necessary to ensure that the viewer for the file type, reproduces the documentary form of the original, or if not to indicate the type of copy that is produced, e.g., a simple copy that only includes the content of the record and possibly not the intellectual or physical form.

The procedures that are necessary for guaranteeing the authenticity of records when their original file representation must be converted to another file format present the greatest difficulty. The record content and documentary form must be demonstrated to be invariant under file format conversion, and if not invariant, that the content is preserved and that there is an archival record of the conversion and the loss of any of the elements of documentary form.

Almost as difficult as the previous problem is that of migrating software viewers to a current computer platform. If a compiler for the programming language in which the viewer was written does not exist on the new platform, the program for the viewer must be rewritten and tested to endure that not only the content but also the documentary form has been preserved.

Also of concern is the technological obsolescence of the database technology for representing metadata about the preserved records. When archival metadata must be transferred to a new database technology, there is a risk of loss of information about the record(s) that is needed to certify authenticity. Among the procedures that are being investigated to address this issues is the persistent object preservation method, which is used not only to provide a software independent representation of records, but the metadata about the records as well.[47]

D. *Can the method of contemporary diplomatic criticism be extended to records in electronic systems?*

This research question was not in the original list of research questions to be addressed. It arose as a result of the Authenticity Task Force's research addressing the conceptual requirements for presuming the authenticity of electronic records. The method of diplomatic criticism is a procedural method for authenticating paper records, and if it could be extended to electronic records, it might be applicable to transferred or preserved electronic records.

One of the primary purposes of diplomatic criticism is to reach a conclusion as to the authenticity of records.[48] However, this was not the reason that it was used by the Authenticity Task Force to analyze the records from the InterPARES case studies:

> The primary purpose of analyzing the case studies from the perspective of contemporary

[47] Reagan Moore, et al., "Collection-Based Persistent Digital Archives—Part 2."

[48] Luciana Duranti, "Diplomatics: New Uses for an Old Science (Part V)," *Archivaria* 32 (Summer 1991): 6-24.

> archival diplomatics was to consolidate information needed for (1) the drafting of the conceptual requirements for assessing the authenticity of electronic records, and (2) the development of a typology of electronic records based on those requirements.[49]

A diplomatic analysis was conducted for forty-four InterPARES case studies, but the results of this analysis failed to support the hypothesis "that intrinsic and extrinsic elements of documentary form and annotations would play key roles in establishing the identity and demonstrating the integrity of electronic records." The Authenticity Task Force Final Report goes on to say:

> At the same time, however, it was possible to identify certain commonalities in the means used by creators to protect record authenticity from one institution to the next. The diplomatic analysis and the analysis of elements relating to identity and integrity revealed that record creators tend to rely on procedural means for protecting authenticity and to treat it as part of the management of the electronic system as a whole rather than as part of the management of individual records within the system. The commonest means identified were access privileges (including passwords, user IDs, and user profiles), followed by the use of audit trails and back-up procedures.[50]

The traditional method of diplomatic criticism can be represented as a procedure and is shown in Appendix B to this report. The procedure in the appendix does not include considerations of contemporary diplomatic criticism, such as determination of archival bond. In outline, the major sections of the procedure are:

1. Determine extrinsic elements (physical form)
2. Determine intrinsic elements (intellectual form)
3. Determine persons
4. Determine type of act
5. Determine the name of the act
6. Determine relationship between document and the procedure
7. Determine type of document
8. Determine diplomatic description
9. Comments regarding document as a whole (diplomatically authentic or inauthentic document, counterfeit, false document, reliable document, complete document, archived document)

The question arises: Can additional sections be added to the procedure to address the procedural and technological contexts (access privileges, audit trails, record profiles, backup procedures) of the electronic document in order to determine whether an electronic document is a diplomatically authentic or inauthentic document? If one can't extend the procedure in this way, why can't it? A positive answer to this research question would certainly be significant, as it would provide a method for verifying the authenticity of electronic records when there was only a weak presumption of authenticity from an assessment of the presumption of authenticity according to the Benchmark Requirements.

E. *How can the method for assessment of authenticity based on the Benchmark Requirements be more precisely specified and tested so that a preserver could be confident that he could apply the method and be confident in its result?*

This research question was not in the original list of research questions to be addressed by the Preservation Task Force. It arose as a result of the Authenticity Task Force's research addressing the

[49] Authenticity Task Force Report, 11.
[50] Authenticity Task Force Report, 14.

conceptual requirements for presuming the authenticity of electronic records. The method of assessment is a procedural method of assessing the presumption of authenticity of the records, which is to be performed by the preserver. The Authenticity Task Force report says:

> A presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created, handled, and maintained. The evidence supporting the presumption that the creator created and maintained its electronic records authentic is enumerated in the Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records (Requirements Set A). A presumption of authenticity will be based upon the number of requirements that have been met and the degree to which each has been met. The requirements are, therefore, cumulative: the higher the number of satisfied requirements, and the greater the degree to which an individual requirement has been satisfied, the stronger the presumption of authenticity. This is why these requirements are termed 'benchmark' requirements.[51]

The ATF and PTF have not actually tried to assess the authenticity of a creator's electronic records using the Benchmark Requirements. Experiments should be conducted to determine whether the Benchmark Requirements and the method of assessment actually achieve what is intended.

While the method of assessment is expressed in simple terms, there are substantial pitfalls inherent in subjective probability assessment due to psychological biases and common misunderstandings of probabilistic reasoning.[52] Furthermore, the conditional dependencies between requirements and between the evidence needed to conclude that a requirement is met can be quite complex.

- Suppose a preserver has just seen that the name of the addressee was not included in the metadata associated with a record. The fact of that discovery is more available to the preserver than other facts. Thus the missing name seems to be a more prominent feature of the record system than it should be. The judgment that the record system is more likely to contain other such records, based on that fact alone, is likely to be in error.

- Where does uncertainty lie? When assessing the authenticity of electronic records, does the uncertainty concerning their authenticity lie in the preserver, or is it a property of the records system? Does uncertainty come from within the preserver, or is it an intrinsic property of events in the environment? If you opt for the choice that holds that uncertainty is a property of events in the environment, then you are subject to the fallacy of denying uncertainty. You believe you can control it. The answer is that uncertainty is attributable to you. It is due to your incomplete knowledge.

- Suppose that the preserver observes the metadata attributes associated with a record. The metadata should contain the name of the author and the name of the addressee of a record, but not all metadata for records includes the name of the author and the name of the addressee. Which is more probable: The metadata contains both the name of the author and the name of the addressee, or the metadata contains the name of the addressee? If you selected the first option you're incorrect. From elementary probability theory, the probability of a conjunction P(A & B) cannot

[51] Authenticity Task Force Report, 22.

[52] Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgement under Uncertainty: Heuristics and Biases*. (Cambridge: Cambridge University Press, 1982). George Wright and Peter Ayton, eds., *Subjective Probability* (New York: John Wiley & Sons Ltd., 1994).

exceed the probability of either of its constituents, P(A) or P(B). This is the conjunction rule. However, it is often the case that the conjunction is more representative of its class than either of its constituents, or more available in some way, and therefore judgments of its probability are subject to one of the representativeness or availability.

- Overconfidence occurs when accumulating evidence, for example, from case-study material about authentic records from which certain predictions are then made. There is a point in the information-gathering process when predictive accuracy reaches a ceiling. Nevertheless, confidence in one's conclusions continues to rise as more information is received. Towards the end of the information-gathering process, most judges are overconfident about their judgments.

- Two preservers applying the Benchmark Requirements to a record creator's procedures, and having the same evidence can have a different degree of belief as to the presumption of authenticity that should be accorded a record creator's electronic records. One reason this can occur is that they have different preferences with regard to risk. Risk takers will tend to overestimate and risk adverse people will tend to underestimate.

The Bayesian approach to reasoning under uncertainty is one approach to reasoning with degrees of belief while dealing with the complexity of conditional dependencies. Combined with Bayesian Belief Networks, it can also expose and overcome some of the common psychological biases and fallacies in reasoning due to misunderstanding of probability.

Bayesian probability is a formal notation and theory that allows one to reason about beliefs under conditions of uncertainty. If we have observed a specific event, then there is no uncertainty. However, suppose H is the hypothesis:

> All of the 40 million email records of the Executive Office of the President that were transferred to the National Archives are authentic.

Since no one will examine each of these records at the time of transfer, nobody can state with certainty whether or not the statement H is true. Different people may have different beliefs in the statement depending on their specific knowledge of factors that might affect its likelihood.

A person's subjective belief in a statement H will depend on some body of knowledge K. This can be represented as the conditional probability P(H|K) that a hypothesis H is true (e.g., that a requirement is met) given available evidence or knowledge K. The expression P(H|K) is a measure of a person's belief in the truth of H warranted by the K.[53]

The definition of the conditional probability of A given that B is true or known is the joint probability of A and B divided by the probability of B.

$$P(A|B) = P(A, B)/P(B)$$

It follows as a theorem (known as Bayes rule) that

$$P(A|B) = P(B|A)\ P(A) / P(B)$$

[53] Colin Howson and Peter Urbach, *Scientific Reasoning: The Bayesian Approach*, 2d ed. (Chicago: Open Court, 1993).

Bayes rule can be thought of as a means of updating ones belief about a hypothesis A in light of new evidence B. Specifically, one's posterior belief P(A|B) is calculated by multiplying the prior belief P(A) by the likelihood P(B|A) that B will occur if A is true.

In those cases where P(A|B) = P(B), A and B are said to be *independent*. If P(A|B,C) = P(A|C), A and B are said to be *conditionally independent given C*.

A Bayesian Belief Network (BBN) is a graphical notation with an associated set of probability tables. The network (or graph) consists of nodes and arcs representing conditional dependencies $P(A_1|A_2, \ldots A_n)$. The key feature of BBNs is that they enable one to model conditional dependencies of variables and to reason using degrees of belief. BBN's provide an intuitive visual representation that can aid in clarifying implicit assumptions made by an expert. With BBNs, it is possible to articulate expert beliefs about the dependencies between different variables. BBNs can also expose and overcome some of the common psychological biases and fallacies in reasoning due to misunderstanding of probability. However, the most important use of BBNs is in revising probabilities in light of actual observations of events. Furthermore, there are software tools that implement the algorithms for propagating the results of new evidence through the BBN, as well as providing a graphical user interface to draw the graphs and fill in the probability tables.

Bayesian Probability Theory and Bayesian Belief Networks are a promising approach to more precisely specifying the conditional dependencies of the Benchmark Requirements and providing a rigorous method for assessment of authenticity based on those requirements.

3. *What are the technical methods of authentication for preserved electronic records?*

The Authenticity Task Force states in its final report:

> Because of the ongoing developments in the area of authentication technologies, it is necessary to clarify the distinction between *authenticity*—which is the focus of InterPARES—and *authentication*. In common usage, *authentication* is understood as a declaration of a record's authenticity at a specific point in time by a juridical person entrusted with the authority to make such a declaration. It takes the form of an authoritative statement (which may be in the form of words or symbols) that is added to or inserted in the record attesting that the record is authentic.
>
> Digital signatures are an example of an authentication technology that has been developed to address the need to secure electronic communication across open networks such as the Internet. Digital signatures, which identify the sender of a data object and verify that it has not been altered in transmission, can support the authentication of electronic records, but they are not sufficient to establish the identity and demonstrate the integrity of an electronic record over the long term. Further research is needed to determine the specific impact of digital signatures on the long-term preservation of authentic electronic records.[54]

During the project, progress was made in developing procedures that use authentication technologies to ensure the authenticity of electronic records *over time*.[55] The Java ARchive (JAR) format is a platform-independent file format that aggregates many files into one. JAR was developed so that Java applets and their components could be bundled into a single file (package) and quickly downloaded to a browser in an http transaction. The Java application launcher can launch one of the files, e.g., a class with a method,

[54] Authenticity Task Force Final Report, 2.

[55] W. E. Underwood, Steps Toward a Logical Theory of Record Integrity and Authenticity, US-InterPARES Report, February 2002.

in the package. A JAR provides the capability to verify the origin of the components in the JAR so that only programs authored by persons or organizations trusted by the user will be executed. JAR is an open industry standard.[56]

The JAR file format was adapted to store digital records. A procedure for storing digital records and metadata in JARs was developed, and a procedure for retrieving records from JARS and verifying their authenticity was developed. Figure 9 shows an example of the directory structure of a JAR. The META-INF(ormation) directory contains three files. The digital files corresponding to digital records stored in the JAR follow the META-INF directory.

```
META-INF/MANIFEST.MF
META-INF/SIGNATURE.SF
META-INF/SIGNATURE.DSA
wp/corr/file1.wp5
wp/corr/file2.wp5
lotus/schedule.wks
lotus/budget.wks
photo/image1.jpg
photo/image2.gif
```

Fig. 9. Directory Structure of package.jar

Figure 10 shows an example of a manifest file MANIFEST.MF. The manifest file consists of a set of path/file names of files and annotations of these files.[57] The annotations corresponding to a path/filename are called a section of the manifest. The message digests in the manifest file are hash codes created from the files and used to check their integrity.

```
Manifest-Version: 1.0
Organization: Executive Office of the President
Organizational-Unit: Legislative Affairs, Office of
Name-of-record creator: “Richard Breeden”
Series-title: “Richard Breeden's Files”

Name: Chronological Correspondence/file1.wp5
SHA1-Digest: TD1GZt8G11dXY2p40lSZPc5Rj64=
File-format: wp5.1
Document-type: letter
Name-of-author: Breeden, Richard
Name-of-addressee: Kristol, W; Kolb, C
Chronological-date: 01/12/92
Archival-date: 01/12/92

Name: …
```

Fig. 10. Manifest File

[56] Sun Microsystems, Inc., Java™ 2 SDK, Standard Edition Documentation, Version 1.3.1, 2001

[57] The manifest file of a JAR represents attributes and values in the form “header: value”.

Figure 11 shows the contents of the signature file (SIGNATURE.SF). The file includes the message digest (hash code) for the entire manifest file. It also contains digest values created from sections of the manifest file.

```
Signature-Version: 1.0
SHA1-Digest-manifest: “hlyS+K9T7DyHtZrtl+LxvqgaMYM=“
Created-By: Signature File JDK 1.2
Name: wp/corr/file1.wp5
SHA1-Digest: r58H40lDL39d6a2tU6T38Letz64=
```

Fig. 11. Signature File

The file SIGNATURE.DSA is associated with the signature file with the same file name, but has a different file extension (DSA). This file stores the digital signature of the corresponding signature file and an X.509 certificate for the public key of the signature.

A procedure was developed for storing the records of a records creator in JARS, for retrieving those records, and for verifying their authenticity. Another procedure was developed that uses JARs for transferring the records of a records creator to a central archives where their authenticity can be verified. A third procedure was developed for storing archival records in JARs and verifying their authenticity when retrieved.

The concepts of record, record series and archival integrity, and of record and record series authenticity were expressed as axioms or definitions in a logical language. This formal theory attempts to express the diplomatic and archival concepts involved in authenticating a document, a record and a record series. This theory and axioms of belief and communications security were used to prove the correctness of the procedures for maintaining the integrity and authenticity of electronic records stored in JARs.[58]

Among the assumptions of the theorems is that there are no preservation transformations on the records. If there are conversions of digital record format to current or standard file formats, demonstration of record integrity will require demonstrating that the preservation transformations preserve the content and essential elements of documentary form.

4. *What are the principles and criteria for media and storage management that are required for the preservation of authentic electronic records?*

Storage technology spans a broad spectrum of price, performance, and capacity. Fast access and rapid transfer rates require the use of hard-disk drives or RAIDs (Redundant Array of Independent Disks), but for capacity, magnetic tape is still the mechanism of choice. It is both inexpensive (when measured in terms of $/GB) and volume efficient (GB/m^3). Data that is not accessed often is an ideal candidate for magnetic tape residency.[59]

Data integrity is the property whereby data has not been altered in an unauthorized manner since the time it was created, transmitted, or stored by an authorized source. The digital format for media should include checksums, or error detecting codes/error correction codes to provide protection against non-malicious or accidental threats to data integrity, such as media deterioration or transmission errors. In contrast to checksums, data integrity mechanisms based on hash functions are designed to preclude undetectable

[58] For details, see William Underwood, “Steps Toward a Logical Theory of Record Integrity and Authenticity.”

[59] P. C. Hariharan. *Media*. Report prepared for US InterPARES Research Team Meeting, Washington D.C., June 20-21, 2000. Available at: <http://is.gseis.ucla.edu/us-interpares/>.

intentional modification. Hash functions are used to ensure data integrity as follows. The hash value of a file X is computed at time T_1. The integrity of the hash-value is protected in some manner such as separate storage from the data. At a subsequent time T_2, the following test is conducted to determine whether X' is the same as the original file, that is, whether the file has been altered. The hash-value of X' is computed and compared to the protected hash-value of X. If they are equal, then X and X' are equal, and thus the file has not been altered.

Hierarchical Storage Management (HSM) provides seamless access to data while minimizing cost of storage. This is accomplished by employing a hierarchy of storage devices (see figure 12) with the most-expensive ones at the top of the pyramid providing record access and high transfer rates. Policy-based rules regularly migrate data sets to less expensive, slower media. Location transparency ensures that users do not need to be aware of the exact storage media unit on which the data set resides when it is retrieved.

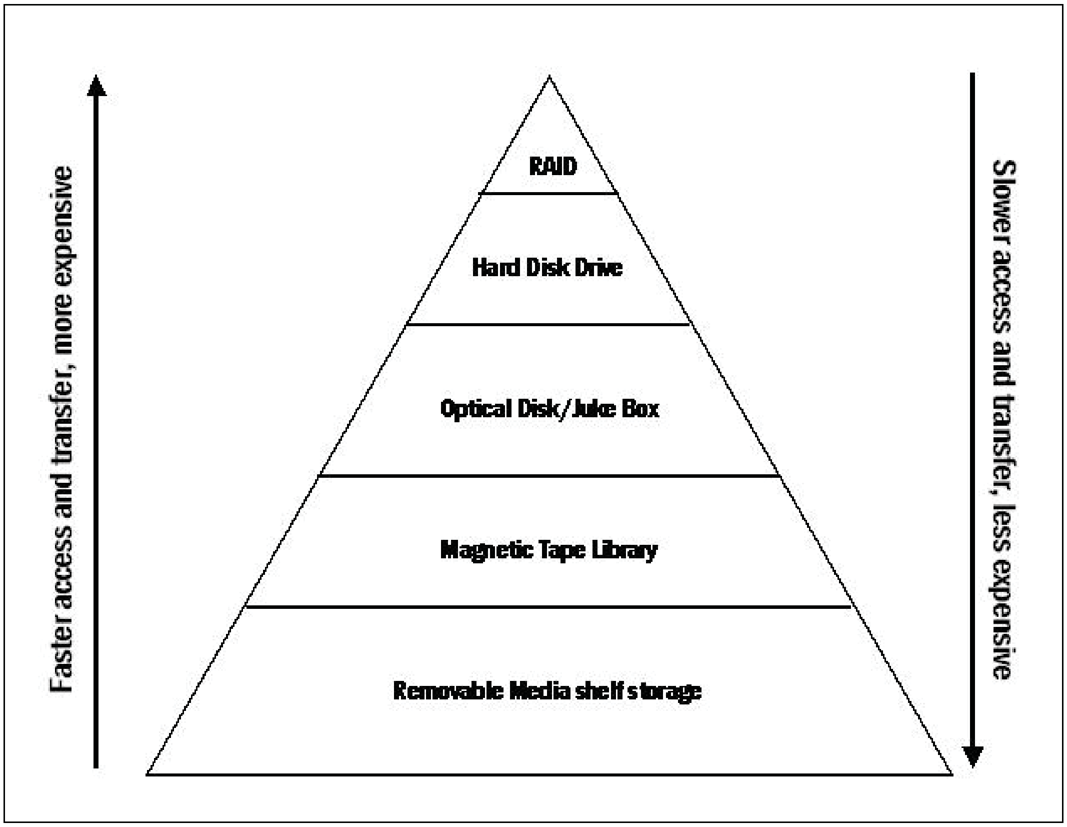

Faster access and transfer, more expensive
RAID
Hard Disk Drive
Optical Disk/Juke Box
Magnetic Tape Library
Removable Media shelf storage
Slower access and transfer, less expensive

Figure 12. Storage Pyramid

5. What are the responsibilities for the long-term preservation of authentic electronic records?

This question is answered by reviewing the PTF Preservation Model to identify roles that need to be assigned to persons and responsibilities for preserving authentic electronic records that must be assigned to roles. First, one identifies those activities at the lowest level of decomposition that have personnel as mechanisms. Second, the outputs of each activity represent the goals of that activity, and the description of the transformation of inputs to outputs represents the sequence of actions that must be performed to achieve the goals. If the activity is not to be automated, then the procedures corresponding to that activity are the responsibility of archival personnel. If the activity is partially automated, then part of the procedures corresponding to that activity is the responsibility of archival personnel. Each procedure is a responsibility assigned to some archival role.

