Appendix 6

How to Preserve Authentic Electronic Records

Preservation Task Force

8 October 2001

Introduction

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

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

Electronic Records

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

Recording Information

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

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

It is not possible to store an electronic record in the form of a record. The "form of a record" is the

---

[1] <http://www.interpares.org/researchplan.htm>

[2] <http://is.gseis.ucla.edu/us-interpares/bib_pres.htm>

[3] P. C. Hariharan. Media, A Presentation for the InterPARES Panel at UCLA. 2 December 1999.

documentary form that enables it to achieve its intended purpose in the first place and to communicate the same information over time. Electronic records are stored in forms that differ substantially from those in which they can serve their intended purpose as records.

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

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

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

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

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

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

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

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

One technique that is used to ensure the fixity of both the content and the appearance of digital documents is to convert them from word-processing files to formats, such as portable document format (PDF) files, which include the data necessary to form font characters properly, eliminating dependence on dynamic link libraries.

While the data content of a record may be stored in many different tables in a relational database, the data must be brought together in order to present the record in its proper documentary form. Once the record is reconstituted from its different data components, it may be possible, and in fact simplest, to save it as a single digital file, for example as a word-processing file, without in any way impacting its identity or integrity.

Because there is no necessary mapping between digital files and records, in many cases it is possible to change the way the record is stored in digital files without changing the record itself.

The inevitable, and often repeated, changes between the storage representation and the presentation for use of digital information create paradoxical elements in the preservation of electronic records.

Keeping Electronic Records

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

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

Keeping electronic records is more complex and difficult. Like all records, an electronic record must transmit the message intended by its creator; however, the fixity of the message carried by an electronic record is at risk because of the changes between the way it is represented in storage and the way it is

presented for use. Both placing an electronic record in storage and retrieving it for use entail transforming the way the content, structure, and appearance of the record are inscribed on a physical medium. Both storage and retrieval transformations create risks that the record may be altered. These risks are compounded by the software, hardware, and the physical media used, including both the storage media and the media on which the records are presented in record form. While maintenance and conservation of the stored information remain essential, the integrity of an electronic record depends on guaranteeing that none of the changes between storage representation and presentation for use, in either direction, has altered the message the record was intended to convey from the time it was first filed as a record to the point of any subsequent use. The necessity of ensuring that transformations between storage and use do not corrupt the records adds a new focus to the preservation of electronic records. An electronic record cannot be said to have been preserved unless it can be delivered in authentic form.

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

Foundation Concepts

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

Digital Components of Electronic Records

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

Every electronic record has at least one digital component: a stream of bits representing information contained in the record. Each digital component has one or more associated methods, or programs, for decoding the bitstream and/or presenting it for use. In the simple case of a record with a single component, the record and the digital component are congruent: the component *is* the record. However, the contents of a record may be stored in several bitstreams; for example, a single record may be stored as a compound document, with different parts of the record stored in different digital files. In such cases, the record has as many digital components as it has bitstreams. When a record has more than one component, each digital component *is part of* the record. Digital components may be distinguished from one another based on the fact that they are stored separately. If a physical file contains all of the components of an electronic record, it *contains* the record. If a physical file contains some, but not all of the components, it *contains parts of* the record.

Physical storage as a criterion for identifying digital components is a containment relationship: a physical file contains all or part of one or more records. In fact, the containment relationship may be either physical or logical, or both. An electronic record may contain one or more distinct digital components stored entirely within a file (physical), or a file may only contain links pointing to distinct components stored elsewhere (logical). For example, in a relational database, a record—in the archival sense—is likely to

comprise data "stored" in several different database tables. Each table that contains part of the record is a digital component of that record. Relational database management systems typically store all of the tables that constitute a database in a single physical file. Nonetheless, reconstructing the record with all of its data content requires locating the row or rows in each table that contains such data and retrieving the record content from it. Thus the logical tables are distinct digital components.

Digital components may be identified based on other criteria besides separate storage or containment relationships. A *record* may consist of many digital components, which may be stored together in one physical file. A *textual record* may contain non-textual content, such as a spreadsheet, an image or even a voice annotation. In such cases, it does not matter whether the different types of content are stored together or separately. Whether the units of non-textual content constitute separate digital components depends on the format or data-type used to represent the contents. In a word-processing file, a picture would constitute a distinct digital component because the word-processing content is character-type data while the picture is binary or raster data. However, if the textual record was a scanned document image, both the picture and the text would be binary image data and constitute only a single digital component. Thus, *data-type* appears as a second criterion for distinguishing digital components. However, this criterion needs to be refined. Specifically, following the definition of digital component given above, a digital component has one or more methods associated with it. A word-processing document and a spreadsheet both use character data, but each requires different software for proper processing and presentation. Conversely, alphanumeric characters and linear graphics are two different types of data, but most word-processing applications are capable of vertical and horizontal lines; therefore, a word-processing document containing only text and such graphic lines has only one digital component determined by data-type. Data-type is a criterion for distinguishing digital components of records if and only if different data-types have different *methods* associated with them.

An *electronic record* may also have one or more distinct digital components necessary to render the record correctly. A bitstream may contain data that indicate how the information content is to be presented for use; for example, codes indicating fonts, type sizes, line endings, paragraph indentations, etc. However, codes indicating presentation features may also be stored in separate bitstreams, such as dynamic link libraries, report templates, or style sheets. Each bitstream indicating presentation features is an additional digital component of the record. In such cases, the components containing presentation specifications may be processed by the same software used to process the bits representing the content of the record. Even though the same link library or other such file may be used in many different documents, each such file required to present a record correctly should be considered as a digital component of the record. These components are distinguished on the basis of separate storage.

Both containment and data-type criteria may apply to the same record. A word-processing file containing two digital photographs has only two data-types, but three digital components.

In sum, each digital component is a logical or physical object that the system processes as an unit. In other words, each component has a specified method or *methods*. In fact, when several components are stored together in a physical file, there must be specified methods for locating and extracting the components. Thus the physical file is itself a component, in addition to the components it contains.