Activity	Role	Responsibility
A1.1	Preservation Manager	Determine preservation requirements for records selected for preservation.
A1.1.1	“	Identify types of properties of records selected from preservation that must be preserved.
A1.1.2	“	Determine how records selected for preservation are composed from digital components.
A1.1.3	“	Determine how records selected for preservation are arranged.
A1.1.4	“	Determine how archival bonds are expressed.
A1.1.5	“	Synthesis requirements for preservation of the body of records selected for preservation.
A.1.1.6	“	Determine basis for certifying authenticity.
A1.2	“	Select preservation technologies.
A1.2.1	“	Identify preservation options.
A1.2.2	“	Evaluate preservation options.
A1.2.3	“	Select preservation methods.
A1.2.4	“	Acquire capability to apply selected preservation methods.
A1.3	“	Specify preservation strategies and actions.
A1.3.1	“	Articulate preservation strategy for a body of records selected for preservation.
A1.3.2	“	Plan for implementing preservation strategy
A1.3.3	“	Assess the strategy and plan
A1.3.4	“	Evaluate execution of preservation
A2.3	Preservation Archivist	Examine transferred electronic records
A2.3.1	“	Map records and digital components within transferred material
A2.3.2	“	Verify that the records in the transfer can be preserved and reproduced.
A2.3.3	“	Take action needed to preserve the record
A3.2.4	“	Correct storage problems

Figure 13. Responsibilities for the Long-Term Preservation of Electronic Records.[60]

This use of an IDEF0 model to identify responsibilities of human personnel is an innovative extension of the IDEF0 modeling methodology.

3. Conclusions and Future Research

The goal of research in the preservation domain was to identify and develop the procedures and resources required for the implementation of the conceptual requirements and the criteria identified in the first two research domains. The PTF analyzed the preservation problem and decomposed it into subproblems

[60] The responsibilities corresponding to activities and roles shown in figure 13 are derived from version 5.1 of the Preservation Model.

(activities). The activities at the lowest level of decomposition can be performed by people or automated procedures, and thus represent a solution to the preservation problem. The PTF model for the “Manage Preservation Function” provides a framework for preserving electronic records. Within that framework, a variety of preservation strategies can be developed by archival institutions that are dependent on the characteristics of the selected, transferred and accessioned records, institutional requirements, and the current and changing state of information technology.

The walkthrough of the PTF preservation model helped refine and partially validated the model. During InterPARES 2, additional walkthroughs should be conducted using case studies with other types of electronic records. This will ensure that the model can be realized in the real world, that is to say, that the model is sound. The kinds of information created, maintained, and used in preserving electronic records were also identified during the walkthrough of the case study. This information can be used during InterPARES 2 to create a data model for the preservation process.

It was also demonstrated that the preservation model provides a framework for implementing archival procedures that satisfy the Baseline Requirements and implement the assessment of the presumption of authenticity of the creator's records. The Terms and Conditions of Transfer identify the kinds of evidence that the record creator must provide to support a presumption of authenticity for the transferred records. Thus the general goal of this research domain was achieved.

In addition to the research questions that the PTF began with, two new questions emerged. The first is “Can the method of contemporary diplomatic criticism be extended to records in electronic systems?” A positive answer to this research question would certainly be significant, as it would provide a method for verifying the authenticity of electronic records apart from an assessment of the presumption of authenticity according to the Benchmark Requirements.

The second new research question derives from the fact that the ATF and PTF have not yet tried to assess the authenticity of a creator's electronic records using the Benchmark Requirements. Experiments should be conducted to determine whether the Benchmark Requirements and the method of assessment actually achieve what is intended.

The ATF's proposed method of assessment is based on the preserver's degree of belief that the Benchmark Requirements are satisfied. Difficulties arise in subjective assessments because of psychological biases and common misunderstandings of probabilistic reasoning. Furthermore, the conditional dependencies between the Benchmark Requirements, the evidence needed and the knowledge required to conclude that a requirement is met can be quite complex.

The Bayesian approach to reasoning under uncertainty is one approach to reasoning with degrees of belief while dealing with the complexity of conditional dependencies. Combined with Bayesian Belief Networks, it can also expose and overcome some of the common psychological biases and fallacies in reasoning due to misunderstanding of probability. This suggests the research question: Can the method for assessment of authenticity based on the Benchmark Requirements be more precisely specified and tested using Bayesian Probability and Bayesian Belief Networks so that a preserver could be confident in the result of an assessment?

The final report of the Preservation Task Force includes four recommendations that are important to summarize here. First, analysts and institutions should use the “Preserve Electronic Records” model as a framework for developing solutions to the challenges of preserving electronic records. Second, use of the model should be based on an understanding of the particular characteristics of electronic records and what those characteristics mean for preserving these records. Six foundation concepts are described in

"How to Preserve Authentic Electronic Records," a document prepared by the PTF and included as an appendix to the International Team Report. The six foundation concepts are:

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,

Central 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.

The third recommendation in the PTF report is that solutions to the preservation of specific bodies of electronic records should be inherently dynamic. This is so for two different reasons. The specific properties of the records brought into the archives will change over time. Thus the preservation system must be capable of being expanded, adapted, or modified to accommodate these changes to records and the new ways of organizing, accessing, and presenting them. The second reason has to do with the fact that the ultimate reason for keeping electronic records is to make the available to persons who need to use them. The continuing evolution of information technology will impact the availability of authentic electronic records. Therefore, 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.

The fourth and final recommendation is really a recommendation to build on the work of InterPARES in several ways. For example, more work is needed to analyze the data and information requirements for executing the processes defined in the preservation model. As suggested earlier, more tests of the model need to be carried out to both validate it and enrich it. Additional research in developing methods for the analysis and categorization of the documentary forms of records should be conducted, since establishing the documentary form of records is an element in the Benchmark Requirements and assessing the impact of reproduction processes on the form of records is part of the Baseline Requirements. The report also recognized the potential for InterPARES to contribute to the enrichment of the Open Archival Information System (OAIS),[61] particularly through adding an archival understanding of authenticity to the OAIS model. Several of these research themes can be pursued in the second phase of the InterPARES Project.

[61] CCSDS 650.0-B-1: *Reference Model for an Open Archival Information System* (*OAIS*), Blue Book. Issue 1, January 2002. This Recommendation has been adoped as ISO 14721:2002.

III. Translating Research Outcomes into Practice

One aspect which is often conspicuously lacking in large research projects, especially those that have a strong theoretical basis, is consideration of how the findings and products of the research can be effectively operationalized and integrated into the wider world of legislation, practice, and professional education. This section discusses ways in which US-InterPARES researchers have addressed and are continuing to address the translation of research outcomes into practice.

A. What policies, strategies, and standards will protect the authenticity of electronic records over time?

The InterPARES Strategy Task Force was charged with creating an intellectual framework to help archivists, records managers, lawmakers, and others to develop policies, strategies and standards that integrate InterPARES recommendations into their diverse national contexts and organizational situations. The Task Force's perspective in this work was that distinctions would need to be drawn between international, national, and organizational policies, strategies, and standards based upon recognition that each cultural, legal and organizational environment has its own needs and contingencies. The key overarching strategy of the Task Force, therefore, was to ensure that these disparate national policies, strategies, and standards be inspired by the same overarching set of principles.

The Strategy Task Force developed a framework and set of principles to be used in formulating new standards and evaluating existing policy and standards environments. These were subsequently contextualized by the national research teams participating in InterPARES into the national domains of participating members.[1] As part of this effort, US-InterPARES researchers evaluated the current United States legislative and standards environment to assess the extent to which it addresses the recommendations emanating out of the InterPARES Domain Task Forces. This evaluation is discussed in this section. A more detailed survey of United States law is included in Appendix C.

A defining characteristic of the United States' record-keeping context is its heterogeneity. Individual sectors, most notably the federal government and certain industries, are relatively homogeneous and controlled by specific records legislation, industry standards and regulations, and accreditation or licensing requirements; however, the United States consciously has not embraced a national information policy that requires uniform approaches to electronic records management. This is the result of several factors, including the common law juridical base of the United States; the distributed structure of the federal, state, and local systems of government; the traditional autonomies of the academic and religious sectors; and the increasing emphasis on enterprise and digital government facilitated through the implementation of state-of-the-art technology.

Records produced by federal, state, and local agencies are generally retained by a designated preserving agency in accordance with statutory requirements and associated records retention policies.[2] Private sector records (e.g., those of businesses, religious organizations, museums, and private universities) are either retained by the creating agency in accordance with statutory, regulatory, and organizational records

[1] The complete report of the Strategy Task Force is contained in *Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* available on the InterPARES Website at: <http://www.interpares.org/book/index.htm>.

[2] For example, responsibility for the archival management of federal records resides with the U.S. National Archives and Records Administration, while in some states responsibility for local government records from clusters of counties is devolved to archival repositories at universities situated in those counties. The records of individual state universities are usually managed by their archives according to state and institutional records management requirements.

retention requirements or are, usually after a period of time, deposited in or donated to archival or manuscript repositories for the purposes of historical research.

The range of practices employed by the archival community reflects this pluralism in recordkeeping requirements and responsibilities (e.g., both life-cycle and continuum approaches and custodial and non-custodial management of electronic records). This pluralism is also reflected in the differences in the resources available for archival management and in the level of expertise of institutional archivists in electronic records management. As a result, electronic records preservation policies have been developed and implemented primarily on an ad hoc basis as needed within individual organizational contexts. Few systematic electronic records programs exist outside of the National Archives and Records Administration (NARA) and certain state archives, despite the National Historical Publications and Records Commission's funding emphasis on developing electronic records management principles, practices, and programs, and a measurable increase in the number of graduate education programs offering coursework in electronic records management.

This report reviews existing national and international legislation and standards that have implications for electronic records management, and specifically the preservation of authentic electronic records within one or more sectors in the United States. The review indicates the extent to which these standards or pieces of legislation address the principles for preservation polices, standards, or strategies identified by InterPARES. It offers some commentary about the current situation and how it might be improved. For example, the principles could be used by different sectors and interest groups to augment, qualify, or tighten the legislation and standards as sources of warrant; to suggest new legislation, standards, and policies; or to recognize and nurture best practices through professional education. It should be noted, however, that the preferred approach in many non-governmental sectors has been to enhance professional education in order to inculcate best practices, and to work in concert with professional archival associations to develop and support professional standards, rather than to respond to externally imposed standards.

1. International and National Legislation and Standards

There are few national standards that relate specifically to the authenticity and long-term preservation of electronic records, although the corporate sector has warrant in the form of the ISO 15489 Records Management Standard as well as regulatory requirement, such as those of the Food and Drug Administration that affect approval of new products. The following legislation and standards are referred to in the subsequent analysis:

Digital Millennium Copyright Act (DMCA) 17 USC Section 101 et seq. (title IV amending §108, §112, §114, chapter 7 and chapter 8, title 17, United States Code)

> President Bill Clinton signed the Digital Millennium Copyright Act into law on 28 October 1998. The legislation implements the 1996 World International Treaty and two World International Property Organization (WIPO) treaties: the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty. Key provisions of the *DMCA* concern the circumvention of copyright protection systems, fair use in a digital environment, and online service provider (OSP) liability (including details on safe harbours, damages, and "notice and takedown" practices). [3]

[3] Though not specifically referenced in the chart that follows, the DMCA and other copyright legislation can pose significant challenges to the design and implementation of the long-term preservation of electronic records. Long-term preservation of authentic electronic and digital records may require copying that is outside the scope of current copyright protection. Future InterPARES policy research will include the examination of existing copyright and intellectual property regimes and their relationship to proposed digital preservation strategies and implementation.

Electronic Communications Privacy Act (ECPA), Title 18 of the United States Code, Section 2701 et seq.

> This was adopted to address the legal privacy issues that were evolving with the growing use of computers and other innovations in electronic communications. The *ECPA* updated legislation passed in 1968 that had been designed to clarify what constitutes invasion of privacy when electronic surveillance is involved. The *ECPA* extended the privacy protection outlined in the earlier legislation to apply to radio paging devices, electronic mail, cellular telephones, private communication carriers, and computer transmissions.

E-Sign (Electronic Signatures in Global and National Commerce Act), Title 15 of the United States Code, Section 7001 et seq. See also state digital signature legislation—for example, the *Electronic Uniform Electronic Transactions Act ("UETA").*

> *E-Sign* (Public Law 106–229), enacted on 30 June 2000, eliminates legal barriers to the use of electronic technology to sign and form contracts, collect and store documents, and send and receive notices and disclosures. *E-Sign* applies broadly to federal and state statutes and regulations governing private sector (including business-to-business and business-to-consumer) activities. *E-Sign* authorizes the substitution of electronic notices for paper notices, including most, but not all, types of consumer notices. *E-Sign* also includes a number of important protections to ensure that consumers can receive, keep, and use electronic notices provided to them.

Government Paperwork Elimination Act, Title 44 of the United States Code, Section 3504 note *(GPEA).*

> The *GPEA*, enacted on 21 October 1998, requires that by October 2003, all executive branch agencies are to provide for the use and acceptance of electronic signatures in communications with the public, where practicable.

Federal Records Act, Title 44 of the United States Code, Chapters 21, 29, 31, 33, and NARA regulations, Title 36 of the Code of Federal Regulations (CFR) Part 1234 (Electronic Records Management). Records Management by Federal Agencies. See also state records and information management legislation.

Presidential Records Act, Title 44 of the United States Code, Sections 2201 et seq.

> The *PRA,* enacted in 1978, changed the legal ownership of the official records of the President from private to public.

Federal Rules of Evidence (FRE) govern admissibility of evidence in administrative proceedings in federal courts. The general requirements address relevance, authentication, and hearsay aspects of evidence. While the *FRE* do not apply to suits in state courts, the rules of many states have been closely modeled on these provisions.

Food and Drug Administration (FDA) Title 21 of the CFR Part 11: Electronic Records; Electronic Signatures.

> The Electronic Records and Electronic Signature Rule (21 CFR Part 11) was established by the U.S. Food and Drug Administration and put into effect on 20 August 1997. The rule defines the requirements for controlling electronic records and submitting documentation in electronic form, and the criteria for approved electronic signatures. It is designed to assist laboratories in the areas of improved data management, simplified regulatory compliance, and increased data security and integrity. The final rule relating to this title provides criteria

under which *FDA* will consider electronic records to be equivalent to paper records, and electronic signatures equivalent to traditional handwritten signatures. Part 11 (21 CFR Part 11) applies to any paper records required by statute or agency regulations and supersedes any existing paper record requirements by providing that electronic records may be used in lieu of paper records. Electronic signatures that meet the requirements of the rule will be considered to be equivalent to full handwritten signatures, initials, and other general signings required by agency regulations. Section 11.2 provides that records may be maintained in electronic form and electronic signatures may be used in lieu of traditional signatures. Records and signatures submitted to the agency may be presented in an electronic form provided the requirements of Part 11 are met and the records have been identified in a public docket as the type of submission the agency accepts in an electronic form. Unless records are identified in this docket as appropriate for electronic submission, only paper records will be regarded as official submissions.

Freedom of Information Act (FOIA), Title 5 of the United States Code, section 552, as amended by the Electronic Freedom of Information Act Amendments of 1996. <http://www.usdoj.gov/04foia/>. See also state *FOIA* legislation.

Provides that any person has the right to request access to federal agency records or information, and that agencies shall make reasonable efforts to search for records in electronic formats and provide to requesters records in any format (including electronic). All agencies of the United States government are required to disclose records upon receiving a written request for them, except for those records that are protected from disclosure by the nine exemptions and three exclusions of the *FOIA*. This right of access is enforceable in court. The federal *FOIA* does not, however, provide access to records held by state or local government agencies, or by private businesses or individuals. All states have their own statutes governing public access to state and local records; state agencies should be consulted for further information about them.

International Organization for Standardization (ISO) 15489 Records Management Standards

ISO 15489 focuses on the business principles behind records management and how organizations establish a framework to enable a comprehensive records management program. The new standard identifies key issues involved in retaining the information and making it available in a usable and reliable way. ISO 15489 is aimed at individuals responsible for setting policies, standards, and guidelines for information management within organizations. These include records managers, archivists, librarians, knowledge management professionals, database managers, and business administrators within organizations who are responsible for the oversight of record-keeping processes.

Reference Model for an Open Archival Information System (OAIS)—CCSDS

The *Reference Model* drafted by the National Aeronautics and Space Administration (NASA), is an ISO technical recommendation relating to the preservation of digital information by digital archives and their producers and consumers. The *Referencing Model (OAIS)*, White Book, Issue 4, Don Sawyer / NASA and Lou Reich / CSC. Among the components of OAIS are the following: the reference model identifies a minimum set of responsibilities for an archive to claim it is an OAIS; establishes common terms and concepts for comparing implementations, but does not specify a specific implementation; provides detailed models of both archival functions and archival information; and discusses OAIS information migration and interoperability among OAISs.

United States Department of Defense (DoD) 5015.2 Records Management Standard

> The DoD standard was created for use by agencies of the United States government. The standard is designed and expressed in terms of compliance with U.S. laws and regulations. The purpose of the DoD standard is to prescribe mandatory baseline functional requirements, and to identify non-mandatory features deemed desirable for Records Management Application (RMA) Software. Within the context of the U.S. government, 5015.2 is a procurement standard requiring government agencies to purchase RMAs that are compliant with at least the minimum specifications.

The authenticity of archival records (i.e., that those records are indeed what they purport to be) is an aspect largely ignored in the legal context of the United States.[4] Issues of system integrity and data reliability for active records are more common areas of concern when evidentiary value is an issue. The notion of authenticity of records in the sense used in diplomatics is largely alien to the corporate and legal records management communities in the United States.

2. State and Local Context

State and local authorities have also not systematically addressed the preservation of authentic electronic records, although legislation that parallels federal legislation often exists at the state level. The National Historical Publications and Records Commission has funded several initiatives to address electronic preservation issues at the state and local government level (e.g., in Minnesota, Mississippi, and the City of Philadelphia); and at individual academic institutions (e.g., Indiana University) that have sought to develop model solutions and policies in the absence of more specific legislation and standards.

The InterPARES research outcomes (i.e., principles, requirements, and models) will have an impact in the United States only to the extent that the authenticity and preservation of electronic records are considered universally pressing issues by archival, records management, and legal professionals. Although system integrity and access to reliable information are critical components of an effective electronic documentary record, the ability to establish and document the continued authenticity of electronic records is crucial to implementing an effective preservation plan.

The following table presents the principles that should govern any preservation policy, standard, or strategy for ensuring the long-term preservation of authentic electronic records. The principles are drawn from the report of the InterPARES Strategy Task Force. Each principle is paired with references to relevant legislation or standards that affect the application of the principle in the U.S. environment. Commentary on the application of the relevant legislation or standards or on the absence of any such legislation or standards is provided as appropriate.

[4] The InterPARES Glossary defines an *authentic record* as, “a record that is what it purports to be and that is free from tampering or corruption.”

U.S. commentary on principles that should govern preservation policies, standards, and strategies for the long-term preservation of authentic electronic records

Principle	**U.S. References and Commentary**
Address records specifically rather than digital objects generally; that is, [any preservation policy, strategy, or standard] should address documents made or received and set aside in the course of a practical activity.	The U.S. National Archives is specifically charged with the archival management of the records of the federal government as defined by the *Federal Records Act* 44 U.S.C. Chapter 31. See also *National Archives and Records Administration* 44 U.S.C. Chapter 21; *Records Management by the Archivist of the United States,* 44 U.S.C. Chapter 29; *Disposal of Records* 44 U.S.C. Chapter 33;[5] *Coordinator of Federal Information Policy* 44 U.S.C. Chapter 35; *Information Technology Management Reform Act (ITMRA)* 40 U.S.C. Section 1401 et seq.; *Paperwork Reduction Act* 44 U.S.C. Chapter 35; *Administrative Procedure Act,* 5 U.S.C. Chapter 5, the *Freedom of Information Act,* 5 U.S.C. Section 552,[6] the *Privacy Act,* 5 U.S.C. Section 552a.[7] Electronic Records: Electronic Signatures 21 CFR 11.1 (c).[8] See also examples from State records and information law such as Ohio[9] and New Mexico.[10] In common practice, however, records and digital objects are typically undifferentiated in litigation and in business activities. Moreover, in non-federal repositories such a those of universities and local historical societies, records, manuscripts, and sometimes other library or artifact collections are often co-administered without explicitly addressing the distinctive preservation and authenticity needs of electronic records. The Open Archival Information System (OAIS) Reference Model refers only to *information objects* and not to records. See also *FRE*, *DoD*, and *ISO*.

[5] The Federal Records Act, 44 U.S.C. 3301, defines federal records to include all books, papers, maps, photographs, machine-readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States government under federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions procedures, operations, or other activities of the government or because of the informational value of data in them.

[6] The Freedom of Information Act, 5 U.S.C. 552(f)(2), defines record to include "any information that would be an agency record subject to the requirements of this section when maintained by an agency in any format, including an electronic format." In general, the definition of "agency record" under FOIA is broader than the definition of "record" under the Federal Records Act.

[7] The *Privacy Act,* 5 U.S.C. 552a(a)(4) defines *record* to mean "any item, collection, or grouping of information about an individual that is maintained by an agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph."

[8] The FDA Electronic Records: Electronic Signatures, 21 CFR Part 11 defines electronic records as "any combination of text, graphics, data, audio, pictorial, or other information representation in digital form that is created, modified, maintained, archived, retrieved, or distributed by a computer system."

[9] Ohio Administrative Code Section 149.01.1 (G) defines records to include "any document, device, or item, regardless of physical form or characteristic, created or received by or coming under the jurisdiction of any public office of the state or its political subdivisions, which serves to document the organization, functions, policies, decisions, procedures, operations, or other activities of the office."

Focus on authentic electronic records.	*E-Sign (Electronic Signatures in Global and National Commerce Act*) digital signature legislation addresses aspects of the reliability of electronic records; however, digital signatures only provide a means for assuring authenticity *in time*, and not preserving authenticity *over time*. See also *FRE*, *DoD*, and *ISO*.
Recognize and provide for the fact that authenticity is most at risk when records are transmitted across space (i.e., when sent between persons, systems, or applications) or time (i.e., either when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced).	This is not currently addressed in the U.S. context.
Recognize that preservation of authentic electronic records is a continuous process that begins with the process of records creation and whose purpose is to transmit authentic records across time and space.	*U.S. Department of Defense (DoD) Directive 5015.2,* Records Management Program Directive, March 2000. ISO 15489 Records Management Standard. OAIS Reference Model.
Base on the concept of trust in records keeping and record preservation and specifically on the concepts of a trusted record keeping system and the role of the preserver as a trusted custodian.	The Joint Interoperability Test Command's (JITC) software testing program for 5015.2-STD, Design Criteria Standard for Electronic Records Management Software Applications, November 1997 <http://jitc.fhu.disa.mil/recmgt/>.
Predicate on the understanding that it is not possible to preserve an electronic record as a stored physical object; it is only possible to preserve the ability to reproduce the record.	This is a new construct in the U.S. context.
Recognize that the physical and intellectual components of an electronic record do not necessarily coincide and that the concept of digital component is distinct from the concept of element of documentary form.	Archival practice in the United States has not traditionally examined elements of documentary form in establishing record-keeping protocols and requirements. The concept of a digital component is a new construct in the U.S. context.

Ohio Administrative Code Section 1306.21 (1) defines "the minimum requirements of creation, maintenance, and security of electronic records and electronic signatures; (2) If electronic records must be signed by electronic means, all of the following: (a) the type of electronic signature required; (b) the manner and format in which the signature must be affixed to the electronic record; (c) the identity of, or criteria that must be met by, any third party used by the person filing a document to facilitate the process. (3) Control processes and procedures as appropriate to ensure adequate preservation, disposition, integrity, security, confidentiality, and auditability of electronic records; (4) Any other required attributes for electronic records that are specified for corresponding non-electronic records or reasonably necessary under the circumstances. (B) (1) The department of administrative services may adopt rules in accordance with section 111.15 of the Revised Code to ensure consistency and interoperability among state agencies with regard to electronic transactions, electronic signatures, and security procedures."

[10] New Mexico Title 1 General Government Administration Chapter 13 Public Records Part 70 Section 7 (A) defines records as "information preserved by any technique in any medium, now known, or later developed, that can be recognized by ordinary human sensory capabilities either directly or with the aid of technology."

Specify the requirements that a copy of a record should satisfy to be considered equivalent to an original.	See 17 USC 101 et seq.
Integrate records appraisal in the continuous process of preservation.	U.S. archivists are increasingly involved in the design of record-keeping systems as well as scheduling electronic records. Both of these activities provide opportunities to integrate appraisal and description requirements into electronic record keeping at a pre-archival stage. U.S. archivists need increased education and training in how best to effect this integration in their own institutional contexts.
Explicitly state that the entire process of preservation must be thoroughly documented as a primary means for protecting and assessing authenticity over the long term.	This principle underlies *FRE*. As with the previous principle, this is in part an issue of ensuring best practices through increased archival education and training in electronic records management. However, there is no current metadata framework that U.S. archivists could impose on record-keeping system design, or require of record-keeping procedures.
Explicitly recognize that the traditional principle that all records relied upon in the usual and ordinary course of business can be presumed to be authentic needs to be supplemented in the case of electronic records by evidence that the electronic records have not been inappropriately altered.	Not explicitly recognized in the U.S. context.
Recognize that the preserver is concerned with both the assessment and the maintenance of the authenticity of electronic records. The assessment of the authenticity of electronic records takes place before the records are transferred to the custody of the preserver as part of the process of appraisal, while the maintenance of the authenticity of copies of electronic records takes place once they have been transferred to the preserver's custody as part of the process of long-term preservation.	This could in part be ensured through increased archival education and training in electronic records management and the development of professional best practices. Since individual U.S. archival repositories in the United States espouse both life cycle and continuum models of archival management, archivists need to understand how to apply this principle within their own institutional contexts.
Draw a clear distinction between the preservation of the authenticity of records and the authentication of a record.	This distinction is not currently made in the U.S. context.

B. Who should care about the authenticity of permanent electronic records?

In the United States today, electronic infrastructures for government, commerce, and research are being built that are predicated on records that are created and exchanged electronically. It is crucial, therefore, that systems and strategies exist for ensuring that these records remain accessible and trustworthy for as long as they are needed. Three constituencies have a major stake in addressing the development of such systems and strategies:

- *Those who have already recognized the critical nature of this issue and are invested in addressing it*, e.g., Federal funding agencies such as the National Historical Publications and Records Commission and the National Science Foundation; national institutions such as the National Archives and Records Administration and the Library of Congress, industry organizations such as the Collaborative Electronic Laboratory Notebook Systems Association (CENSA)

- *Those who can help to solve the challenges associated with the issue*, e.g., technology researchers, software developers, policy developers, educators

- *Unwitting stakeholders—those who should be concerned but are as yet not fully aware of the scope of the issue*, e.g., records creators and records user communities.

As discussed in the previous section, the InterPARES Strategy Task Force Report outlines an intellectual framework for the articulation of international, national, and organizational policies, strategies, and standards for the long-term preservation of authentic electronic records. That report acknowledges that any policy, strategy, or standard is created within, and must be articulated in a manner suitable for a particular environment. In order to ensure that research outcomes such as those of InterPARES 1 have the best potential for influencing and guiding the stakeholder constituencies outlined above, it is key that archival requirements for authentic electronic records are framed within particular environmental constraints. This section discusses how an understanding of the contexts, perspectives, and needs of the stakeholders was addressed by US-InterPARES through the development of an extensive Stakeholder Analysis of some of the most significant U.S. stakeholder groups.[11] Because the U.S. context is very complex and not yet fully defined, stakeholder relationships in the analysis were examined according to the scope of InterPARES 1 research, digital preservation research in general, and specific environments in which electronic records preservation might be implemented.

Currently, there is no standard approach to identifying and analyzing stakeholders. Stakeholder analysis often seeks only to identify some major stakeholders, using very general definitions of what a stakeholder might be. While the definition of a stakeholder as anyone who can affect or be affected by an organization's actions is commonly accepted, it is important to separate apparent stakeholders from actual stakeholders. Utilizing a theory of stakeholder identification and salience used in the management literature, the Stakeholder Analysis identified and grouped parties concerned with the long-term preservation of electronic records.[12] Each group was analyzed to determine why it is, or should be concerned about digital preservation. Each stakeholder group was then analyzed to determine its significance as a stakeholder in terms of *legitimacy*, *urgency*, and *power*:

[11] The Stakeholder Analysis is available on the US-InterPARES Website at: <http://is.gseis.ucla.edu/us-interpares>.

[12] Ronald K. Mitchell, Bradley R. Agle, and Donna J. Wood, "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts," *Academy of Management Review* 22 (October 1997): 853–86.

> *Legitimacy* is a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions.
>
> *Urgency* is the degree to which stakeholder claims call for immediate attention. A stakeholder group has an urgent interest in the research when its needs are of a time-sensitive nature, and when they are important or critical to its mission.
>
> *Power* is the probability that one individual or group within a relationship is in a position to carry out its own will despite resistance, bearing in mind that powerful stakeholders may be able to exert influence which will affect a project either negatively or positively. Power is the ability to control which decisions are made, and to facilitate the implementation of these decisions. Power may be coercive, based on the use of force or the threat of force; utilitarian, relying on material persuasion or incentives; or normative, involving more symbolic influence. When evaluating power, it is important to consider whether each stakeholder group has the resources—whether mental energy, time, expertise, or technology—to manage electronic records preservation.

Each stakeholder will have one or more of these qualities to varying degrees. In general, the more qualities, and the more of each quality, a group has, the more seriously its needs and interests should be considered. It is important to remember, however, that a legitimate stakeholder with urgent interests may have no power to influence the project, while not every powerful stakeholder has a legitimate, urgent claim.

The Stakeholder Analysis examined eleven primary stakeholder groups: archivists, records managers, government, IT professionals, lawyers, for-profit industry, educators, librarians, artists, scientists, and research funding organizations. Each one of the groups has some sort of stake in the InterPARES research and holds some degree of power to influence it. These groups are the most obvious of all potential stakeholders. The groups are listed in order of the group's significance according to the analysis of its legitimacy, urgency, and power in respect to the electronic records preservation research of the InterPARES Project. The Stakeholder Analysis also examined the potential interests of several other groups, including auditors, charitable and not-for-profit organizations, religious institutions, professional organizations, standards bodies, and the public.

This Stakeholder Analysis is not the final statement on the needs and concerns of stakeholders of electronic records preservation research projects like InterPARES. Much of the analysis is based on secondary sources or assumptions, and future research projects would greatly benefit from considering stakeholder needs in more detail throughout the project, perhaps by conducting focus group research with a group of experts representing the concerns and interests of the various stakeholder groups.

C. How do we prepare records creators, records managers, and archivists to assist in the creation and preservation of authentic electronic records?

While standards, policies, and best practice guidelines are all key tools for changing practices and technologies related to records creation and preservation, another important aspect is the education of all stakeholders, from archivists through to the general public, about the need to create and preserve trustworthy electronic records. There are at least three roles that such education can play:

- Building awareness, and changing the behaviors, outlook and expectations of society in general and of stakeholder groups in particular;
- Training records creators and preservers to implement standards, policies, and best practice guidelines; and,
- Creating new generations of archival educators and researchers expert in electronic records management and preservation.

US-InterPARES has sought to address these components not only through the Stakeholder Analysis discussed earlier, but also through the development of educational workshops and model archival curricula and the incorporation of doctoral students as junior researchers into the work of the project. Several workshops have been developed that will be made available online on the US-InterPARES Website and will continue to be modified and updated as the work of InterPARES 2 progresses. These include workshops on the diplomatic method, activity modeling as an analytical tool, and a modular workshop on all the work of InterPARES that can be subdivided by each domain area into explanations and outcomes (that is, Authenticity, Appraisal, Preservation, and Strategy). US-InterPARES researchers used a variety of research methods during the course of the project, and we felt that a greater understanding of these methods in the archival community would help advance the research basis of the profession. Therefore research team members have also worked on development of a model curriculum on archival research methods and research design. The Method of Diplomatic Criticism included in Appendix B is an outline of how to employ one of these research methods.

In the course of InterPARES 1, six doctoral students participated from the Department of Information Studies, University of California, Los Angeles, designing research instruments, collating and analyzing data, and presenting results.[13] At the conclusion of InterPARES 1, one doctoral dissertation had also been completed that was developed out of the Authenticity case studies of student records systems.[14]

Finally, recognizing that InterPARES is a complex research project, we produced *InterPARES Interpreted*. This short pamphlet serves as a simple guide to the project. It conveys the premises and basic findings of the research in language readily accessible to individuals who are not experts in the field of

[13] A full list of presentations made by US-InterPARES researchers, including research assistants, is on the U.S. team website at: <http://is.gseis.ucla.edu/us-interpares>.

[14] Eun G. Park, "Developing a Framework for Authenticity Requirements in University Student Records Systems: An Exploratory Study" (Ph.D. diss., University of California, Los Angeles, 2002).

electronic records. This group might include archivists and records managers or members of any of the many stakeholder groups who have a vested interest in solutions to the problem of preserving authentic electronic records over time.[15]

[15] The text of this pamphlet is available on the U.S. team website. Printed copies are available from the U.S. InterPARES staff at the School of Information Science and Policy, University at Albany, State University of New York.

IV. Conclusion and Areas for Further Research

The work conducted in each InterPARES research domain—authenticity, appraisal, preservation, and strategies—significantly advanced our understanding of the complex issues involved in the long-term preservation of authentic electronic records. The work of the Authenticity Task Force was pivotal to the research project in that it underscored the importance of the concept of the authenticity of records, explicated what that concept means, and detailed the ways in which authenticity is so endangered in today's recordkeeping environments. The Authenticity Task Force's case studies of electronic record systems elucidated the hybrid nature of most contemporary electronic recordkeeping systems and demonstrated the absence of explicit measures to ensure authenticity. Methodologically, the Task Force's findings underscored the importance of both "top-down" approaches such as diplomatic analysis and "bottom-up" approaches such as systems and functional analyses. The Task Force learned the value of analyzing electronic records from the diplomatic perspective of the individual record, from the perspective of archival science that looks at record aggregates, and from the perspective of system analysis. It finally determined that an overall systems approach, that takes into account the range of contexts of the recordkeeping environment—juridical-administrative, provenancial, procedural, documentary, and technological—is required.

The *Benchmark Requirements Supporting the Presumption of Authenticity of Electronic Records* developed by the Authenticity Task Force provide a clear set of measures for records creators and archivists to apply in determining the degree to which any given set of records can be regarded as authentic. Similarly, the *Baseline Requirements Supporting the Production of Authentic Copies of Electronic Records* provide measures for those charged with preserving authentic electronic records to adhere to as they make copies of records in their custody available to users.

The development of activity models by both the Appraisal and the Preservation Task Forces provide a clear delineation of archival practices on which data models for use in software development, metadata models for ensuring adequate process documentation and archival description, and information policy can be formulated and evaluated The Appraisal Task Force's "Select Electronic Records" model identifies and clarifies the steps involved in the appraisal of electronic records. In addition to the traditional practice of basing appraisal decisions on judgments of continuing value, it recommends that appraisal of electronic records needs to take into account an assessment of the authenticity of the records and the feasibility of their preservation, specifically the feasibility of their preservation from a technological standpoint. It recommends early appraisal of electronic records and regular monitoring of appraisal decisions to ensure that changes to the records and their contexts over time have not negatively affected their identity or integrity or the ability to preserve them.

The Preservation Task Force produced a model of the process of preserving electronic records that identifies the procedures and resources needed to implement the requirements and criteria set forth by the Authenticity and Appraisal Task Forces. The "Preserve Electronic Records" model does not prescribe a particular computer system or recommend specific technological tools or methods for preservation. It does, however, provide a detailed and coherent framework for analyzing the problem of preserving such records and for guiding the choice and evaluation of various technological options and specific strategies for ensuring the continuing availability of authentic copies of records over time and over different generations of technology.

All of the research findings on the preservation of electronic records will be of little use, however, unless ways can be found to apply those findings in the real world. The work of the InterPARES Strategy Task Force, therefore, was to devise an intellectual framework that would support the development of policies, strategies, and standards facilitating the long-term preservation of authentic electronic records. To this end it developed a set of principles and criteria that should govern the development of any records preservation policy, strategy, or standard. At part of this work, the US-InterPARES research team also prepared a detailed analysis of existing laws and regulations that are directly relevant to archivists who are working to preserve electronic records in the United States.

The exploration of any set of research questions inevitably begets new questions, even as the research produces substantial results. It is not surprising, therefore, that the InterPARES Project has identified numerous areas where further research is needed. The Authenticity Task Force has raised questions about the possibility of further developing an analytical framework integrating aspects of contemporary diplomatics and archival theory that would address both the individual document and record aggregates and identify the role of different contexts in relation to both individual records and record aggregates. The question of developing a typology of electronic records based upon individual creators and the acts, procedures, or functions they carry out was judged to be the most fruitful approach to this problem, but one that was beyond the scope of the current project. Questions also remain about how specific technologies might support the implementation of the Benchmark and Baseline Requirements in specific recordkeeping environments, as well as the more general question of how the principles developed can be applied to other kinds of digital objects, such as records generated for cultural, as opposed to administrative, purposes,

Further research is also needed in analyzing the data and information requirements for creating a more seamless and detailed representation of the inherently connected functions of appraisal and preservation. Additional testing of the models should be carried out in order to validate and enrich them. There is also the possibility of using the InterPARES research work to enrich the OAIS Reference Model by incorporating into it the important concept of the authenticity of the information objects being preserved.

More focus needs to be placed in the area of implementation, translating the research outcomes into practice. For example, it would be valuable to develop a collaborative applied research environment in which InterPARES conceptual requirements and activity models could inform the development of functional tools, such as those being developed as part of the San Diego Supercomputer Center Archival Workbench Project. A publication that translates the results of the project into a methodology that people can adapt to their own particular institutional contexts would be useful. Further work in the policy area is needed to differentiate the multivariate needs of the various stakeholders in the United States. The requirements on government, private corporations, and nonprofit organizations may be significantly different in some instances, and individuals working in those different sectors need to understand the requirements that they must meet. Thus other policy-related products, such as risk-management assessment tools, refined cost-benefit analyses and cost models, and more educational products will be essential in advancing an understanding of the issues that InterPARES has addressed.

Many of these lines of inquiry will be pursued in the second InterPARES Project (InterPARES 2), which began this year. InterPARES 2 will address problems involved in the creation, maintenance, and preservation of authentic records created in emerging experiential, dynamic, and interactive systems. In addition to an expanded examination of long-term authenticity issues, this second phase of InterPARES will address issues relating to the reliability and accuracy of electronic records across the whole life cycle, from creation to permanent preservation. The research will be looking particularly at records created in these newer systems as the result of the activities in the artistic, scientific, and governmental sectors.

Appendix A. Template for Analysis

Documentary Form

Definition: The rules of representation according to which the content of a record, its administrative and documentary context, and its authority are communicated. Documentary form possesses both extrinsic and intrinsic elements.

Extrinsic Elements of Documentary Form

Definition: The elements of a record that constitute its external appearance.

1. Presentation Features

Definition: A set of perceivable features (graphic, aural, visual) generated by means of encoding and program instructions, and capable, when used individually or in combination, to present a message to our senses.

Overall Presentation

Definition: The record's overall information configuration, i.e., the manner in which the content is presented to the senses.

Text

Definition: Words, numbers, or symbols.

Graphic

Definition: A representation of an object or outline of a figure, plan, or sketch by means of lines. A representation of an object formed by drawing.

Image

Definition: An artificial imitation or representation of the external form of any object, or an optical appearance or counterpart of an object, such as is produced by rays of light, refracted as through a lens, or falling on a surface after passing through a small aperture. A subset of image is moving images which are visual images, with or without sound, that, when viewed, present the illusion of motion.

Sound

Definition: Aural representation of words, music, or any other manifestation of sound.

Combination of More than One of the Above

Specific Presentation Features

Definition: Specific aspects of the record's formal presentation that are necessary for it to achieve the purpose for which it was created.

Examples: Specific presentation features might include but are not limited to the following:

- special layouts
- deliberately employed type fonts
- deliberately employed colours
- hyperlinks
- graphic indication of attachments
- sample rate of sound files
- resolution of image files
- scales of maps

2. Electronic Signature

Definition: A digital mark having the function of a signature in, attached to, or logically associated with a record, which is used by a signatory to take responsibility or give consent to the content of that record, and which may be used to verify its authenticity.

Electronic Seal

Definition: Specific electronic means of authenticating a record or ensuring that it is only opened by the intended addressee. It is a distinct type of electronic signature.
Example: An electronic seal might include but is not limited to the following:

- digital signature, i.e., an electronic signature based on public key cryptography.

Authentication Certificate of Trusted Third Party (TTP)
Definition: An attestation issued by a TTP for the purpose of authenticating the ownership and characteristics of a public key. Such attestation appears in conjunction with the digital signature of the author of a record and is itself digitally signed by the TTP.

3. Digital Time-Stamp Issued by a Trusted Third Party (TTP)

Definition: An attestation by a TTP that a record was received at a particular point in time.

4. Special Signs

Definition: Symbol that identify one or more of the persons involved in the compilation, receipt, or execution of the record.
Examples: Special signs might include but are not limited to the following:

- digital watermarks
- organization crest
- personal logo
- originator identifier

Intrinsic Elements of Documentary Form

Definition: The elements of a record that convey the action in which the record participates and its immediate context.

1. Name of Author

Definition: Name of the physical or juridical person having the authority and capacity to issue the record or in whose name or by whose command the record has been issued.
Note: In traditional records, the name of the author typically appears as the name expressed in the letterhead (*entitling*), in the initial wording of the record (*superscription*), and/or at the bottom of the record (*subscription*). It may be the same name as that of the writer, and, with records that are electronically transmitted, may correspond to the name of the originator. However, the name of the author only validates the record when it has the function of an attestation.