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

Preservation Control

A technological boundary exists between any two states of a system or of interoperating systems when the transition from one state to another does, or can, entail significant changes in the attributes or methods of a digital object. For records, significant changes are those that affect identity or integrity. Technological boundaries exist at macro and micro levels. Macro level boundaries occur at the interfaces between systems, subsystems or applications, such as during system, media, or data format migrations

or in transfers between the "live" systems in which the records are created, and other applications in which they are transmitted over space or stored over time. Micro level boundaries occur when a record is decomposed into separate digital components or is reconstituted from its components, and when different methods are invoked to process distinct components. Transitions from storage representation to presentation for use can involve both macro and micro boundaries.

Preservation control is critical in transitions across technological boundaries. Preservation control consists of actions, conditions, and constraints designed to ensure the preservation of records and their continued authenticity. While preservation controls during maintenance of the records in storage must be adequate and effective, the risks of corruption or loss of records are more frequent and complex during transitions across technological boundaries. Thus preservation controls can be divided into two types: *systemic controls* are those that ensure records remain unchanged over time within a given system or subsystem; *dynamic controls* are those that ensure records remain authentic across technological boundaries.

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

Archival Requirements for Preservation

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

The requirement for records to remain unchanged, thus, needs to be qualified. Even in the case of hard-copy records, this requirement effectively means that the record should not be changed in any way that relates to its essential record nature, rather than to its existence as a physical object. For electronic records, this qualification has been aptly stated in the InterPARES report *Requirements for Assessing and Maintaining the Authenticity of Electronic Records*: “When we refer to an electronic record, we consider it essentially intact and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered.”[4]

The archival requirement for integrity, and therefore for authenticity, depends on the message intended by the record creator. The interpretation of a record depends on the reader as much as the creator; therefore it is beyond the control of the record preserver. Whatever interpretation is made, it must be consistent with the creator’s intended communication, and that undeniably involves the information content of the record, and it may involve the way the content is presented. Here again, neither requirement is absolute. If an accident, such as water damage, resulted in the loss or blurring (content or presentation) of a few words in a document, we would say that the record is damaged, but not that it is inauthentic. The strength of the requirement for unaltered content and presentation depends on the intended use of the record. The requirement is greatest when the user wants to see the "original," regardless of interpretation. But for many valid uses, it is sufficient if the content—of all or part of the record—is substantially intact and unaltered and its presentation is basically consistent with the original.[5] If this were not true, it would be impossible for historians, political scientists, and others to cite *original*

[4] Requirements for Assessing and Maintaining the Authenticity of Electronic Records, Appendix 2.

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

sources in analytic works.

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

"Original" Electronic Records

The transformations entailed by storage, retrieval, and presentation of electronic records make the concept of *original record* of uncertain applicability in the domain of electronic records. The traditional concept of an original record is tightly coupled with its inscription on a specific physical medium. The same record inscribed on any other unit of physical medium is not considered to be an original, but a copy. Given that electronic records are not stored even on the same type of medium used to present them to humans,[6] and that the physical inscription of the record on a storage medium is fundamentally different from the inscription on a display device, a strict application of the traditional concept of original record would mean that the original ceases to exist at the moment it is committed to storage and deleted from the video device the writer used to create it. Therefore, the closest we can come to an electronic original, once the record has been set aside, is a copy in the form of the original. Given that electronic records are not stored in their original form, to produce a copy in the form of the original we need to maintain information about that form and also about the methods that are needed to translate between the storage representation and the presentation for use.

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

Depending on the software used, simply changing the size of the window in which a document is viewed can change properties such as character size or line length. Similarly, changing the magnification or "zoom" ratio, or switching between draft and print image modes in word processors alters the appearance of a document on a screen.

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

The Need to Reproduce Electronic Records

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

Reproducing the record involves both its intrinsic and extrinsic elements of form. With respect to the bitstreams that are maintained in storage over time, an electronic record contains one or more digital components. The first step in reproducing an electronic record is to reconstruct it by assembling all of its digital components in the proper arrangement. The second step is to present or render each of the components individually and all of the components collectively in the proper documentary form. The final

[6] E.g., magnetic and optical storage media are drastically different from CRT or LCD displays.

process in reproducing the record is to re-establish its immediate context. The immediate context of a record is its archival bond: the relationship between a record and other records. This is a two-step process. The first step consists of re-establishing the structure of the set of records in which the record belongs. The second step is to populate that structure with the relevant records.

The "Chain of Preservation"

An electronic record "in storage" is simply not the same as it was either before being stored or after retrieval. To justify belief that an electronic record retrieved from storage is the same in all essential respects as the record previously placed in storage, the rationale that is applied to hard-copy records—namely, that a physical object has been under continuous control that prevented it from being altered—is not sufficient. Given that the storage and retrieval processes for electronic records inevitably entail physical and representational transformations, the traditional concept of an unbroken chain of custody needs to be expanded to encompass the processes that are necessary to ensure that an electronic record is transmitted over time without inappropriate alteration. This expanded concept can be called the *unbroken chain of preservation:* the entire process of committing an electronic record to storage, maintaining it in storage, retrieving it, and presenting it must adequately preserve all its essential attributes in order to support a credible claim that the retrieved electronic record is authentic. In addition to what is entailed in the chain of custody, the chain of preservation will include information about the records creator's practices to support a presumption of authenticity, in accordance with the benchmark requirements for authenticity, information about the processes of bringing the records into the archives and maintaining them over time, and information about the reproduction of records, in accordance with the Baseline Requirements Supporting the Production of Authentic copies of electronic Records. In the digital environment, where records are not affixed in stable fashion to durable media in the forms in which they are presented for use, it is necessary to invoke an additional principle: the unbroken chain of preservation. It is not sufficient to hold on to the records. We must also ensure that any action that affects the way the records are presented protects their integrity.