2. Name of Originator

Definition: Name of the person assigned the electronic address in which the record has been generated and/or sent.
Note: When the name of the originator is different from the name of the author of the record, the law usually considers the originator's name as the indication of the person responsible for issuing the record.

3. Chronological Date

Definition: The chronological date is the date, and possibly the time, of the record's compilation included in the record by the author or the electronic system on the author's behalf.

4. Name of Place of Origin of Record

Definition: The name of the geographic place where the record was generated, included in the content of the record by the author or the electronic system on the author's behalf.

5. Name of Addressee(s)

Definition: The name of the person(s*)* to whom the record is directed or for whom the record is intended.
Note: In traditional records this element corresponds to the *inscription* and usually occurs at the top of the record. With electronic mail records, the name of the addressee(s) continues to appear in the top portion of the record (i.e., in a header).

6. Name of Receiver(s)

Definition: The name of the person(s) to whom the record is copied for information purposes.

7. Indication of Action or Matter

Definition: The *subject* line(s) and/or the *title* at the top of the record.

8. Description of Action or Matter

Definition: Presentation of the ideal motivation (*preamble*) and the concrete reason (*exposition*) for the action as well as the action or matter itself (*disposition*).

9. Name of Writer

Definition: The name of the *person* having the authority and capacity to articulate the content of the record.
Note: In traditional records, the name of the writer usually appears at the bottom of the record and is typically constituted by the *subscription*. The name of the writer may be the same as the name of the author (and perhaps of the originator).

10. Corroboration

Definition: Explicit mention of the means used to validate the record.
Note: *To validate* means to make legally valid; to grant official sanction to by marking; to support or corroborate on a sound or authoritative basis.

11. Attestation

Definition: The written validation of a record by those who took part in the issuing of it (author, writer, counter-signer) and by witnesses to the action or to the signing of the record.
Note: In traditional records, the attestations usually appear as *signatures* at the bottom of the record (the eschatocol). However, some records have the attestation in the protocol. For example, memoranda may be signed or initialled beside the *superscription*. With electronic records, such as electronic mail messages, the attestation appears in the header of the message.

12. Qualification of Signature

Definition: The mention of the title, capacity and/or address of the persons signing a record.
Note: Qualification of signature may follow either a *subscription* or a *superscription*.

Annotations

Definition: Additions made to a record after it has been created.

Annotations Made in the Course of Executing the Record

Definition: Additions made to a record after it has been created as part of the formal execution phase of an administrative procedure.
Note: Normally this sort of annotation is used only for the authentication and registration of legal records whose form is required by law, e.g., the registration number added to a land deed by the land registry office, or the statement of the authenticity of the signatures in a will.

Examples: Such additions might include, but are not limited to the following:

- **Priority of Transmission**

Definition: Indication of the priority in which a record is to be transmitted.

- **Transmission Date, Time and/or Place.**

Definition: The *date*, *time*, and/or *place* when the record leaves the space in which it was generated.
Note: Transmission date, time and/or place is usually added by the electronic system.

- **Indication of Attachments**

Definition: Mention of autonomous items that have been linked inextricably to the record before transmission (i.e., added during its execution) in order for it to accomplish its purpose.

Annotations Made in the Course of Handling the Business Matter to Which the Record Relates

Definition: Additions made to the record in the course of handling the business matter in which the record participates and reflecting actions taken subsequent to the creation of the record for the purpose of handling the action or matter in which the record participates.
Such additions might include, but are not limited to the following:

- Received Date and Time
- Name of Handling Office
- Action Taken
- Dates and Times of Further Action or Transmission

Annotations Made in the Course of Managing the Record for Records Management Purposes

Definition: Additions made to a record for the purpose of handling the record itself and reflecting actions taken subsequent to the creation of the record for the purpose of managing it as part of the agency's records.
Such additions might include, but are not limited to the following:

- **Archival Date**

Definition: The date on which a record is officially incorporated into the creator's records.

- **Draft or Version Number**

Definition: The unique identifier assigned to sequential drafts or versions of the same record, added to the record when it is saved.

- **Record Item Identifier**

Definition: The progressive number of the record within the dossier or, in the absence of dossiers, within the specific class.

- **Dossier Identifier**

Definition: The identifier for the dossier in which the record belongs.
Note: It may be constituted by the name of a person or organization, a symbol, a progressive number, a date, or a specific topic within the class's general subject.

- **Class Code**

Definition: The code of the class to which the record belongs, as it appears in the classification scheme, thus connecting it to other records in the same class.

- **Registration Number**

Definition: The consecutive number added to each incoming or outgoing record in the protocol register, which connects it to previous and subsequent records made or received by the creator.

- **Name of Creator**

Definition: The name of the *person* in whose archival fonds the record exists.

Medium

Definition: The physical carrier of the message.
Note: The medium is considered an essential component of the record inasmuch as a record does not exist until it has been affixed to a physical carrier.

Context

Definition: The framework of action in which the record participates.

Juridical-Administrative Context

Definition: The legal and organizational system in which the creating body belongs.
Note: Indicators of juridical-administrative context are laws, regulations, etc.

Provenancial Context

Definition: The creating body, its mandate, structure, and functions.
Note: Indicators of provenancial context are organizational charts, annual reports, the classification scheme, etc.

Procedural Context

Definition: The business procedure in the course of which the record is created.
Note: In some organizations, the business procedures are integrated with documentary procedures. Indicators of procedural context are work-flow rules, codes of administrative procedure, classification schemes, etc.

Documentary Context

Definition: The fonds to which the record belongs and its internal structure.
Note: Indicators of documentary context are classification schemes, record inventories, indexes, registers, etc.

Technological Context

Definition: The characteristics of the technical components of the electronic system in which the record is created.

Hardware

1. Storage
Definition: The medium that stores data in the system.

Main Memory (aka primary memory)
Note: This type of storage is fast, different parts of it can be accessed randomly (rather than sequentially) and directly by the CPU/microprocessor. Thus, for a process to run or a file to be accessed, it must be loaded, at least partially, into the main memory. Main memory is provided via integrated circuit chips and does not involve mechanical movements. It is “volatile” in that its contents will be lost when a computer system is shut down.
Example: random access memory (RAM), cache memory.

Secondary Storage (aka secondary memory)
Note: This type of storage is slower than main memory and is cheaper. It involves mechanical parts and movements that contribute to its low speed of access. It is non-volatile in that shutting down the system will not result in loss of data on the secondary storage. Compared to magnetic tapes, secondary storage devices are randomly accessible.
Examples: hard disks, magnetic or optical disks, CD-ROMs, DVDs.

Tertiary Storage
Note: This type of storage is sequentially accessible only, and is used for long-term file preservation.
Examples: magnetic and digital tapes.

Storage for Security/Recovery Purposes
Note: This type of storage is used as a protective measure against the possibility of catastrophic loss. It tends to be overwritten at regular intervals and is not intended to serve the purpose of long-term file preservation.
Examples: magnetic and digital tapes.

2. CPU/Microprocessor
Definition: The primary resource for job/instruction execution.
Note: This resource can be broken down further into its own sub-systems (e.g., registers and logic units). Its speed of executing instructions is considerably higher than the speed of accessing main memory. It interfaces directly with main memory, so a record must be loaded into main memory from secondary or tertiary storage before it can be readable.

3. Network
Definition: The primary source of communication between systems or components thereof.
Note: Network encompasses its own types of hardware, software, and architectures.
4. Peripheral Devices
Examples: Mouse, monitor, keyboard, printer.
5. Architecture
Definition: The configuration of hardware components and their interfaces.
Examples: CPU architecture, motherboard architecture, system architecture (i.e., serial, pipelined, parallel, distributed, client-server), network architecture.

Software

1. Operating System
Definition: The system that manages, controls, protects and facilitates the use of hardware resources in the electronic system.
Note The following can be identified as functions and main modules of an operating system: process management (scheduling, switching), deadlock management, memory management, secondary storage management, storage scheme (data mapping), disk scheduling, virtual memory management, file system (distributed, file format, directories), interrupt handling, user interface, device and network interface. The way an operating system is configured (parameterized), may affect certain aspects of data and files in the system. For example, there may be a limit imposed on the size of a data file.
2. System Software
Definition: Software that creates an environment for application programs to be created, executed, and maintained, typically through system calls to the operating system.
Note: System software is sometimes referred to as system utilities or system tools.
Examples: languages (machine language, high-level languages), compilers, interpreters and translators, coding (compression, encryption), system utilities (i.e., hard disk defragmentation tools, virus detectors, etc.).
3. Network Software
Definition: Network software manages networks and their resources in order to meet the communication requirements of one or more applications.
Examples: protocols, routing, and switching software.
4. Application Software
Definition: Software that constitutes any type of program that is tailored to satisfy real-world needs and requirements.
Note: Application software varies widely in nature and complexity, as the range of applications using this type of software is quite diverse. Application software may be developed in-house by the organization that uses it, custom-made by another company or contractor for the organization that uses it, or purchased as an off-the-shelf package. It is important to know whether the software includes source code, documentation, and other components, in addition to the executables. As in the operating system, a set of parameters or characteristics may be associated with the application software whose values affect the number, format and size of the records that are handled.
Examples: Microsoft Word, Lotus 1-2-3, Netscape Communicator, database management system (DBMS) software, computer-aided design (CAD) software.

Data

Definition: numbers, characters, images or other methods of recording that represent values that can be stored, processed, and transmitted by electronic systems.

1. File Structure
Definition: The relationship and organization of files within a system.
Note: File structure includes the directory structure of a file system. The physical structure and organization of files in a file system may also constitute an aspect of the file structure and data format. This can include the mapping of files onto disk blocks of each disk plate, and among a set of disks.
2. Data Format/File Format
Definition: The organization of data within files. These are organizations that are usually designed to facilitate the storage, retrieval, processing, presentation, and/or transmission of the data by software.

Note: Data format is concerned with the representation of each piece of data and the relationship between pieces of data. Within a file, it includes standardized data formats such as ASCII text, as well as proprietary file formats such as Microsoft's Word97 and Adobe's PDF file formats. It also includes structures such as the tabular format of data files in a database management system, and the format (using tags) of data files used by mark-up languages.
Examples: portable document format (PDF), rich text format (RTF), ASCII text.

System Models

Definition: System models are abstractions that represent the entities, activities and/or concepts in the system as well as their attributes, characteristics, and the functional relationship between them.
Note: “Functional relationship” refers to a relationship involving two or more entities/objects that it is important to represent explicitly in order for the application to function correctly. System models contrast with data format and file structure in that they represent behavioural, procedural, and/or functional aspects of a system or software application. They may, however, affect directly or indirectly the way files are conceived in an application and the way data are organized within the files in an application. A model is usually represented graphically (e.g., as in entity-relationship, object-hierarchy, data-flow, control-flow, and state-transition diagrams). Modelling languages (e.g., IDEF, UML) and their associated software tools serve as aides in creating model specifications. The model usually becomes part of an application’s requirements, specifications, and/or design document. Parts of the model can also be represented and used in an application’s data dictionary.
Examples: entity-relationship models, object domain diagrams, IDEF(0) process models, UML use-case models, data-flow diagrams.

System Administration

Definition: System administration is a set of procedures that ensure correct, secure, reliable, and persistent operation of the system.
Examples: Providing access privileges; ensuring security, availability, reliability and integrity of the system over time; configuring the system; backing up files; system maintenance; and upgrading hardware, software and storage systems.

Appendix B. The Method of Diplomatic Criticism

This appendix shows the method of diplomatic criticism represented as a procedure. This procedure was created as part of an early exercise of the InterPARES project to introduce the researchers to the method of diplomatic criticism and to apply it to various kinds of electronic records.[1] The procedure does not include considerations of contemporary diplomatic criticism, such as determination of archival bond. Nor does the procedure include extensions needed to analyze records in electronic systems.

1. [Determine extrinsic elements (physical form)]

Describe the medium of the document, preparation of the medium for receiving the message (edging, ruling), and shape or size of the medium.
Describe the script of the document, that is, layout, pagination, formatting; types of scripts; different hands; typefaces or inks, paragraphing; punctuation, abbreviations and initialisms; erasures and corrections; computer software; formulae.
Describe the language of the document, that is, the vocabulary, composition and style.
Describe special signs, such as, signs of writers and subscribers, or signs of chanceries and records offices.
Describe seals as to their material, shape and size, typology, legend for inscription and method of affixing.
Describe seals included in the execution phase for authentication or registration.
Describe the annotations included in the handling phase, that is signs beside text, previous or following actions, dates of hearings or readings, notes of transmission, disposition, subject, "Urgent", "Bring Forward".
Describe annotations included in the management phase, that is, registry number, classification number, cross-references, date and office of receipt, and archival identifiers.

2. [Determine intrinsic elements (intellectual form)]

Describe the protocol of the document, that is, from the beginning of the document up to the beginning of the text. This may include entitling, title, date, invocation, superscription, inscription, salutation, subject, *formula perpetuitatis*, appreciation, or other elements.
Describe the text of the document, that is, the part that manifests the will of the author, the evidence of an act, or the memory of it. This may include, preamble, notification, exposition, disposition, final clauses or other elements.
Describe the eschatocol of the document, that is, the part that follows the text of the document. This may include corroboration, date, appreciation, salutation, complimentary clause, attestation, and qualification of signature or secretarial notes.

3. [Determine persons]

Determine the author of the act, author of the document, addressee of the act, addressee of the document, writer, counter signer(s), or witnesses.

[1] W. E. Underwood. "Diplomatic analysis of electronic military messages." *Archivi per la storia*, XII, 1-2, (December 1999): 121-29.

4. [Determine type of act]

If the agent of an act is one individual or organ, then the type of act is a simple act.
If the agents of an act are two or more individuals, then the type of act is a collegial act.
If the agents of an act are two or more interacting parties (individuals, public bodies, states, states-and individuals,), then the type of act is a contract.
If the agents of an act are two or more individuals with identical intents, and produce one document, then the type of act is a collective act.
If the agent of an act is a single individual or organ, but is directed to different individuals or organs, and results in one document, then the type of act is a multiple acts.
If the agent of an act is one or more individuals or organs, and the act is composed of many different acts each resulting in documents that are necessary to the function of the final document, then the type of act is compound act.
If the type of action is compound act, and a single individual or organ repeatedly performs the same act but produces different documents all necessary to the final, definite act, then the compound act is a continuative act.
If the type of act is a compound act and more than one individual or organ perform similar acts, all necessary to the accomplishment of the final act, then the compound act is a complex act.
If the compound act is made up of a series (in sequence or parallel) of different acts, which may be simple, or compound, collegial or collective, performed by different individuals and/or organs which may have similar or different motivations and perform different functions, all making possible the accomplishment of a final act, the compound act is called an act on procedure.

5. [Determine the name of the act]

Examine the disposition component of the text to determine the name of the act.

6. [Determine relationship between document and procedure]

If a document is constituted of acts that initiate a procedure, then the document relates to the initiative phase of a procedure.
If a document is constituted of acts that are necessary to evaluate the situation, then the document relates to the inquiry phase of a procedure.
If a document is constituted of acts of opinion or advice after the relevant data have been assembled, then the document relates to the consultative phase of a procedure.
If a document is constituted of act(s) that reflect final decisions are being made, then the document relates to the deliberative phase of a procedure.
If a document is constituted of acts by a person other than the author of a deliberative document, then the document relates to the controlling phase of a procedure.
If a document is constituted of actions that give formal character to the transaction, e.g., validation, communication, notification, publication, then the document is related to the executive phase of a procedure.

7. [Determine type of document]

a. [Determine form name of document]

If a document has the physical and intellectual form (extrinsic and intrinsic elements) of a document type previously named and defined, then it is of that form name.

If a document has a different physical and intellectual form than previously defined document names, then name and define a new type of document type in terms of the documents form, and the document is of this new form name.

b. [Determine nature of document]

If a document is created by a public person or by his command or in his name, the nature of a document is public.

If a document is created by a private person or by his command or in his name, then the nature of the document is private.

c. [Determine function of document]

If a document puts a juridical act into existence, then the function of the document is dispositive.

If a document produces evidence of a juridical act that came into existence before the document was written, the function of the document is probative.

If a document constitutes written evidence of a juridically relevant activity that does not result in a juridical act, then the function of the document is supporting.

If a document constitutes written evidence of an activity that is juridically irrelevant, then the function of the document is narrative.

d. [Determine status of transmission]

If a document is a temporary version prepared for purposes of correction, then the state of transmission is draft.

If a document is the first complete and effective document, then the state of transmission is original.

If a document is a reproduction of a document, then its state of transmission is copy.

If a document is a copy and is a transcription of the content but not the form, then the document's state of transmission is simple copy.

If a document is a copy and reproduces both the form and content of the original document, e.g., a photocopy, then the status of transmission is an imitative copy.

If a document is a copy and is a complete and effective record, but is not the first to be created, then the document's status of transmission is a copy in the form of the original.

If a document is a copy, and an officer who is certified to perform such a function certifies the document, the state of transmission is authentic copy.

8. [Determine Diplomatic Description]

The document description consists of:

(i) The year, month, day and place that the document was created.
(ii) A sentence indicating the persons and actions involved
(iii) The form name, nature, function and status of transmission

9. [Comments regarding the document as a whole]

If document is written according to the practice of the time and place indicated in the text, and signed with the name(s) of the person(s) competent to create it, then it is a diplomatically authentic document.

If a document lacks written physical or intellectual elements of form of the practice of the time and place indicated in the text, then it is diplomatically inauthentic.

If signature or seal of a document is counterfeit, then document is counterfeit.

If document is counterfeit, forged or somehow tampered with, then it is a diplomatically false document.

If document is created or received by a physical or juridical person in the course of a practical activity, then it is an archival document.

If a reliable person wrote the document, then document is reliable.

If document has all the required elements of documentary form, then document is complete.

If an archived document is reliable and complete, then document is a record.

Appendix C. Survey of Applicable U.S. Law

> A good overview of the legal issues related to electronic records in the U.S. context has been provided earlier in this report in the section *Translating Research Outcomes into Practice*. This appendix, however, provides a more detailed commentary on United States law that relates to the work of the InterPARES Project.

United States law consists of federal law plus the law of 51 jurisdictions (the 50 states plus the District of Columbia), presenting a wide array of approaches to the subject of electronic records. This Appendix is not intended to be an exhaustive compilation or analysis of such laws; rather, it constitutes brief commentary on those federal laws, regulations, and cases that appear most relevant to the important and innovative issues raised by the combined InterPARES projects concerning preserving electronic records and ensuring their authenticity over time.[1] For example, U.S. law relating to "records" neither explicitly embraces nor precludes the important principle arising out of the work of the Preservation Task Force that it is possible only to preserve the ability to reproduce the electronic record, rather than to preserve the electronic record "itself." With respect to the remaining principles derived from the work of the InterPARES project, and subject only to limitations imposed by copyright law,[2] no insuperable barriers to ratification and utilization of the principles exist in the context of laws affecting U.S. records and recordkeeping practices; indeed, as shown below, there are particular points of congruence.

There are a number of important sources of U.S. law which affect the way in which electronic records are utilized in federal courts, used and preserved in federal agencies, and used in business and consumer transactions involving the United States as a party. Additionally, there are federal statutes that address electronic communications and transactions taking place in the private sector, including recent U.S. legislation concerning the acceptance of electronic and digital signatures in global commerce.

1. Federal Rules of Evidence

The Federal Rules of Evidence (FRE) are congressionally mandated for use in ninety-four federal district courts throughout the U.S. A settled body of precedent under these rules exists involving the admissibility and evidentiary use of records in federal judicial proceedings, including electronic records. Assuming that the relevance of a particular electronic record has been established, its admissibility in court depends on proof of its identity or authenticity, and whether it is allowable into evidence by an exception to the hearsay rule.

With respect to authentication, the federal "best evidence" rule, as set out in FRE 1002, states that "[t]o prove the content of a writing, recording or photograph, the original writing, recording, or photograph is required, except as otherwise provided in these rules . . ." Electronic data is encompassed within what constitutes an "original" writing or recording: "If data are stored in a computer or similar device, any printout or other output readable by sight, shown to reflect the data accurately, is an original."[3]

[1] In focusing on federal law, we will only have brief occasion to reference recently proposed uniform and other state laws addressing digital signature issues which may be preempted in whole or in part by recent Congressional enactments. Nevertheless, a rich array of state law recently has come into existence on the topic of digital signatures, as well as on other emerging issues affecting electronic records. *See generally*: <http://www.in.gov/digitalsignatures/statedslaws.html>.

[2] As described *infra*, federal copyright law precludes certain preservation activities involving the reproduction of records that are not in the public domain (regardless of how one cares to characterize the "preservation" function involved).

[3] FRE 1001(3).

This definition of what consists of an “original” record would appear sufficiently broad enough to incorporate the notion of electronically preserving the *ability* to accurately reproduce the stored record (in the form of a printout or other output readable by sight), rather than insisting on the requirement of producing “the record” itself. [cf. STF Report, 6th Principle.]

In turn, the definition of what constitutes a “duplicate” includes the phrase “a counterpart produced by . . . electronic re-recording . . . or by other equivalent techniques which accurately reproduce[ ] the original.”[4] “Duplicates” are admissible in court to the same extent as an original, unless a genuine question is raised as to the authenticity of the original or under other circumstances where it would be unfair to admit the duplicates into evidence.[5]

In addition to the various means of authentication of records set out in FRE 901, the “self-authentication” provisions in FRE 902 provide that extrinsic evidence of authenticity as a condition precedent to admissibility is not required with respect to certain forms of records of regularly conducted activities, where a written statement of a custodian or other qualified person can be obtained certifying that the record was (A) made at or near the time of the occurrence of the matters set forth by a person with knowledge of those matters, (B) was kept in the course of the regularly conducted activity, and (C) was made by the regularly conducted activity as a regular practice.[6]

Even if deemed authentic, records created in the course of business are “statements,” *i.e.*, “written assertions,” under FRE 801(a), which if offered into evidence to prove the truth of the matter being asserted constitute hearsay.[7] However, under one or more business record exceptions to the hearsay rule, they may be admitted into evidence. For example, “Records of regularly conducted activity,” defined in FRE 803(6) as constituting an exception to the hearsay rule, include “a memorandum, report, record, or data compilation, in any form.” “It is well-settled that computer compilations may constitute business records for purposes of [FRE 803(6)] and may be admitted at trial if a proper foundation is established.”[8] As stated by one leading treatise written in 1994, recent cases “appear to accept the reliability of computerized recordkeeping systems virtually without question.”[9]

Whether this general presumption in favor of the admissibility of electronic records kept in the usual course of business will remain intact, where courts (and juries) are faced with authentication and non-repudiation issues arising out of new forms of electronic media—including the use of electronic and digital signature technologies and the anticipated emergence of a public key infrastructure—is an open question awaiting further case-by-case development.

[4] FRE 1001(4).

[5] FRE 1003.

[6] FRE 902(11).

[7] FRE 801(c).

[8] *United States v. Croft*, 750 F.2d 1354, 1365 (7th Cir. 1985). *See also United States v. Russo*, 480 F.2d 1228 (6th Cir. 1973) (court recognizing that a liberal construction of the rules of admissibility is required in a case of computer-stored evidence).

[9] Donald S. Skupsky and John C. Montana, *Law, Records and Information Management: The Court Cases* (Denver: Information Requirements Clearinghouse, 1994), 59.

2. Federal Recordkeeping Law

a. The Federal Records Act.

The "Federal Records Act" (FRA) in actuality consists of a series of statutes passed by Congress beginning in 1943.[10] The definition of what constitutes a "federal record" has remained stable since enactment of the Federal Records Act of 1950. As codified at 44 U.S.C. 3301, a "federal record" consists of

> all books, papers, maps, photographs, machine readable materials, or other documentary materials, regardless of physical form or characteristics, made or received by an agency of the United States under Federal law or in connection with the transaction of public business and preserved or appropriate for preservation by that agency . . . as evidence of the organization, functions, policies, decisions, procedures, operations, or other activities of the Government or because of the informational value of the data in them.[11]

The FRA governs all aspects of the life cycle of records in more than 300 departments, agencies, and other components of the federal government as identified by NARA. Each agency head is expected to "maintain an active, continuing program for the economical and efficient management of the records of the agency."[12] The Archivist provides "guidance and assistance to the Federal agencies with respect to ensuring adequate and proper documentation of the policies and transactions of the Federal government and ensuring proper record disposition."[13] The Archivist "signs off" on the destruction of temporary records either by approving individually-submitted agency records schedules, or through the use of general records schedules covering the most routine, administrative federal records commonly used by federal agencies, both pursuant to a notice and comment process.[14]

In *Armstrong v. Executive Office of the President*,[15] the government argued that the FRA allowed federal agencies within the Executive Office of the President the discretion to designate electronic versions of e-mail records situated on "live," networked e-mail systems, as "nonrecord" material kept only for the convenience of the end-user, so long as policies were in place that hard-copy versions of electronic mail were printed out and placed in traditional paper files. The district court disagreed, holding that paper versions of e-mail records were incomplete insofar as they failed to contain "who, what, where, when" information, such as a full list of recipients of the email, and when the e-mail was received (if important), and thus the email could not be said to constitute merely "extra copies" kept for convenience.[16] This holding was affirmed in the Court of Appeals, which found that paper and electronic versions of e-mail were at best "kissing cousins," differing in that the hard copies constituted amputated versions of the

[10] *See* Records Disposal Act of 1943, 57 Stat. 380; Federal Records Act of 1950, 64 Stat. 583; Government Records Disposal Amendments of 1970, 84 Stat. 320; Federal Records Management Amendments of 1976, 90 Stat. 273; and the National Archives and Records Administration Act of 1984, 98 Stat. 2280.

[11] 44 U.S.C. 3301 sets out three nonrecord categories of documents, including library and museum materials acquired for exhibition purposes; stocks of publications; and "extra copies of documents preserved only for convenience of reference." NARA regulations further provide that working files, such as preliminary drafts and rough notes, constitute "records" and shall be maintained for adequate documentary purposes (for however short or long that might be), under two conditions. First, the drafts must be circulated to other agency employees, including for comment or action; and second, the drafts must contain unique information, such as substantive annotations or comments, that add to a proper understanding of agency policies, decisions, actions, or responsibilities. 36 C.F.R. 1222.36(c).

[12] 44 U.S.C. 3102.

[13] 44 U.S.C. 2904(a).

[14] 44 U.S.C. 3303a(a) & 3303a(d), respectively

[15] 810 F. Supp. 335 (D.D.C.), *aff'd in relevant part*, 1 F.3d 1274 (D.C. Cir. 1993).

[16] 810 F. Supp. at 341.

complete records in electronic form.[17] The government's further argument in the Court of Appeals that any additional information contained in electronic versions was itself nonrecord was likewise rejected.[18]

The rulings in *Armstrong* established the principle that certain contextual metadata in electronic records is valuable and should be managed under an agency's overall FRA obligations; the case left open, however, *how* such metadata should be managed, including the basic question of whether electronic records must be preserved in their electronic form (for any and all purposes, including establishing authenticity). In the successor case of *Pubic Citizen v. Carlin*,[19] the Court of Appeals narrowed the import of *Armstrong* by upholding the validity of NARA's General Record Schedule 20, which allows electronic mail (with annotations showing transmission and receipt data, per *Armstrong*), as well as word processing documents, to be deleted off "live" systems *provided* the agency preserves a recordkeeping copy in either a traditional paper system of official files or in an electronic recordkeeping system.

The *Carlin* court recognized that the act of requiring the "setting aside" of electronic records created on a "live" system (*i.e.*, an "electronic information system" as defined under NARA regulations[20]), with the intent of placing the records in an official recordkeeping system of whatever kind (paper or electronic) prior to their deletion, had a rational basis under the FRA.[21] [cf. STF Report, 1st principle.] Nothing the Court said precludes, however, the possibility that proprietary software meeting the requirements of what constitutes an "electronic recordkeeping system" (as defined by NARA at 36 C.F.R. 1234.22) may be fully integrated with "live" desktop applications in one system.

Neither *Armstrong* nor *Carlin* purport to address in a systematic way what types of contextual "metadata" are worthy of preservation in an agency's recordkeeping system, for purposes of establishing authenticity over time or otherwise.[22] Nor does this case law directly confront the records management concerns arising with respect to newer forms of electronic records in multimedia (*e.g.*, hypertext links using HTML, videoconferencing, integrated voice mail, etc.), especially with respect to what form "setting aside" such records into future agency recordkeeping systems will take.

b. U.S. Department of Defense Directive 5015.2

U.S. Department of Defense (DoD) Design Criteria Directive 5015.2-STD provides a minimum set of functional requirements for proprietary software to meet in order to perform "electronic recordkeeping" functions.[23] NARA has provided a non-exclusive endorsement of the DoD Directive and the Joint

[17] 1 F.3d at 1285, 1286.

[18] 1 F.3d at 1292. In an interesting side-passage, the *Armstrong* panel spells out in a footnote that its holding would not change even if the extra data or metadata of interest (*e.g.*, directories, distribution lists), appearing only in electronic form is not assumed to be an "integral part[] of the electronic record," but rather form separate electronic records themselves. Id. at 1284 & n.8.

[19] 184 F.3d 900 (D.C. Cir. 1999), *cert. denied*, 529 U.S. 1003 (2000).

[20] *See 36* C.F.R. 1234.2

[21] In the Court's words: "The district court concluded . . . that the 'common feature of the records scheduled under GRS 20--the fact that they have been generated by electronic technology--has no relation to each record's value.' That captures only half the matter, however. GRS 20 does not authorize disposal of electronic records per se; rather, such records may be discarded only after they have been copied into an agency recordkeeping system. Therefore, GRS 20 seems to us to embody a reasoned approach. . . ." 184 F.3d at 905 (internal citations and quotations omitted).

[22] NARA is currently considering the issue of whether to have regulations explicitly provide that the "content, context, and structure" of electronic records must be managed, and what such a requirement would entail. See 66 Fed. Reg. 51739 (Oct. 10, 2001) (Advanced Notice of Proposed Rulemaking).

[23] *See* <http://jitc.fhu.disa.mil/recmgt/#standard>.

Interoperability Test Command's (JITC) software testing program,[24] finding that they collectively meet the set of baseline recordkeeping requirements contained in NARA regulations, at 36 C.F.R. 1234.22.

As of May 20, 2002, the JITC register reports that 31 separate software programs have been tested and approved as compliant with the 5015.2-STD.[25] It is not known how many components of federal agencies are currently utilizing software on the register at the present time.

c. Presidential Records Act

The Presidential Records Act (PRA) was enacted in 1978 as a post-Watergate reform, to ensure that at the end of a President's term(s) in office, legal ownership of the President's public papers passes to the U.S. government, rather than continuing to reside in the private hands of the former President.[26] The Act covers the records of the Reagan presidency and all subsequent Administrations. "Documentary material" covered by the Act include "audio, audiovisual, or other electronic or mechanical recordations."[27] In turn, "presidential records" are defined to mean those documentary materials created or received by the President, his immediate staff, or units or individuals of the EOP whose function is to solely advise and assist the President in the course of carrying out constitutional, statutory, official, or ceremonial duties.[28]

Upon leaving office, a President may restrict certain designated categories of records for up to twelve years, including records relating to national security, consisting of confidential communications between the President and his advisers, and for personal privacy.[29] All presidential records are presumptively closed from public access for five years, except that the Archivist may open particular non-restricted collections of records earlier.[30] After the end of these restriction periods, access is obtained through the Freedom of Information Act, as incorporated into the PRA.[31] Special access to PRA records may be obtained under court order or pursuant to congressional authorization, and a former President may access the records of his Administration.[32]

At the end of the Clinton Administration, NARA took in approximately 32 million e-mail records created in the Clinton Executive Office of the President, approximately 16 million of which constitute "Presidential records" (with the remainder designated as federal records). This body of presidential e-mail was captured in the EOP's Automated Records Management System (ARMS), created in the Clinton EOP in response to the *Armstrong* decision. The system did not allow for digital or electronic signatures being appended to e-mail or word processing documents.

3. Federal Access & Privacy Statutes

a. The Freedom of Information Act

The Freedom of Information Act, originally enacted in 1966,[33] and significantly amended in 1974,[34] provides a statutory right of access to government information.[35] The Act has been termed "a model for

[24] *See* <http://www.archives.gov/records_management/policy_and_guidance/joint_interoperability_letter.html>.
[25] *See* <http://jitc.fhu.disa.mil/recmgt>.
[26] *See* 44 U.S.C. 2201-2207.
[27] 44 U.S.C. 2201(1).
[28] 44 U.S.C. 2202(2).
[29] 44 U.S.C. 2204(a).
[30] 44 U.S.C. 2204(b).
[31] 44 U.S.C. 2204(c).
[32] 44 U.S.C. 2205.
[33] 80 Stat. 250.
[34] 88 Stat. 1561.
[35] 5 U.S.C. 552.

governmental transparency throughout the world."[36] Any individual has the right to obtain access to federal agency records, except to the extent that the records are protected from disclosure by nine statutory exemptions.[37] However, the original FOIA made no reference to electronically stored information, nor did the Act contain an explicit definition of what constituted an agency "record." In the intervening years, case law arose interpreting the right to access electronic information narrowly. *See, e.g.*, *SDC Development Corp. v. Mathews,* 542 F.2d 1116 (9th Cir. 1976) (medical database compiled by the National Library of Medicine did not constitute agency "records" and thus was outside scope of FOIA).

In enacting the Electronic Freedom of Information Act Amendments of 1996,[38] Congress rejected the holding of *SDC Development Corp.*,[39] and made explicit that the public's access to government information generally includes the right to request that electronic records are disclosed in their electronic form, at the requestor's option. Under the EFOIA's definition, a "record" consists of "any information that would be an agency record subject to the requirements of [the EFOIA] when maintained by an agency in any format, including an electronic format."[40] The EFOIA further provides that:

> In making any record available to a person under this paragraph, an agency shall provide the record in any form or format requested by the person if the record is readily reproducible by the agency in that form or format. Each agency shall make reasonable efforts to maintain its records in forms or formats that are reproducible for purposes of this section.[41]

Lastly, the EFOIA also requires that agencies "shall make reasonable efforts to search for the records in electronic form or format, except when such efforts would significantly interfere with the operation of the agency's automated information system.[42] Case law has yet to arise as to what constitutes "reasonable efforts" by agencies to maintain records in electronic formats (or to search for electronic records), including with respect to records containing electronic or digital signatures and as part of an emerging federal public key infrastructure.

b. Privacy Act

The Privacy Act of 1974[43] regulates the collection, maintenance, use and dissemination of personal information by federal agencies, principally by requiring agencies to provide notice about what information is contained in "systems of records" and how such information may be stored, accessed, and used. In enacting the Privacy Act, Congress had in mind curbing illegal surveillance of individuals by federal agencies, as well as the use and abuse of computer technology.[44] A "record" for purposes of the Privacy Act is

> any item, collection, or grouping of information about an individual that is maintained by an agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history, and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph.[45]

[36] Henry H. Perritt, Jr*., Electronic Freedom of Information*, 50 Admin. L. Rev. 391, 391 (1998).
[37] 5 U.S.C. 552(b).
[38] 110 Stat. 2422.
[39] *See* H.R. Rep. 104-795, at 20 (1996), reprinted in 1996 U.S.C.C.A.N. 3448, 3463.
[40] 5 U.S.C. 552(f)(2).
[41] *Id.*, 552(a)(3)(B).
[42] *Id.,* 552(a)(3)(C).
[43] *See* 5 U.S.C. 552a.
[44] *See Thomas v. U.S. Dep't of Energy*, 719 F.2d 342, 345 (10th Cir. 1983).
[45] 5 U.S.C. 552a(a)(4).

A Privacy Act "record" can include merely one descriptive item concerning an individual, so long as it is linked to that individual through an identifying name, number, symbol, or other identifying particular in a "system of records" as defined by the Act.[46] Individuals have both an express right of access to agency records maintained on themselves,[47] and a right to seek amendment of agency records, including to the extent those records are not "accurate" or "complete," *i.e.*, authentic.[48] The Act provides an express right of judicial review should an agency decline to amend a record in accordance with a request,[49] as well as the right to sue for damages due to an agency's maintenance of inaccurate records.[50]

c. Family Educational Rights and Privacy Act

The Federal Family Educational Rights and Privacy Act of 1974 (FERPA),[51] provides limited access to student educational records.[52] The pur- pose of the Act is: "(1) to create a right of access to student records for parents and students in order to ensure their accuracy and (2) to protect the privacy of those records by preventing unauthorized access by third parties." [53] The two main concerns of FERPA, privacy and reliability of student educational records are highlighted in the section of the statute that denies federal funding for failure to comply.

> No funds shall be made available under any applicable program to any educational agency or institution unless the parents of students who are or have been in attendance at a school of such agency or at such institution are provided an opportunity for a hearing by such agency or institution, in accordance with regulations of the Secretary, to challenge the content of such student's education records, in order to insure that the records are not inaccurate, misleading, or otherwise in violation of the privacy or other rights of students, and to provide an opportunity for the correction or deletion of any such inaccurate, misleading, or otherwise inappropriate data contained therein and to insert into such records a written explanation of the parents respecting the content of such records.[54]

FERPA defines "education records" as "those records, files, documents, and other materials which . . . contain information directly related to a student . . . and . . . are maintained by an educational agency or institution or by a person acting for such agency or institution."[55] The information kept in a record does not need to be in written form; any permanent recording such as a tape, picture, or computer file can be a record.[56] Records must meet both prongs of this definition to be exempt from disclosure as student education records.

Given the statute's broad definition of what constitutes an "education record" and emphasis on the authenticity of educational records, FERPA provides another possible application of the findings and recommendations arising from the InterPARES Preservation and Authenticity Task Forces.

[46] 5 U.S.C. 552a(a)(5); *Quinn v. Stone,* 978 F.2d 126, 133 (3d Cir. 1992)
[47] 5 U.S.C. 552a(d)(1).
[48] *Id.* 552a(d)(2). The term "authentic" is not, however, used in the Act.
[49] *Id.* 552a(g)(1)(A).
[50] *Id.* 552a(g)(1)(C).
[51] 20 U.S.C. 1232g, available at: <http://www4.law.cornell.edu/uscode/20/1232g.html>.
[52] 20 U.S.C. 1232g
[53] *United States v. Miami Univ*., 91 F. Supp. 2d 1132, 1150 (S.D. Ohio 2000).
[54] 20 U.S.C. 1232g(3)(C)(2) (1994).
[55] 20 U.S.C. 1232g(a)(4)(A) (1994).
[56] *See* 34 C.F.R. 99.3

4. Electronic and Digital Signature Laws

a. The Government Paperwork Elimination Act

In 1998, Congress enacted the Government Paperwork Elimination Act (GPEA),[57] seeking to "preclude agencies or courts from systematically treating electronic documents and signatures less favorably than their paper counterparts."[58] The GPEA sets a deadline of October 21, 2003, by which time federal agencies are required to provide (1) "for the option of the electronic maintenance, submission, or disclosure of information, when practicable as a substitute for paper"; and (2) "for the use and acceptance of electronic signatures, when practicable."[59] The term "electronic signature" is defined as meaning "a method of signing an electronic message that (A) identifies and authenticates a particular person as the source of the electronic message; and (B) indicates such person's approval of the information contained in the electronic message.[60] GPEA states that electronic signatures or other forms of electronic authentication conforming to federal guidance interpreting GPEA "shall not be denied legal effect, validity, or enforceability because such records are in electronic form."[61]

GPEA also provides that procedures to be developed for the use and acceptance of electronic signatures shall include the requirement to "ensure that electronic signatures are as reliable as is appropriate for the purpose in question and keep intact the information submitted."[62] In guidance issued by the Office of Management and Budget (OMB), OMB has further interpreted this provision to include the duty of "due consideration" to

> ensuring that agencies comply with their recordkeeping responsibilities under the FRA for these electronic records. Electronic record keeping systems reliably preserve the information submitted, as required by the Federal Records Act and implementing regulations.[63]

Accordingly, OMB has informed agencies that, as part of their duties, they should "consider the record keeping functionality of any systems that store electronic documents and electronic signatures, to ensure users have appropriate access to the information and can meet the agency's record keeping needs."[64]

After urging agencies to perform a cost/benefit calculation and risk assessment of utilizing electronic signatures for specific agency transactions, OMB's guidance addresses how agencies should implement electronic signatures and transactions, noting several prominent examples of regulatory guidance concerning use of electronic signatures and other filings, as adopted by the SEC, EPA, FDA, IRS, and other federal entities.[65] OMB recognizes that one aspect of any implementation plan involves ensuring the "chain of custody," by providing for electronic audit trails for secure electronic transactions.[66]

[57] 112 Stat. 2861-749 to 2861-751.

[58] S. Rep. 105-335.

[59] Pub. L. 105-277, § 1704, 44 U.S.C. 3504 note.

[60] *Id.*, § 1710. In subsequent OMB guidance, it is noted that this definition is "consistent with other accepted legal definitions of signature," including under the Uniform Commercial Code, 1-201(39) (1970) ("signature has long been understood as including 'any symbol executed or adopted by a party with present intention to authenticate a writing'"), and the Uniform Electronic Transactions Act (UETA), *see* <http://www.nccusl.org>. *See* "Procedures and Guidance: Implementation of the Government Paper Work Elimination Act" ("OMB GPEA Guidance"), Part II, Section 2, 65 Fed. Reg. 25508 (May 2, 2000), *available at*: <http://www.whitehouse.gov/OMB/fedreg/gpea2.html>.

[61] Pub. L. 105-277, § 1707.

[62] *Id.*, § 1703(b)(C)

[63] *See* OMB GPEA Guidance, Part 1, Section 1.

[64] *Id.*, Part I, Section 3.