Preservation, Conservation, and Maintenance

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

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

The Process of Preserving Electronic Records

The "Preserve Electronic Records" Model

The InterPARES Preservation Task Force has focused its efforts on the articulation of a formal model of the process of preserving electronic records. The process model was articulated using the Integration Definition (IDEF) methodology adopted by the International Team. Specifically, the model was articulated in accordance with the IDEF(0) standard for function modelling.[7]

In IDEF(0), "A function model is a structured representation of the functions, activities or processes within the modeled system or subject area." The "Preserve Electronic Records" model is intentionally generic. It identifies and describes the processes necessary to preserve electronic records, articulates the inputs needed by each process, the controls under which it operates, the mechanisms necessary to accomplish the process, and the output(s) produced by each process. The model defines the relationships among these entities and processes. It should be interpreted as describing a subject area, rather than a specific system. That is, while the model is systematic, it does not prescribe an implementation. Rather than defining a preservation system, the "Preserve Electronic Records" model provides a comprehensive, precise, and coherent road map which institutions and persons concerned with the preservation of electronic records can use in designing, developing, and evaluating systems that address their specific requirements, objectives, and constraints.

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

---

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

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

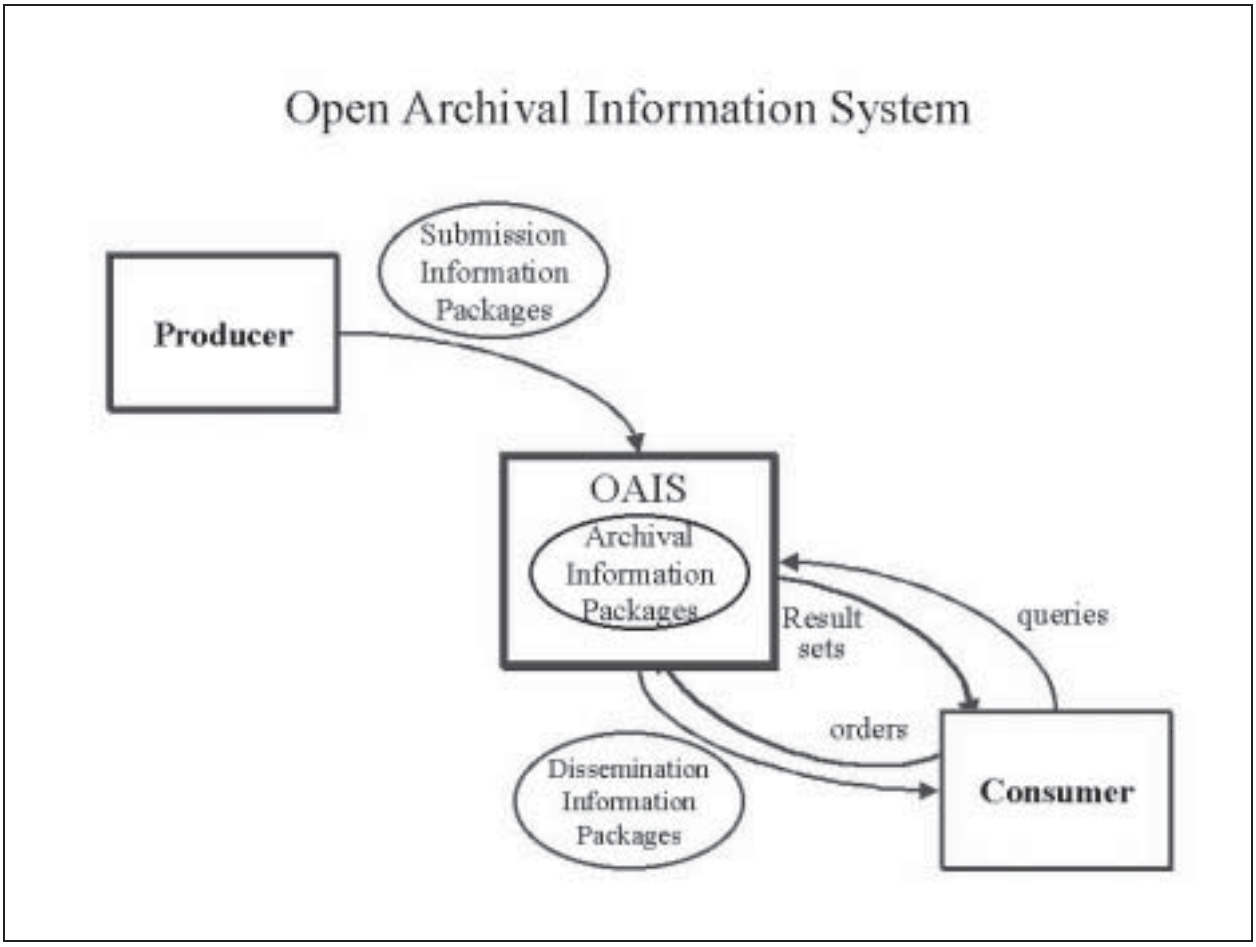

Figure 1. Open Archival Information System

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

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

The boundaries of the preservation function model derive from the viewpoint according to which the model is constructed. IDEF(0) models “functions (actions, processes, operations), functional relationships, and the data and objects.” The relationships between functions are logical, and not necessarily chronological. IDEF(0) does not explicitly model temporal sequences. Moreover, in IDEF(0), the viewpoint determines what can be ‘seen’ within the model context, and from what perspective or ‘slant’. Depending on the audience, different statements of viewpoint may be adopted that emphasize different aspects of the subject. Things that are important in one viewpoint may not even appear in a model presented from

another viewpoint of the same subject.[9]

The horizon for the viewpoint of the preservation model is determined by the scope of the InterPARES Project as whole. The project is concerned with the preservation of electronic records that have been selected for preservation when they are no longer needed for the practical purposes for which they were originally created. The scope of the InterPARES Project corresponds to that of the OAIS box in the centre of Figure 1. Therefore, the process described in the "Preserve Electronic Records" model begins with the transfer of the records from their creator, or from an agent acting for the creator, to a person whose primary responsibility is that of preserving authentic records; that is, the preserver. However, the preserver, as defined by the InterPARES Project, has responsibilities that are broader than the preservation process itself. For example, the preserver is presumed to be responsible for selecting the records that are to be preserved. In the "Preserve Electronic Records" model, the viewpoint is literally and strictly that of "the person responsible for preservation." The model's viewpoint includes only those entities and processes connected with someone, or some organization, carrying out the role of preserving the records. The same person or organization may have other roles or other, coincidental responsibilities, such as appraisal or reference, but coincidental responsibilities are excluded from the "Preserve Electronic Records" model. The role of preserving records includes all and only those activities necessary to ensure the transmission of authentic electronic records over time, according to the concept of preservation as described earlier in this report.

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

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

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

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

---

[9] Op. cit. in fn 7.

Preservation Overview

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

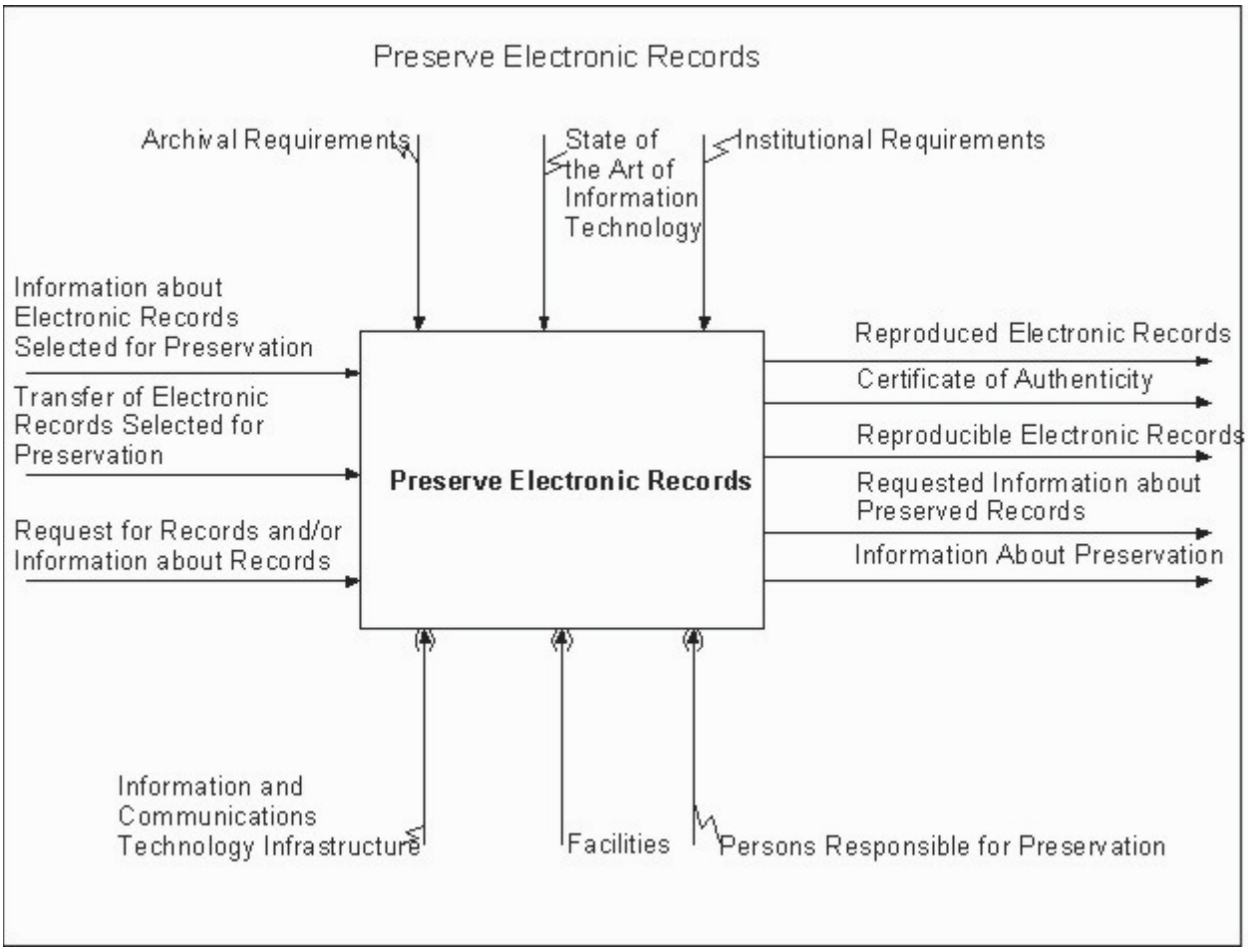

Figure 2. Preserve Electronic Records Context

The context diagram shows that three factors control the process of preserving electronic records. In order to preserve records, and especially to preserve them as authentic, we need to know what the requirements are for doing so. These requirements derive from archival science and principles and related standards and best practices for managing records. These requirements are labelled as "Archival Requirements" in the diagram. Preserving electronic records entails using digital information technology. The possibilities for doing so are limited by the state of the art of information technology, which constitutes the second type of control on the preservation process. The state of the art of technology includes products and services available in the marketplace, and also the feasibility of developing systems or applications for preserving records from these products and services, and also standards. Finally, the exercise of the preservation function will also be governed by requirements of the institution in which this function is carried out.

The diagram shows three mechanisms that are necessary to perform the preservation process. They are