[65] *Id.*, Part 2, Section 8. See especially FDA regulations at 21 C.F.R. Part 11.

[66] *Id.*, Part 2, Section 8(e).

OMB's guidance directed NARA to develop policies and guidance on the management, preservation, and disposal of federal records with particular consideration to records issues associated with the use of electronic signature technologies.[67] NARA in turn issued its "Records Management Guidance for Agencies Implementing Electronic Signature Technologies" in October 2000,[68] which discusses characteristics of trustworthy records, including the concepts of reliability, authenticity, and integrity.[69] According to the NARA guidance, for "a record to remain reliable, authentic, with its integrity maintained, and useable for as long as the record is needed, it is necessary to preserve its content, context, and sometimes its structure."[70]

The NARA document gives two examples of approaches available to agencies to ensure the trustworthiness of their records over time. [cf. STF Report, 3rd Principle.] One approach is for an agency to choose to maintain adequate contextual documentation of the records' validity, gathered at or near the time of record signing, for as long as the electronically-signed record is retained. This approach is less dependent on technology and more easily maintained as technology evolves, but in using this approach, "the signature name may not remain readable over time because of bit-wise deterioration in the record or as a result of technological obsolescence."[71]

NARA's alternative approach is for an agency to choose to maintain the ability to revalidate digital signatures (including the public key used to validate the signature, any certificate related to that key, and the corresponding certificate revocation list from the certificate authority), for as long as the digitally-signed record is retained.[72] Under either approach, NARA requires that agencies ensure that a printed name of the electronic signer and the date when signature was executed be included as part of any human readable form of the electronic record (such as in an electronic display or printout), for any permanent records to be transferred to the legal custody of NARA.[73]

The Justice Department also issued guidance to federal agencies pursuant to GPEA, in which it is noted that GPEA leaves some authenticity issues unresolved:

> GPEA may not necessarily make electronic records valid and enforceable under all circumstances (any more than paper signatures are valid under all circumstances) . . . While th[e] wording [of Section 1707] bars courts from invalidating electronic records and signatures merely because they are in electronic form, it does not require courts to accept electronic records and signatures that are deficient in other respects merely because they are in electronic form. For example, if there are reasons to doubt that it was actually the electronic signature holder who affixed the signature in question, a court might not accept the electronic signature, just as it might decline to accept a paper signature that could not be verified.[74]

[67] *Id.,* Part 1, Section 3(e).

[68] *See* <http://www.nara.gov/records/policy/gpea.html>.

[69] *Id.*, Section 4.1.

[70] *Id.*, Section 4.2.

[71] *Id.,* Section 4.3.

[72] *Id.*

[73] *Id,* Sections 4.3 & 5.6.

[74] *See* "Legal Considerations in Designing and Implementing Electronic Processes: A Guide For Federal Agencies" (U.S. Department of Justice, November 2000), at 19, *available at*: <http://www.usdoj.gov:80/criminal/cybercrime/eprocess.pdf>.

Notwithstanding the thrust of the legislation, it remains to be seen whether a significant percentage of federal agencies will be reporting in October 2003 that they have implemented GPEA's requirements, as opposed to finding that it has not yet been "practicable" for them to do so.[75]

b. The Electronic Signatures in Global and National Commerce Act (E-Sign)

The E-Sign Act, effective October 1, 2000, "promotes the use of electronic contract formation, signatures, and recordkeeping in private commerce by establishing legal equivalence between: contracts written on paper and contracts in electronic form; pen-and-ink signatures and electronic signatures; and other legally-required written documents (termed 'records') and the same information in electronic form."[76] If parties choose to use electronic signatures and records in the course of "transactions" as defined by the Act, E-sign provides that (a) no signature, contract, or other record shall be denied legal effect, validity, or enforceability solely because it is in electronic form; and (b) no contract relating to a transaction shall be denied legal effect solely because an electronic signature or electronic record was used in its formation.[77]

With respect to record retention requirements of Federal agencies, E-Sign provides that if a Federal law or regulation requires that a document or particular information be retained by an individual or a company, retention may be by electronic means so long as the electronic record "accurately reflects the information set forth in the contract or other record."[78] As stated in OMB's guidance, "E-Sign generally preserves an agency's existing authority to specify standards and format for records filed with the agency."[79] The requirements imposed by Federal laws and by State laws administered by a State agency are generally subject to E-Sign beginning March 1, 2001 or, if related rulemaking is underway, June 1, 2001.[80]

E-Sign defines "record" as meaning "information that is inscribed on a tangible medium or that is stored in an electronic or other medium and is retrievable in perceivable form."[81] An "electronic record" under E-Sign means "a contract or other record created, generated, sent, communicated, received, or stored by electronic means."[82] The term "electronic signature" is defined as meaning "an electronic sound, symbol, or process, attached to or logically associated with a contract or other record and executed or adopted by a person with the intent to sign the record."[83]

E-Sign preempts state electronic or digital signature laws to the extent that they exhibit a preference for particular electronic signature technologies.[84] Also, to the extent parties agree to use electronic signatures or records, many federal and state laws otherwise requiring paper notices of the parties' commercial transactions are arguably superseded or preempted by E-Sign. Whether E-Sign can be said to supersede or preempt individual states' adoptions of certain uniform state laws, including the Uniform Computer

[75] *See* text at n.60, *supra.*

[76] OMB Guidance on Implementing the Electronic Signatures in Global and National Commerce Act (E_SIGN) (Sept. 25, 2000), at 2, *see*: <http://www.whitehouse.gov/omb/memoranda/esign-guidance.pdf>.

[77] 15 U.S.C. 7001(a).

[78] 15 U.S.C. 7001(d).

[79] *See* OMB E-Sign Guidance, Section A.4, interpreting 15 U.S.C. 7004.

[80] 15 U.S.C 7007.

[81] 15 U.S.C. 7006(9).

[82] 15 U.S.C. 7006(4).

[83] 15 U.S.C. 7006(5).

[84] 15 U.S.C. 7002(a)(2)(A)(ii).

Information Transactions Act (UCITA), and/or the UETA, *supra* n.62, presents complex legal issues beyond the scope of this survey.[85]

5. Copyright Law

a. Digital Millennium Copyright Act

The Digital Millennium Copyright Act of 1998 (DMCA)[86] implements two World Intellectual Property Organization (WIPO) treaties, and amends the Copyright Act, at Title 17 of the U.S. Code. The legislative history reflects the drafters concern that "the law must adapt in order to make digital networks safe places to disseminate and exploit copyrighted material."[87] The DMCA addressed and revised a number of sections of the Title 17 of the Copyright Act as well as adding several new sections including Section 1201 on anticirumvention.[88] For purposes of the InterPARES research project and findings, the most relevant sections address issues of reproduction of records involving copyrighted material and claims of infringement based on creation of derivative work.[89]

According to a report by the Committee on Intellectual Property Rights and the Emerging Information Infrastructure "[l]arge scale archiving of the cultural record requires resolution of two key legal issues the ability to make copies when migrating from one storage technology to another, and the ability to reformat, thereby creating derivative works when moving from one software technology to the next."[90] In addition, the report highlights the importance of the first sale doctrine[91] to the ability of libraries, museums, archives, and historical societies to acquire, preserve, and archive cultural heritage materials.

Section 108 contains specific provisions for archives and libraries to make and distribute a limited number of reproductions of copyrighted work without obtaining prior permission of the copyright

[85] UCITA, a planned revision of the Uniform Commercial Code (UCC) proposed by the National Conference of Commissioners on Uniform State Laws, *see* <http://www.nccusl.org>, covers transactions in intangible goods, including computer software, online databases, and has been adopted in almost all states.
See generally E-Signatures Basics of the U.S. Structure, 38 Houston Law Review 921 (Fall 2001), available on Westlaw, 38 HOULR 921 (JLR database). *See also* <http://www.nccusl.org/nccusl/UCITA-2001-comm-fin.htm> (Report of the UCITA Standby Committee, Dec. 17, 2001, Recommendation 2) (describing relationship between UCITA and E-Sign). For a good summary of the preemption of state law issues involved with UETA, from a consumer rights organization perspective, *see* Margot Saunders and Gail Hillebrand, "E-Sign and UETA: What Should States Do Now?" *available at* <http://www.consumerlaw.org/e-sign.html>.

[86] Pub. L. 105-304, 112 Stat. 2860 (Oct. 28, 1998).

[87] Senate Committee on the Judiciary, Sen. Rep. No. 105-190*, Report on the Digital Millennium Copyright Act of 1998* ("Senate Report"), Section II: Legislative History, 2, *available at*: <http://www.ipmall.fplc.edu/hosted_resources/S._Rept._105-190.pdf>.

[88] 17 U.S.C. 1201. The anticircumvention provisions could have an impact on the implementation of InterPARES recommendations for long term digital preservation if a digital record or object can only be read through some type of proprietary software and if the proprietary software must migrate from one generation of hardware to the next. Until digital preservation of record content can be insured without any supporting proprietary software the anticirumvention provisions may inhibit some types of long-term digital preservation.

[89] 17 U.S.C. 101 et seq.

[90] Committee on Intellectual Property Rights and the Emerging Information Infrastructure, *The Digital Dilemma: Intellectual Property in the Information Age* (Washington, D.C: National Academy Press, 2000), 119.

[91] 17 U.S.C. 109(a). The first sale doctrine states that once a copyright owner sells a copy of his work to another, the copyright owner relinquishes all further rights to sell or otherwise dispose of that copy. The Supreme Court first adopted the first sale doctrine in the case of *Bobbs-Merrill Co. v. Straus*, 210 U.S. 339 (1908). In that case, the Supreme Court held that the exclusive right to sell copyrighted works only applied to the first sale of a copyrighted work. 210 U.S. at 349-350. While the copyright owner retained the underlying copyright to the expression fixed in the work, the copyright owner gave up his ability to control the fate of the work once it had been sold.

holder.[92] The Senate Judiciary Committee of the 105th Congress report made a point that the exemption for archives and libraries relates to their “functions as repositories of such works for public reference,” and reemphasized that the definition of “archives” and “libraries” as used in Section 108 is restricted to “such institutions only in the conventional sense of entities that are established as, and conduct their operations through physical premises in which collections of information may be used by researchers and other members of the public” and is not intended to apply to all websites, bulletin boards, and homepages on the Internet[93] While Section 108 contains a specific exemption for infringement for limited reproduction for preservation and replacement by archives and libraries, it falls short of allowing the type of systematic ongoing digital preservation activities as envisioned by the Preservation Task Force report. The statute’s exemption is narrowly worded requiring one criterion for published material and another for unpublished work. For example, an archives or library can only reproduce unpublished material if the archives or library already owns a copy in its collection and the purpose of the copying is for preservation and security purposes.

Prior to the passage of the DMCA, Section 108 permitted an archives or library to make and distribute one copy of an unpublished work for preservation, security or for deposit for research in another archives or library if reproduction was done “in facsimile form.”[94] According to the Senate Judiciary Committee’s report on DMCA, “the legislative history to that section makes clear that, when this language was enacted more than twenty years ago, Congress intended to permit the copy to be made by microfilm or electrostatic photocopying process, but not in computerized form.”[95] The amendment to Section 108(b) now permits an archives to make up to three copies and permits the copies to be made in digital or analog formats however, the statute restricts the use of any digital or electronic copies by providing any such copy of a work that the library or archive makes in a digital format must not otherwise be distributed in that format and must not be “made available in that format to the public outside the premises of the library or archives.”[96]

In contrast, to make a copy of a published work, an archives’ or library’s purpose can only be to replace a copy it has or used to have in its collection, where the copy “is damaged, deteriorating, lost, or stolen, or if the existing format of in which the work is stored has become obsolete.”[97] Such published works also must be out of print and the library or archives must have made a reasonable effort to determine “that an unused replacement could not be obtained at a fair price.”[98] The statute provides some guidance as to when a format can be considered obsolete by providing the following definition: “a format shall be considered obsolete if the machine or device necessary to render perceptible a work stored in that format is no longer manufactured or is no longer reasonably available in the commercial marketplace.”[99]

Section 101 of the Copyright Act defines a “derivative work” as “based on one or more preexisting works, such as translation, musical arrangement, dramatization, fictionalization, motion picture version, sound recording, art reproduction, abridgement, condensation, or any other form in which a work may be recast, transformed, or adapted, a work consisting of editorial revisions, annotations, elaborations, or other modifications which, as a whole, represent an original work of authorship, is a ‘derivative work.’”[100] Section 106(2) describes the copyright owners exclusive rights including, “to prepare derivative works based on the copyrighted work.” [101] Given the broad scope of protection and definition of derivative work to include “any other form in which the work may be recast” copyright owners could

[92] 17 U.S.C. 108.
[93] Senate Report, Section V: Section-by-section Analysis, p. 50.
[94] 17 U.S.C. 108 (b) 1976-1998.
[95] Senate Report, Section IV: Section-by section analysis, p. 50.
[96] 17 U.S.C. 108(b).
[97] 17 U.S.C. 108(c)
[98] 17 U.S.C. 108(c)(1).
[99] 17 U.S.C. 108(c)(2).
[100] 17 U.S.C.101.
[101] 17 U.S.C 106(2).

argue that some forms of digital preservation involving copyrighted work may create derivative work, which is also protected by the law.

While the DMCA provides for the additional quantity of digital copying for specific classes of copyrighted work by archives and museums it stops short of providing the type of blanket exemption for archives and libraries engaged in the systematic long term preservation of electronic records and collections. As the results of the InterPARES project begin to become operationalized and used by archives, libraries and museums, the Section 108 exemption for archives and libraries related to digital preservation may need to be revisited in order to provide archives and libraries the ability to preserve over time authentic digital records and collections.

b. The Sonny Bono Copyright Term Extension Act

The Sonny Bono Copyright Term Extension Act of 1998[102] retroactively extended the duration of copyright from the life of author plus 50 years to the life of the author plus seventy years, in the case of individual works, and from 75 years to 95 years in the case of works of corporate authorship and works first published before January 1, 1978. In essence, the term extension act preserves copyright protection for all covered works for an additional twenty years. Libraries and archives won a limited exception to use works in the last 20 years of protection. This provision permits use for purposes of preservation, scholarship or research if the work is not subject to normal commercial exploitation or is not available at a reasonable price.[103]

The Sonny Bono Copyright Term Extension Act is currently the subject of a challenge before the U.S. Supreme Court in the case of *Edred v. Ashcroft*, where the Court accepted *certiorari* of the case in February 2002 from a decision of the D.C. Circuit upholding Congress' right to extend the terms of already governing copyrights.[104] Plaintiffs represent commercial and non-commercial users of public domain works, including Eldred, who posts on the Internet works in the public domain, who claim injury due to the twenty-year extension of copyright.

c. Licensing

Over the past few years, licensing of digital products has become a fast growing avenue for protecting intellectual property as well as furthering market oriented electronic commerce of digital products.[105] In a legal context, licensing is governed by the law of contracts; however, there continue to be some unresolved questions as to whether licensing of copyrighted digital products can preempt federal copyright protection. Regardless of how the overarching issue of contract precedence or federal preemption is resolved, increasingly copyrighted software used in conjunction with electronic records governed by license agreements can present challenges to the implementation of InterPARES digital preservation strategies. For example, a license agreement that prohibits archiving and copying of proprietary software which is used solely for the purposes of viewing, reading, or using digital content of electronic records would restrict the ability for long-term preservation of the electronic record's digital content. To insure the long-term preservation of authentic digital records that require license agreements, archives, libraries,

[102] Pub. Law 105-298, 112 Stat. 2827 (1998), amending 17 U.S.C. 101, 302-305.

[103] 17 U.S.C. 108(h).

[104] *See Eldred v. Reno,* 239 F.3d 372 (D.C. Cir. 2001)

[105] Senate Report, Section III: Discussion (citing International Intellectual Property Alliance statistics that "the U.S. creative industries accounted for 3.65 percent of the U.S. gross domestic product (GDP)--$278.4 billion"), p. 9. *See also ARL Supplementary Statistics 1999-2000*. This survey report of Association of Research Libraries found licensed electronic resources accounted for 12.9% of the Library Materials budget. In 1999-2000, 105 ARL Libraries reported spending almost 100 million, *available at* <http://www.arl.org/stats/sup/sup00.pdf>.

and cultural institutions must incorporate into license agreements clauses that provide for the right to archive the digital content.

6. Electronic Discovery & Spoliation

A large and growing literature exists on the subject of the discovery in civil litigation of electronic documents, including e-mail, word processing, spreadsheets, web page audit trails, and other records created and received during the course of business activities.[106] Without question, records in electronic form (including electronically captured associated information or metadata) are increasingly targets of discovery.

For more than thirty years, the Federal Rules of Civil Procedure have contemplated that parties may request documents in the form of "data compilations from which information can be obtained."[107] The further 1993 and 2000 changes in the Federal Rules setting forth a uniform initial disclosure rule (requiring early disclosure of information supporting a party's claims or defenses, including key documents), only increase the scrutiny both sides apply in litigation to the failure to turn over records in electronic form. Initial disclosure pursuant to Rule 26(a)(1) assumes that litigants will undertake reasonable measures to assess what forms of electronic evidence exist in support of their case, including (implicitly) taking steps to preserve such evidence.

Rule 26 also sets forth limits on what constitutes reasonable discovery.[108] In 1993, the Advisory Committee to the Federal Rules noted the "information explosion of recent decades has greatly increased both the potential cost of wide-ranging discovery and the potential for discovery to be used as an instrument of delay or oppression." Arguably, discovery demanding that an opponent in litigation conduct resource-intensive searches to recover deleted files and/or data on backup tapes has the potential to be considered abusive under the Rules.[109] In recent years, some federal courts have engaged in "routinizing" what constitutes a reasonable and acceptable set of joint-inspection procedures, for parties to mutually utilize when exchanges of computerized information are in order and where preservation of evidence concerns are present.[110] Although as a usual matter each side bears its own discovery costs, courts will engage in cost-shifting when there is a large resource burden attached to responding to discovery requests involving computerized records.[111]

[106] *See, e.g.*, Martin H. Redish, *Electronic Discovery and the Litigation Matrix*, 51 Duke L.J. 561 (2001); Devin Murphy, *The Discovery of Electronic Data in Litigation: What Every Practitioner Needs to Know*, 27 Wm. Mitchell L. Rev 1825 (2001); James J. Marcellino & Anthony A. Bongiorno, *E-Mail Is the Hottest Topic in Discovery Disputes: One Litigant Seeks Facts Buried in a Data Base; the Other Seeks to Avoid Burdens of Production*, Nat'l L.J., Nov. 3, 1997.

[107] F.R.C.P. 34(a).

[108] F.R.C.P. 26(b)(2)(i)-(iii).

[109] *See McPeek v. Ashcroft*, 202 F.R.D. 31, 32 (D.D.C. 2001) ("The purpose of having a backup system and retaining the tapes was to permit recovery from a disaster, not archival preservation."); *cf.* 36 C.F.R. 1234.24(c) (NARA regulations stating "backup tapes should not be used for recordkeeping purposes.")

[110] See, e.g., Rowe Entertainment v. The William Morris Agency, Inc., 205 F.R.D. 421 (S.D.N.Y. 2002); Simon Property Group, L.P. v. mySimon, Inc., 194 F.R.D. 639 (S. D. Ind 2000); Playboy Enterprises, Inc. v. Welles, 60 F.Supp.2d 1050 (S.D. Cal. 1999) (adopting use of "mirror images" using computer forensic procedures and experts trained in evaluating them to preserve the integrity of original evidence). See also model standards proposed in 1999 by the American Bar Association's Section of Litigation (1999), which provide more detailed guidance with respect to the discovery of electronic information.

[111] *See Anti-Monopoly v. Hasbro*, 1996 WL 22976 (S.D.N.Y. 1996); *Simon Property Group, supra*. *But cf. In re Brand Name Prescription Drug Antitrust Litigation*, 1995 WL 3360526 (N.D. Ill. 1995) (court held that $50,000 to $70,000 cost of searching 30 million e-mail messages was held not to constitute undue burden and costs would not be shifted to the requestor). *See generally*, Corinne L. Giacobbe, *Allocating Discovery Costs in the Computer Age: Deciding Who Should Bear the Costs of Discovery of Electronically Stored Data*, 57 Wash. & Lee L.Rev. 257 (2000).

Courts also deal with the intentional (and negligent) destruction of evidence under (i) the rubric of discovery sanctions, or (ii) by inferring "spoliation," and (iii) more recently, utilizing spoliation tort theories of recovery.[112] Once a relatively rare phenomenon in the paper-based world, the law on destruction of evidence and spoliation has become transformed because of the ease by which electronic records are destroyed in the normal course of business. Nevertheless, the exponential increase in e-mail communications, coupled with the ubiquitousness of other forms of electronically stored information preserved within a corporate entity (or federal agency), renders problematic the goal of 100 percent preservation ever being achieved.