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

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

The Main Preservation Processes

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

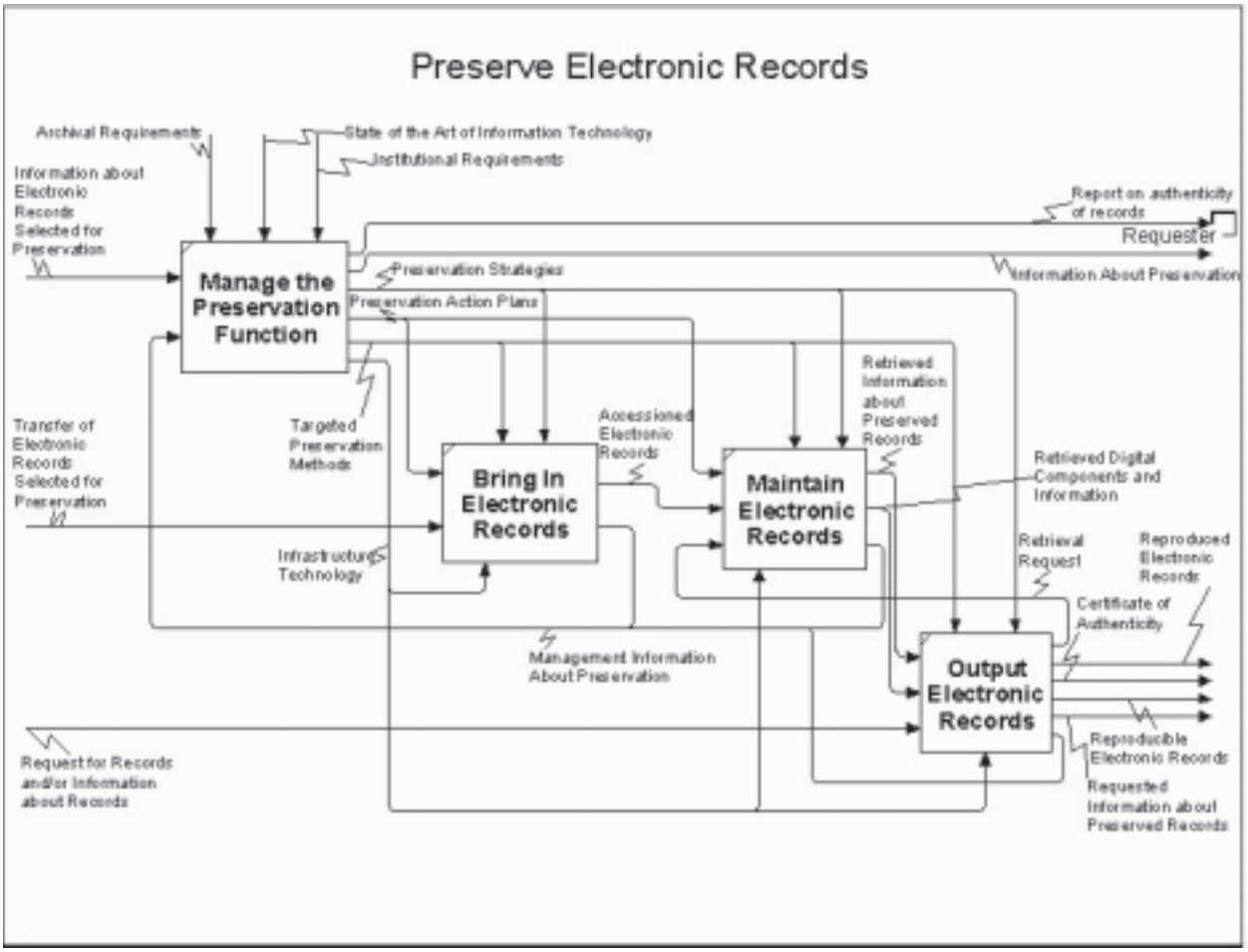

Figure 3. The Main Preservation Processes

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

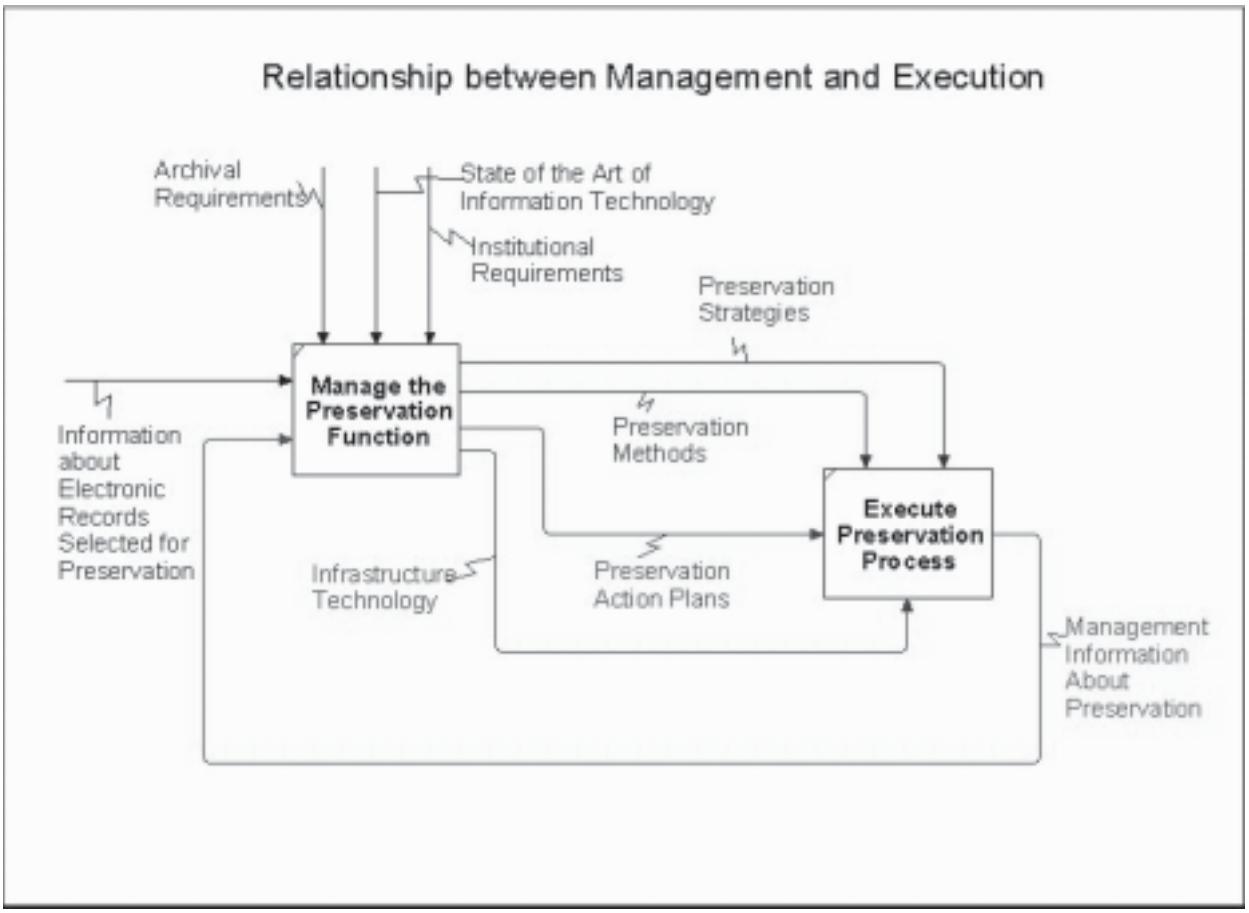

Figure 4. Management and Execution of Preservation Processes

Managing the Preservation Function:

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

subsystems, and communications protocols. The management function selects and acquires the information technology infrastructure that is used in carrying out the three execution processes.

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

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

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

Bringing Records into the Preservation System:

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

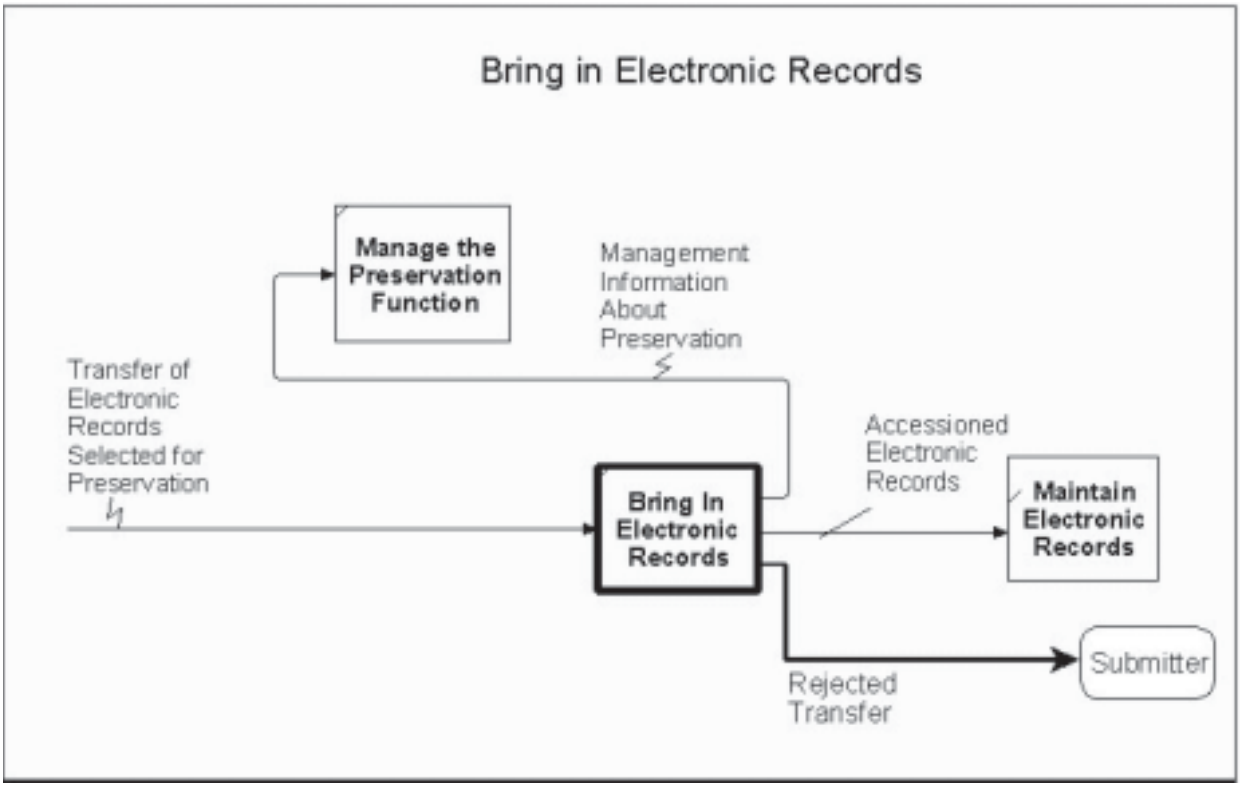

Figure 5. Bringing Records into the Preservation Process

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

However, the model does not include all facets of responding to requests for records or information. It is assumed that certain preliminary steps, and possible subsequent ones, related to the requests are taken by persons responsible for access to the records. For example, the persons responsible for the access function will help

requesters to identify the records or information in which they are interested, and will determine if they have a right to receive the records or information they request. As with the appraisal function, there will be some overlap between the preservation and access functions. If a model of the reference or access function were articulated, it would include activities that coincide with or overlap some of the activities within the Bring In process because the examination of the records and related information during that process would provide the opportunity to develop the information needed for producing finding aids. Similarly, the reference function would probably determine the computer interface used to provide requesters with access to electronic records, but preservation would determine how the records need to be presented in that interface in order to guarantee that the reproduced records are authentic.