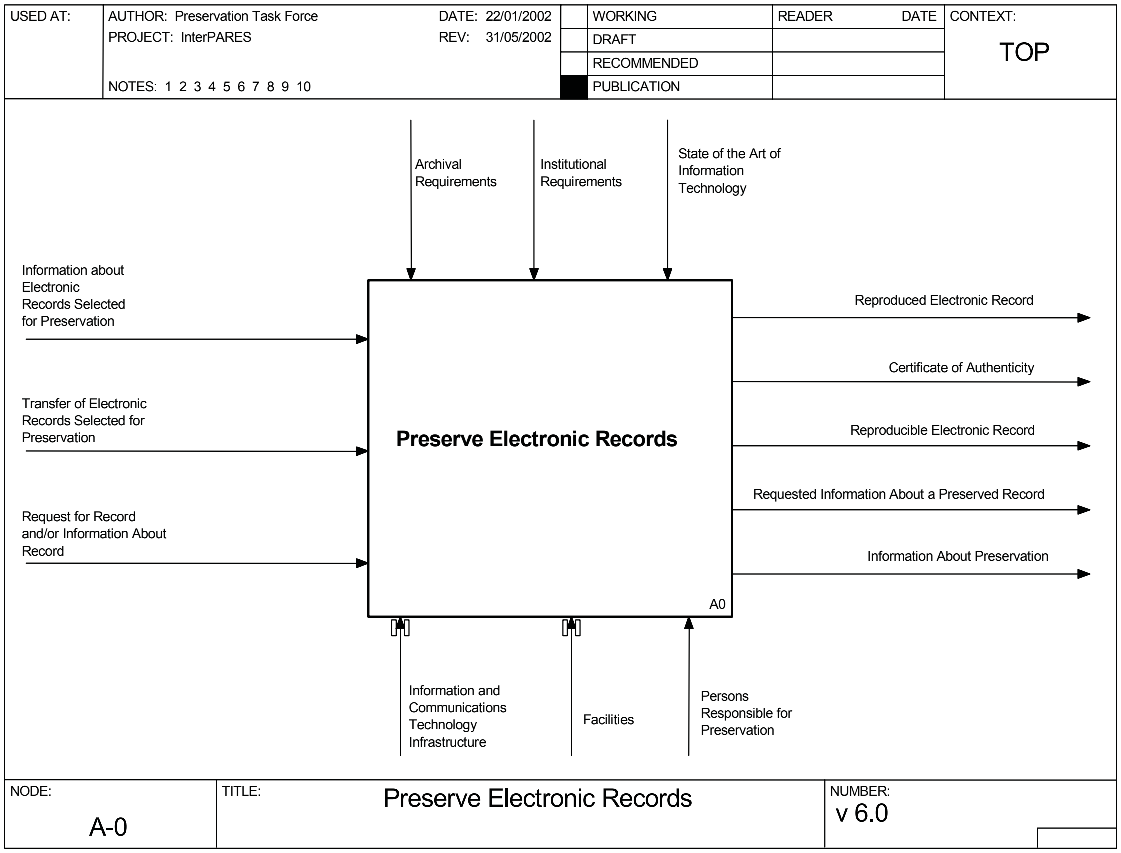
In light of the phenomenon of modern discovery practice, and especially post-Enron, enterprises must give serious consideration to distinguishing what records are appropriate for preservation in electronic form (and for how long), and to taking steps to put in place a records retention and management scheme for the archiving and preservation of key documents which explicitly addresses authenticity issues. Organizing and preserving records consistent with business purposes (including but not limited to using standardized metadata) minimizes litigation risks when faced with burdensome discovery, while at the same time maximizes the possibility of access and retrieval in response to such discovery.

Given the nature of major litigation, it seems certain that it is only a matter of time before authenticity issues such as those addressed in the InterPARES findings will arise during civil discovery (to be more fully addressed at a subsequent stage of litigation) for electronic records that will have been preserved for increasingly lengthy periods of time.

[112] *See generally*, Jamie S. Gorelick, Stephen Marzen, Lawrence Solum, *Destruction of Evidence* (New York: John Wiley & Sons, 1989), chapters 1-4, *with* 2001 Cumulative Supplement.

Appendix D. Preservation Models

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 <http://www.idef.com> for more information.
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.



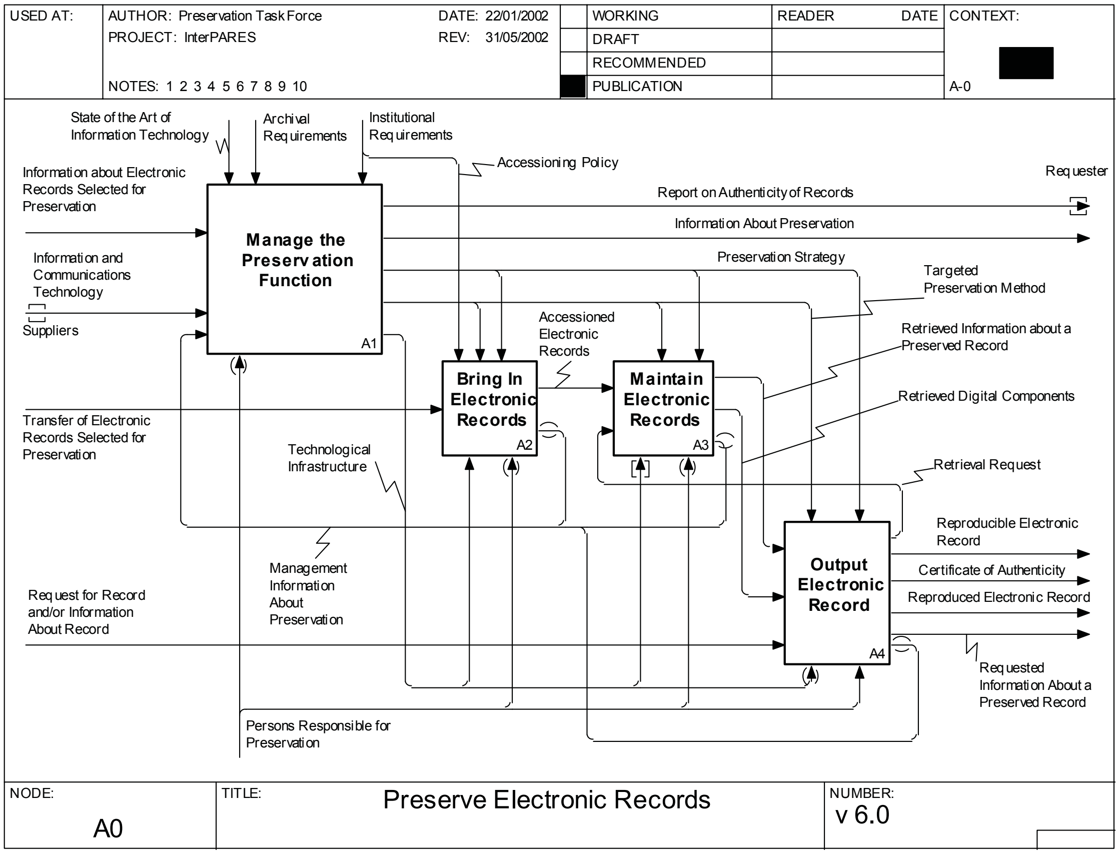
USED AT:
AUTHOR: Preservation Task Force
PROJECT: InterPARES
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 22/01/2002
REV: 31/05/2002

□ WORKING
□ DRAFT
□ RECOMMENDED
■ PUBLICATION

READER DATE
CONTEXT: TOP

NODE: A-0
TITLE: Preserve Electronic Records
NUMBER: v 6.0

Archival Requirements
Institutional Requirements
State of the Art of Information Technology
Information about Electronic Records Selected for Preservation
Transfer of Electronic Records Selected for Preservation
Request for Record and/or Information About Record
Preserve Electronic Records
A0
Reproduced Electronic Record
Certificate of Authenticity
Reproducible Electronic Record
Requested Information About a Preserved Record
Information About Preservation
Information and Communications Technology Infrastructure
Facilities
Persons Responsible for Preservation



USED AT:
AUTHOR: Preservation Task Force
PROJECT: InterPARES
NOTES: 1 2 3 4 5 6 7 8 9 10
DATE: 22/01/2002
REV: 31/05/2002

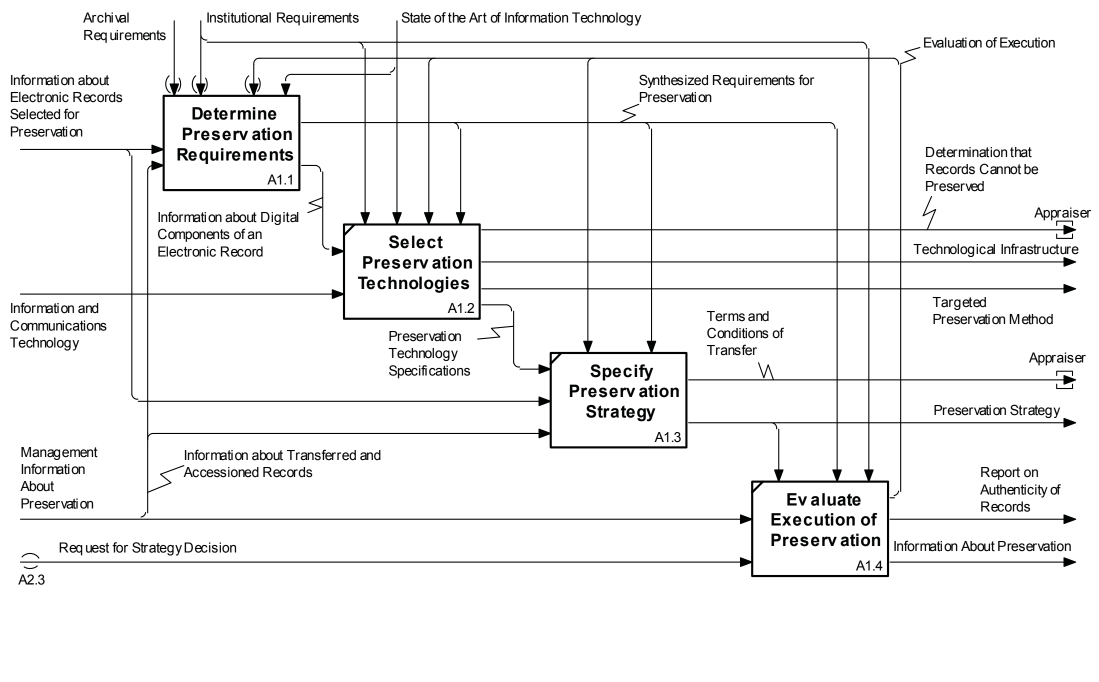
□ WORKING
□ DRAFT
□ RECOMMENDED
■ PUBLICATION

READER DATE
CONTEXT: A-0

NODE: A0
TITLE: Preserve Electronic Records
NUMBER: v 6.0

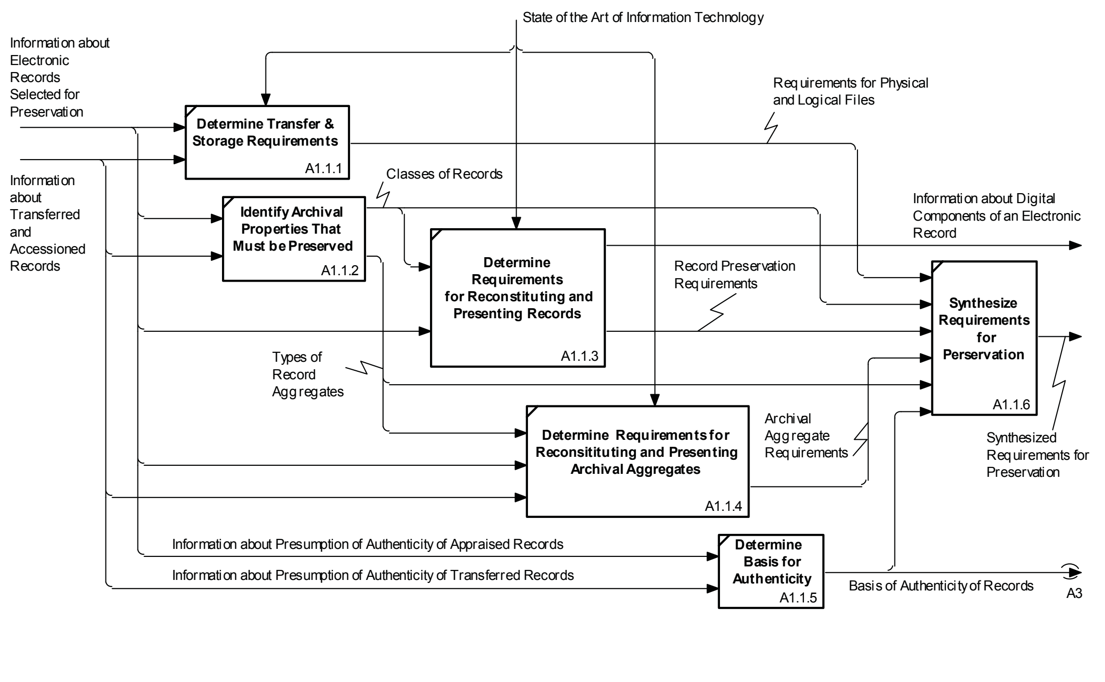
State of the Art of Information Technology
Archival Requirements
Institutional Requirements
Accessioning Policy
Information about Electronic Records Selected for Preservation
Information and Communications Technology
Suppliers
Manage the Preservation Function
A1
Requester
Report on Authenticity of Records
Information About Preservation
Preservation Strategy
Targeted Preservation Method
Accessioned Electronic Records
Retrieved Information about a Preserved Record
Bring In Electronic Records
A2
Maintain Electronic Records
A3
Retrieved Digital Components
Transfer of Electronic Records Selected for Preservation
Technological Infrastructure
Retrieval Request
Reproducible Electronic Record
Certificate of Authenticity
Reproduced Electronic Record
Output Electronic Record
A4
Management Information About Preservation
Request for Record and/or Information About Record
Requested Information About a Preserved Record
Persons Responsible for Preservation

USED AT:	AUTHOR: Preservation Task Force	DATE: 23/01/2002		WORKING	READER	DATE	CONTEXT:
	PROJECT: InterPARES	REV: 31/05/2002		DRAFT			
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10		■	PUBLICATION			A0

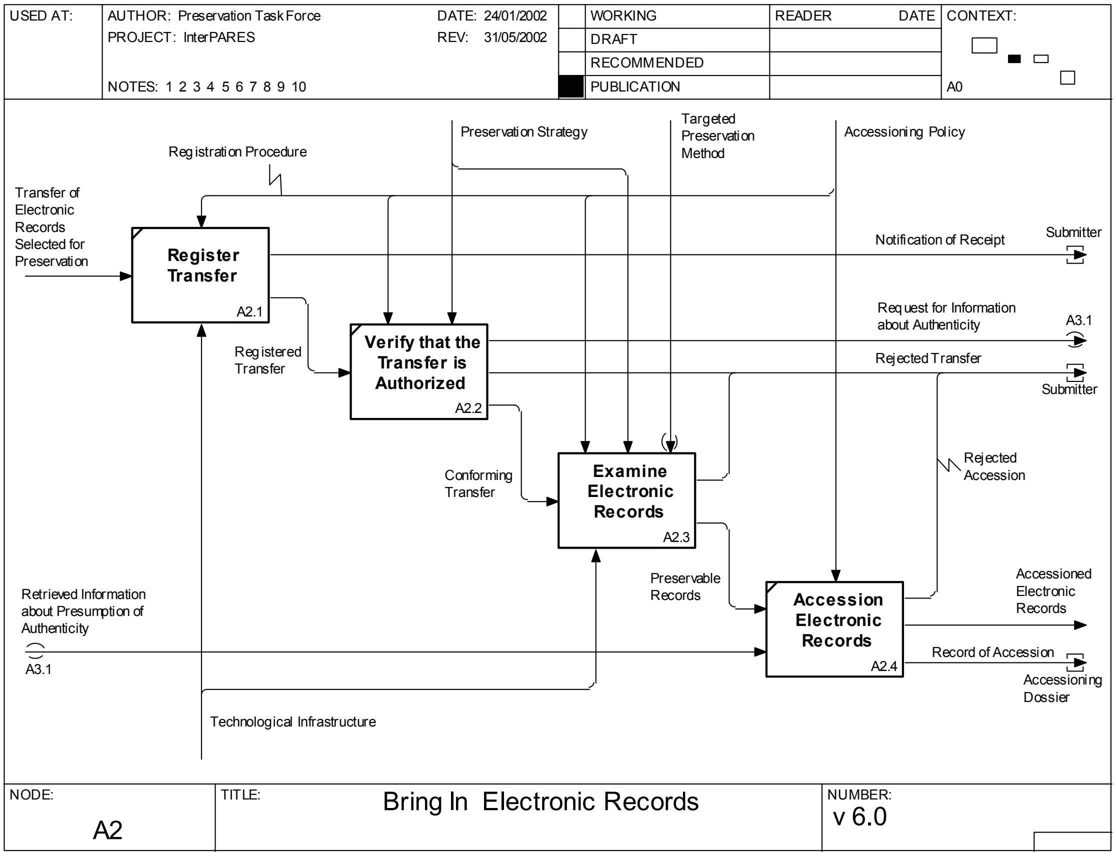

Archival Requirements
Institutional Requirements
State of the Art of Information Technology
Evaluation of Execution
Information about Electronic Records Selected for Preservation
Synthesized Requirements for Preservation
Determine Preservation Requirements
A1.1
Determination that Records Cannot be Preserved
Information about Digital Components of an Electronic Record
Appraiser
Select Preservation Technologies
Technological Infrastructure
Information and Communications Technology
A1.2
Targeted Preservation Method
Terms and Conditions of Transfer
Preservation Technology Specifications
Appraiser
Specify Preservation Strategy
Preservation Strategy
A1.3
Management Information About Preservation
Information about Transferred and Accessioned Records
Report on Authenticity of Records
Evaluate Execution of Preservation
Request for Strategy Decision
Information About Preservation
A1.4
A2.3

NODE: A1	TITLE: Manage the Preservation Function	NUMBER: v 6.0

USED AT:	AUTHOR: Preservation Task Force	DATE: 24/01/2002		WORKING	READER	DATE	CONTEXT:
	PROJECT: InterPARES	REV: 06/02/2002		DRAFT			
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10		■	PUBLICATION			A1


State of the Art of Information Technology
Information about Electronic Records Selected for Preservation
Requirements for Physical and Logical Files
Determine Transfer & Storage Requirements
A1.1.1
Information about Transferred and Accessioned Records
Classes of Records
Identify Archival Properties That Must be Preserved
A1.1.2
Information about Digital Components of an Electronic Record
Determine Requirements for Reconstituting and Presenting Records
A1.1.3
Record Preservation Requirements
Synthesize Requirements for Perservation
A1.1.6
Types of Record Aggregates
Archival Aggregate Requirements
Synthesized Requirements for Preservation
Determine Requirements for Reconsitituting and Presenting Archival Aggregates
A1.1.4
Information about Presumption of Authenticity of Appraised Records
Information about Presumption of Authenticity of Transferred Records
Determine Basis for Authenticity
A1.1.5
Basis of Authenticity of Records
A3

NODE: A1.1	TITLE: Determine Preservation Requirements	NUMBER: v 6.0



USED AT:

AUTHOR: Preservation Task Force

PROJECT: InterPARES

DATE: 24/01/2002

REV: 31/05/2002

NOTES: 1 2 3 4 5 6 7 8 9 10

WORKING

DRAFT

RECOMMENDED

■ PUBLICATION

READER

DATE

CONTEXT:

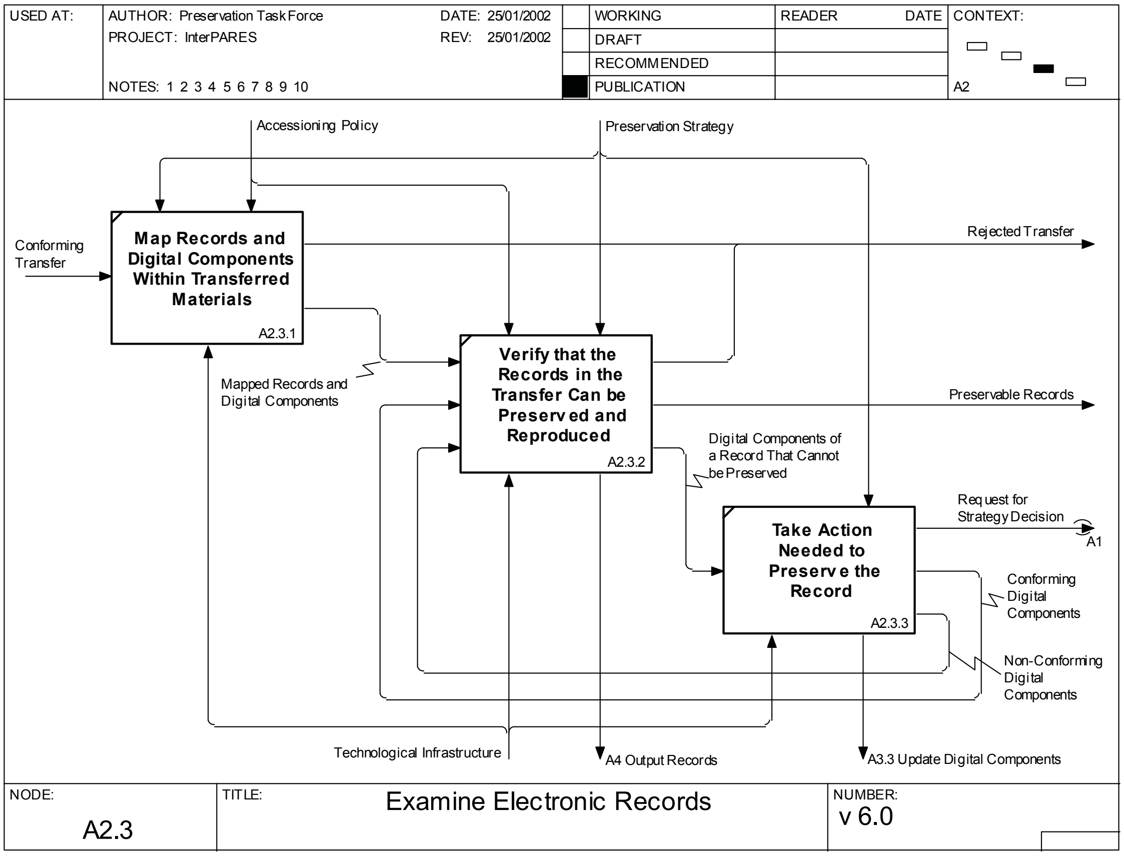
A0

NODE: A2

TITLE: Bring In Electronic Records

NUMBER: v 6.0

Registration Procedure
Preservation Strategy
Targeted Preservation Method
Accessioning Policy
Transfer of Electronic Records Selected for Preservation
Register Transfer
A2.1
Notification of Receipt
Submitter
Registered Transfer
Verify that the Transfer is Authorized
A2.2
Request for Information about Authenticity
A3.1
Rejected Transfer
Submitter
Conforming Transfer
Examine Electronic Records
A2.3
Rejected Accession
Preservable Records
Accession Electronic Records
A2.4
Accessioned Electronic Records
Record of Accession
Accessioning Dossier
Retrieved Information about Presumption of Authenticity
A3.1
Technological Infrastructure



USED AT:

AUTHOR: Preservation Task Force

PROJECT: InterPARES

DATE: 25/01/2002

REV: 25/01/2002

NOTES: 1 2 3 4 5 6 7 8 9 10

WORKING

DRAFT

RECOMMENDED

■ PUBLICATION

READER

DATE

CONTEXT:

A2

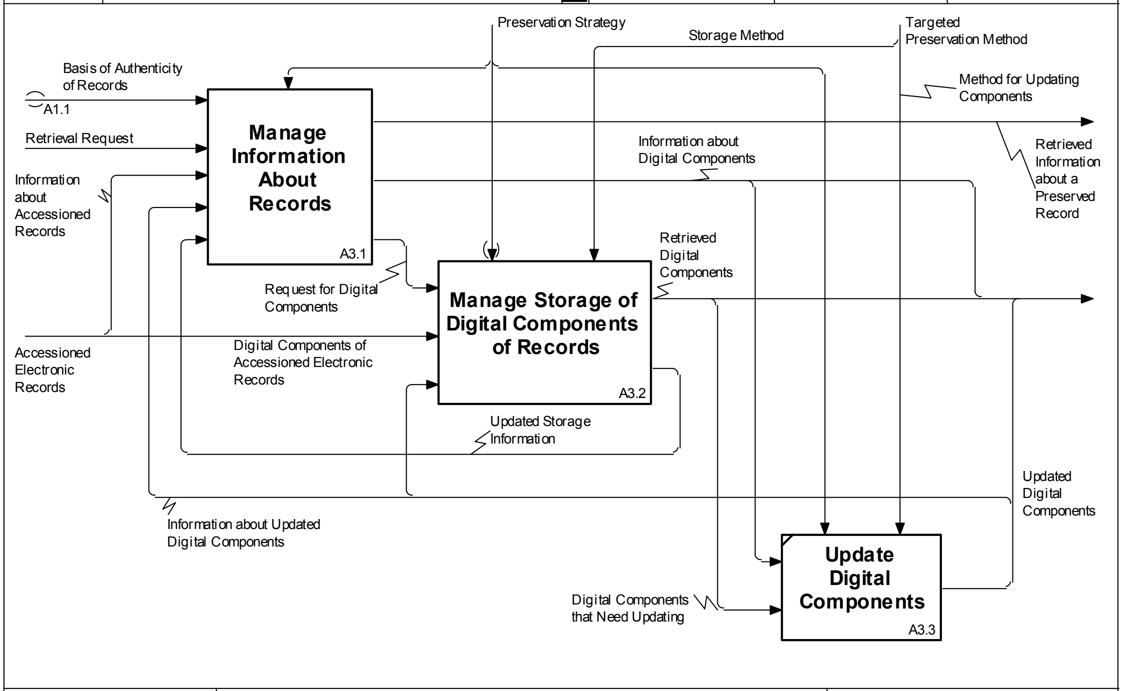
NODE: A2.3

TITLE: Examine Electronic Records

NUMBER: v 6.0

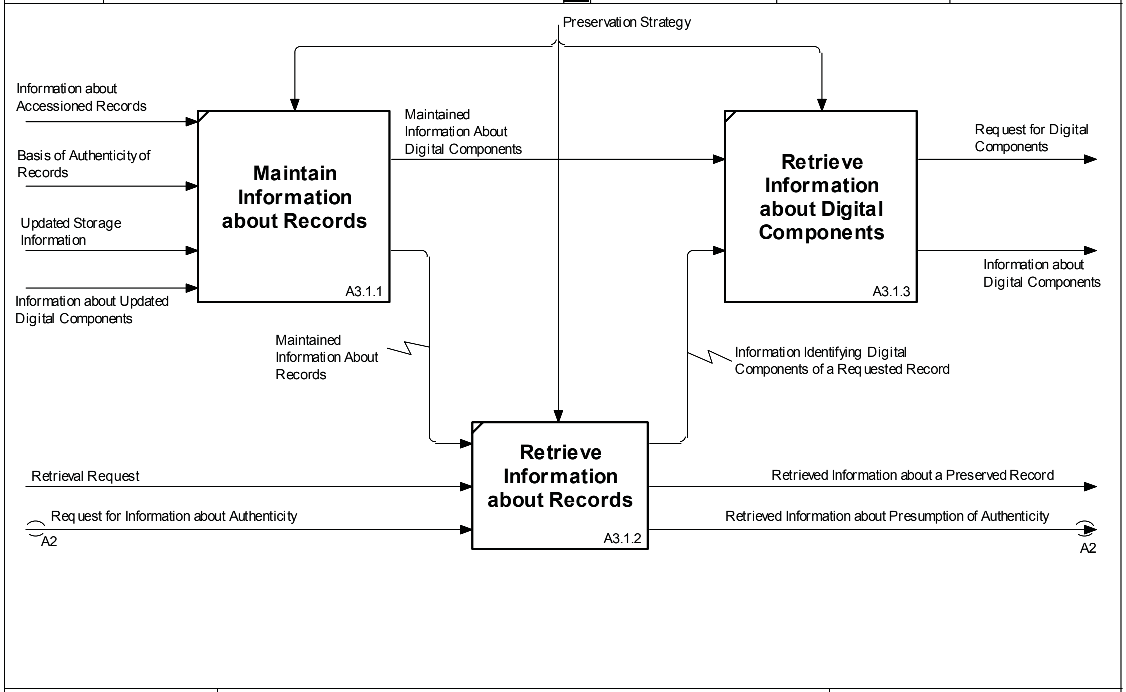
Accessioning Policy
Preservation Strategy
Conforming Transfer
Map Records and Digital Components Within Transferred Materials
A2.3.1
Rejected Transfer
Mapped Records and Digital Components
Verify that the Records in the Transfer Can be Preserved and Reproduced
A2.3.2
Preservable Records
Digital Components of a Record That Cannot be Preserved
Take Action Needed to Preserve the Record
A2.3.3
Request for Strategy Decision
A1
Conforming Digital Components
Non-Conforming Digital Components
Technological Infrastructure
A4 Output Records
A3.3 Update Digital Components

USED AT:	AUTHOR: Preservation Task Force	DATE: 28/01/2002		WORKING	READER	DATE	CONTEXT:
	PROJECT: InterPARES	REV: 31/05/2002		DRAFT			
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10		■	PUBLICATION			A0


Preservation Strategy
Storage Method
Targeted Preservation Method
Basis of Authenticity of Records
A1.1
Method for Updating Components
Retrieval Request
Manage Information About Records
Information about Digital Components
Retrieved Information about a Preserved Record
Information about Accessioned Records
A3.1
Retrieved Digital Components
Request for Digital Components
Manage Storage of Digital Components of Records
Accessioned Electronic Records
Digital Components of Accessioned Electronic Records
A3.2
Updated Storage Information
Updated Digital Components
Information about Updated Digital Components
Update Digital Components
Digital Components that Need Updating
A3.3

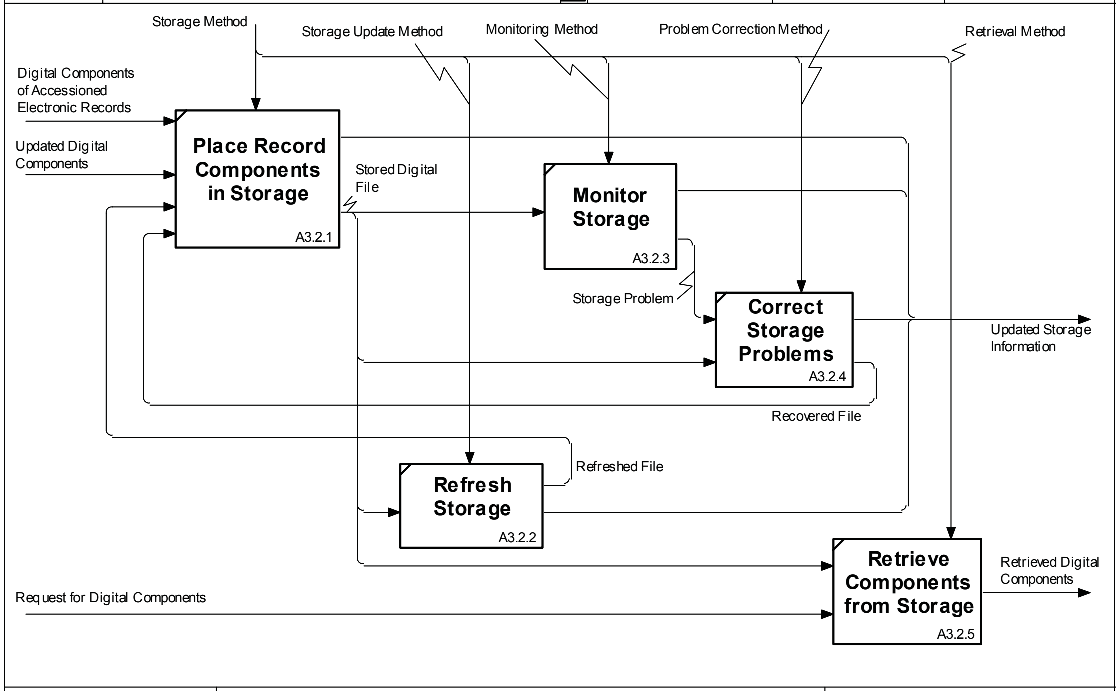
NODE: A3	TITLE: Maintain Electronic Records	NUMBER:

USED AT:	AUTHOR: Preservation Task Force	DATE: 30/01/2002		WORKING	READER	DATE	CONTEXT:
	PROJECT: InterPARES	REV: 31/01/2002		DRAFT			
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10		■	PUBLICATION			A3


Preservation Strategy
Information about Accessioned Records
Maintained Information About Digital Components
Request for Digital Components
Basis of Authenticity of Records
Maintain Information about Records
Retrieve Information about Digital Components
Updated Storage Information
Information about Digital Components
Information about Updated Digital Components
A3.1.1
A3.1.3
Maintained Information About Records
Information Identifying Digital Components of a Requested Record
Retrieve Information about Records
Retrieval Request
Retrieved Information about a Preserved Record
Request for Information about Authenticity
A2
Retrieved Information about Presumption of Authenticity
A2
A3.1.2

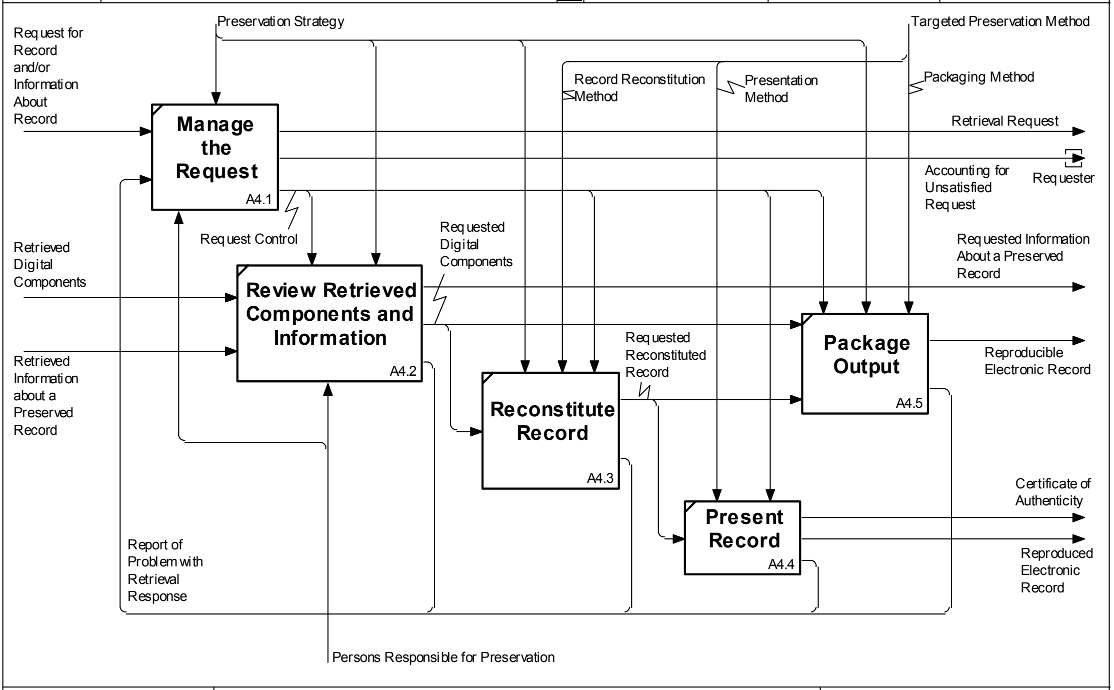
NODE: A3.1	TITLE: Manage Information About Records	NUMBER:

USED AT:	AUTHOR: Preservation Task Force	DATE: 30/01/2002		WORKING	READER	DATE	CONTEXT:
	PROJECT: InterPARES	REV: 31/01/2002		DRAFT			
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10		■	PUBLICATION			A3


Storage Method
Storage Update Method
Monitoring Method
Problem Correction Method
Retrieval Method
Digital Components of Accessioned Electronic Records
Updated Digital Components
Place Record Components in Storage
A3.2.1
Stored Digital File
Monitor Storage
A3.2.3
Storage Problem
Correct Storage Problems
A3.2.4
Updated Storage Information
Recovered File
Refreshed File
Refresh Storage
A3.2.2
Retrieve Components from Storage
A3.2.5
Retrieved Digital Components
Request for Digital Components

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

USED AT:	AUTHOR: Preservation Task Force	DATE: 30/01/2002		WORKING	READER	DATE	CONTEXT:
	PROJECT: InterPARES	REV: 31/05/2002		DRAFT			
				RECOMMENDED			
	NOTES: 1 2 3 4 5 6 7 8 9 10		■	PUBLICATION			A0


Preservation Strategy
Targeted Preservation Method
Request for Record and/or Information About Record
Record Reconstitution Method
Presentation Method
Packaging Method
Manage the Request
A4.1
Retrieval Request
Accounting for Unsatisfied Request
Requester
Request Control
Requested Digital Components
Retrieved Digital Components
Review Retrieved Components and Information
A4.2
Requested Information About a Preserved Record
Retrieved Information about a Preserved Record
Package Output
A4.5
Requested Reconstituted Record
Reproducible Electronic Record
Reconstitute Record
A4.3
Certificate of Authenticity
Present Record
A4.4
Reproduced Electronic Record
Report of Problem with Retrieval Response
Persons Responsible for Preservation

NODE:	TITLE:	NUMBER:
A4	Output Electronic Record	

Appendix E. Bibliography of Publications & Papers

Authentic Records in the Electronic Age: Proceedings of the InterPARES Symposium. Beijing, China, State Archives Administration of China, November 14-15, 2001. [Includes papers by Cloonan and Sanett, Eppard, Gilliland-Swetland, and Underwood.]

Baron, Jason R. "E-Mail Litigation Wars: The U.S. National Archivist Strikes Back," *Authentic Records in the Electronic Age: Proceedings of an International Symposium*, edited by Luigi Sarno (Vancouver: InterPARES Project and Istituto Italiano di Cultura Vancouver, 2000): 156-67.

Chen, Su-Shing. "The Paradox of Digital Preservation," *Computer* 34(March, 2001): 24-28.

Cloonan, Michèle Valerie and Shelby Sanett. "Comparing Preservation Strategies and Practices for Electronic Records." *New Review of Academic Librarianship* 6 (2000): 205-216.

Cloonan, Michèle V. and Shelby Sanett. "Preservation Strategies for Electronic Records: Where We are Now—Obliquity and Squint?" *American Archivist* 65 (Spring/Summer 2002): 70-106.

Eppard, Philip B. "Electronic Records and Democratic Rights: Historical Perspectives and Current International Research," *Approaching a New Millennium, Lessons from the Past, Prospects for the Future: Proceedings of the 7th Conference of the International Society for the Study of European Ideas* (Bergen, Norway: HIT Centre at the University of Bergen, 2000).

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

Gilliland-Swetland, Anne J. "Managing and Providing Access to Electronic Evidence: The U.S. Experience," *Irish Archives* (in press).

Gilliland-Swetland, Anne J. "The Potential of Markup Languages to Support Descriptive Access to Electronic Records: The EAD Standard." *Archivi & computer* 2 (2001): 110-121.

Gilliland-Swetland, Anne J. "Securing Our Identities in the Age of Digital Recordkeeping: The Role of Reliable and Authentic Records in the Establishment and Preservation of Human Rights." *Approaching a New Millennium, Lessons from the Past, Prospects for the Future: Proceedings of the 7th Conference of the International Society for the Study of European Ideas* (Bergen, Norway: HIT Centre at the University of Bergen, 2000).

Gilliland-Swetland, Anne J. "Testing Our Truths: Delineating the Parameters of the Authentic Archival Electronic Record." *American Archivist* 65 (Fall/Winter 2002) (in press).

Gilliland-Swetland, Anne J. and Philip B. Eppard. "Preserving Authentic Electronic Records: The InterPARES Project." *Annotations: National Historical Publications and Records Commission Newsletter*, December 2000, 7-8.

Gilliland-Swetland, Anne J. and Philip B. Eppard. "Preserving the Authenticity of Contingent Digital Objects: The InterPARES Project," *D-Lib Magazine* 6, no. 7/8 (2000). Available at: <http://www.dlib.org/dlib/july00/eppard/07eppard.html>.

InterPARES Interpreted: A Guide to Findings on the Preservation of Authentic Electronic Records. US-InterPARES Project, June 2002.

Park, Eun G. "Developing a Framework for Authenticity Requirements in University Student Records Systems: An Exploratory Study." Doctoral dissertation. University of California, Los Angeles, 2002.

Park, Eun G. "The Significance of Ensuring the Authenticity of Records in Student Information Records Systems: Methodology and Implications," *Proceedings of the First PhD Forum on Archives*, Renmin University, Beijing, China, 2001.

Park, Eun G. "Understanding Authenticity in Records and Information Management: Analyzing Practitioner Constructs." *American Archivist* 64 (fall/winter 2001): 270-91.

Preserving Authentic Electronic Records: Preliminary Research Findings. Proceedings from an International Symposium. Edited by Luigi Sarno. University of British Columbia, February 17, 2001. [Includes papers by Eppard, Farb, Gilliland-Swetland, Thibodeau, and Underwood.] Available at: <http://www.interpares.org/reports.htm>.

Sanett, Shelby. "Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity." *College and Research Libraries* 63(September 2002): 388-404.

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

Shankar, Kalpana. "Preserving the Digital Record: An Introduction to the InterPARES Project." *Storage Management Solutions* 5, no. 1 (2000): 31-33.

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

Underwood, William E. "Diplomatic Analysis of Electronic Military Messages." *Archivi per la storia*, XII, 1-2, (December1999): 121-29.