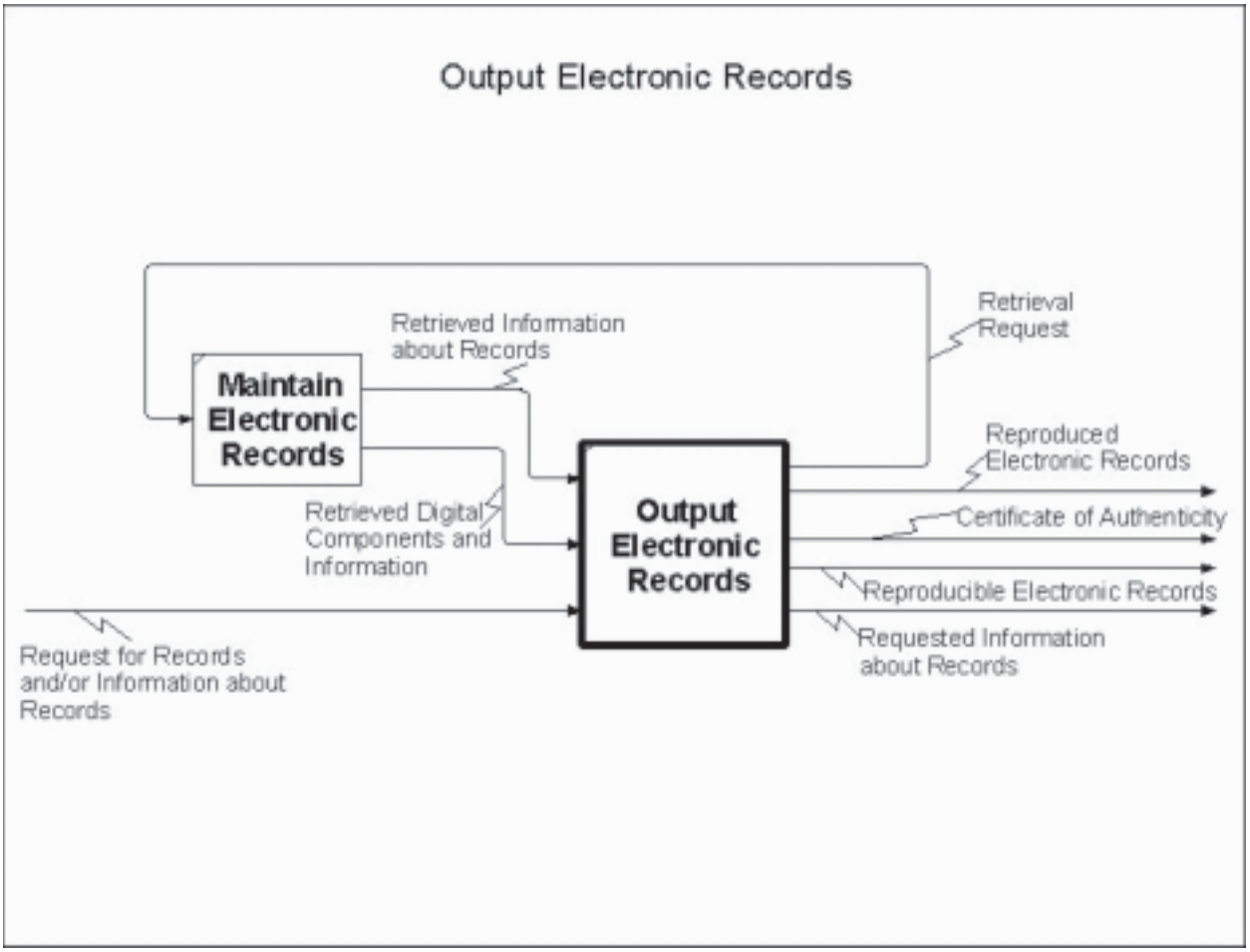

Figure 6. Reproducing Electronic Records

Figure 6 depicts the Output Electronic Records process of the preservation model. The requests that are input to this process can be for records or for information about records, or both. Although the reference function is responsible for helping requesters to identify the records of interest to them, descriptions and

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

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

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

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

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

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

the production of authentic copies. Accordingly, the Output process should also provide the requesters with criteria they can use to determine if the copies are authentic.

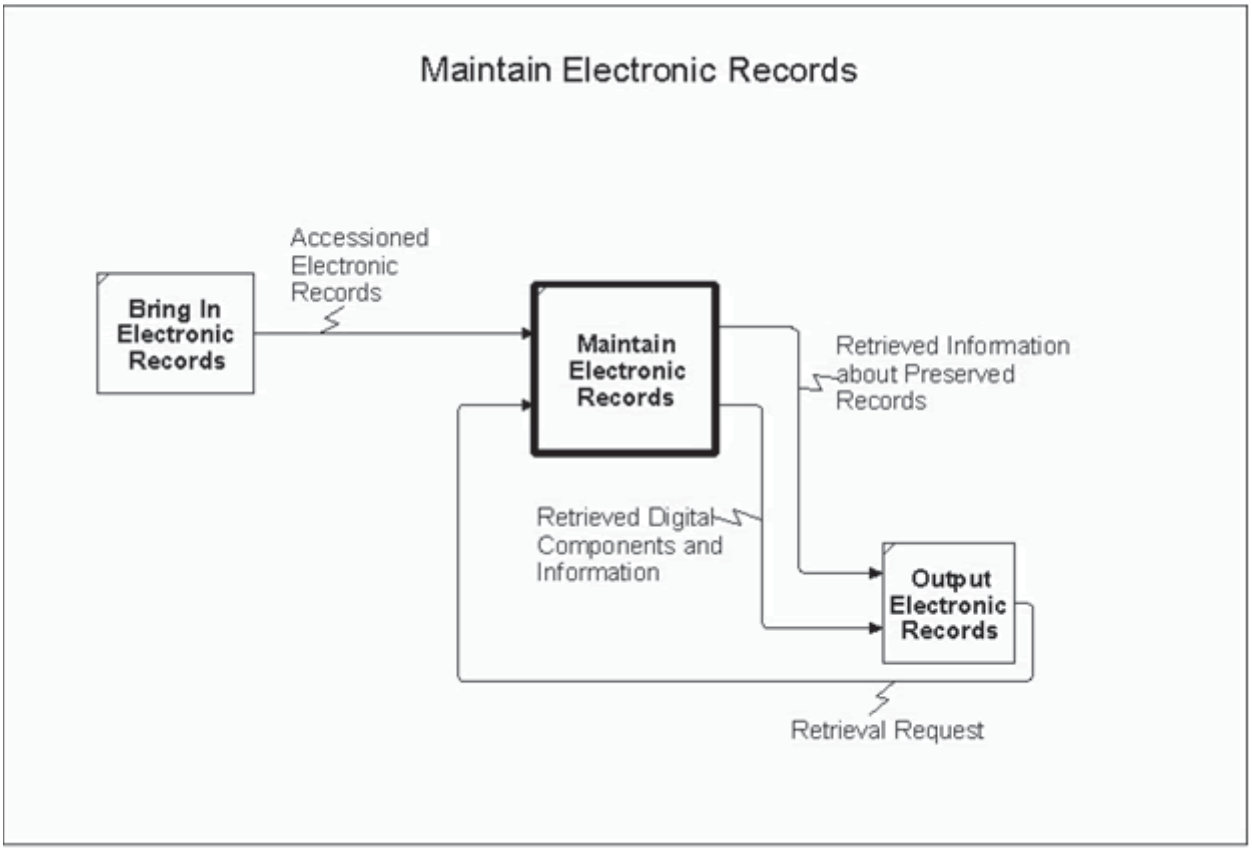

Figure 7. Maintaining Electronic Records

Maintaining Electronic Records over Time:

One of the four basic processes remains to be described, Maintain Electronic Records. This is an internal process: none of the inputs that come into the preservation process from the outside go directly to this sub-process, and the outputs of the maintenance process go to other sub-processes within the function. Nonetheless, the Maintain process is the core preservation process with respect to transmitting electronic records over time. It is depicted in Figure 7. The Maintain process is connected to both the Bring In and Output processes. The digital components of the records—along with information needed to reproduce them, re-establish their archival bonds, and certify their authenticity are received from the Bring In process. Requests for records and information are received from the Output process and the results of executing such requests are returned to Output.

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

To this point we have examined the four preservation processes depicted in Figure 3 as they relate to one

another. In the following sections, we consider the activities that take place within each of the four preservation processes.

Preservation Sub-processes

Manage the Preservation Function

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

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

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

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

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

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

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

and objectives that evolve over time.

Bringing Electronic Records under Preservation Control

In accordance with the preservation strategies established in the preservation management process, electronic records are brought under preservation control. Bringing electronic records under preservation control includes four activities: registering the record transfer, verifying the authority for transfer, examining the records, and accessioning the records.

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

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

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

The examination should also include a final review[10] of information that provides a basis for presuming that the records were maintained as authentic. That information then becomes part of the preservation history of the records. It will be retained, and augmented in the course of maintaining the records over time in order to document the "chain of preservation" and enable copies of records to be certified as authentic.

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

Maintaining Electronic Records over Time

Electronic records are stored as *digital components*, which may be separate *digital files,* or contained in a single digital file. The preservation function aims to deliver records that convey the intellectual content and intent of the creator. To do this, it is necessary to reassemble the components to reconstitute the records and present them in their original documentary form. This has to be done every time the record is accessed simply because an electronic record cannot be stored in the same form in which it is presented to humans. Strictly speaking, maintaining electronic records over time means maintaining the ability to output them. Maintaining electronic records, thus, requires storing their digital components and

[10] This information was reviewed previously as part of the appraisal process.

maintaining information about the records, such as what digital components they contain, how those components are related to each other and the records, and how the records should be presented. In order to be able to output authentic electronic records, it is also necessary to maintain information that justifies an assertion of authenticity.

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

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

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

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

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

Evaluating the success of such technical interventions requires reconstituting and rendering the record(s) and/or archival set(s) based on the reproduction strategies and authenticity requirements. In both cases,

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

Reproducing Electronic Records

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

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

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

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

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

Conclusion and Recommendations

Much attention to the preservation of electronic records has focused on the twin problems of the relatively short life expectancy of digital media and the rapid obsolescence of hardware and software. The InterPARES Project started with recognition of these problems and cast the preservation issue in terms of evaluating practical methods for solving them. The research plan called for the Preservation Task Force “to identify and develop the procedures and resources required for the implementation of the conceptual requirements [for preserving authentic electronic records] and criteria [for appraising electronic records] identified in the first two domains."[11] This formulation of the problem of preserving electronic records clearly situates it not in technology, but in the interface between the goal of preserving electronic records and the technology on which they depend. Technology itself is not a problem. If we did not need to preserve records beyond the life expectancies of hardware, software, and digital media, we would not have any preservation problem. Similarly, technology cannot determine the solution. It is archival and records management requirements that define the problem. It must be archival and records management criteria that determine the appropriateness and adequacy of any technical "solution." The question “What

[11] <http://www.interpares.org/researchplan.htm>

is the best technological method for preserving electronic records?" is as meaningless as the question "What is the best medicine for making people healthy?" Neither can be answered without specifying the conditions it is meant to address. The InterPARES Project defined these conditions as the archival requirements for authenticity and the archival criteria for selecting records to be preserved.

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

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

Recommendation 1.

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

Recommendation 2.

Use of the "Preserve Electronic Records" model should be based on an understanding of the particular characteristics of electronic records and what those characteristics entail for preserving these records, as summarized in the foundation concepts that were set out earlier:

- Digital Components of Electronic Records
- Preservation Control
- Archival Requirements for Preservation
- "Original" Electronic Records
- The Need to Reproduce Electronic Records
- The Chain of Preservation.

The key to all of these concepts is the recognition that the chain of preservation for electronic records must extend over their entire life and that the process of preserving electronic records extends to and includes reproducing the records.

Recommendation 3.

Solutions to the preservation of specific bodies of electronic records should be inherently dynamic. The reason is twofold. First, most archives and other preservers will accumulate electronic records over time. Over time, the specific properties of the records brought into the archives will change. The preservation

system must be capable of being expanded, adapted, and modified to accommodate new and different types of electronic records, and new ways of organizing, accessing, and presenting such records. Second, the goal of preserving electronic records is not to keep them, in archives or elsewhere, but to make them available to persons who have a need for, or an interest in, them. While the preserver has a fundamental responsibility for providing access to authentic records, their availability will be impacted by the continuing evolution of information technology. Preservers should assume that future users will want to use the best available technology for access to the records. Preservation solutions should be designed to be able to interface with evolving technologies for information discovery, retrieval, communication, and presentation.

Recommendation 4.

The InterPARES Project has been so fruitful that it has not only provided valuable products in response to the research questions that it originally posed, but it has also raised the threshold of research by articulating issues that are entailed in the original questions, although not explicit in them, by identifying new questions, and by opening up lines of research that should provide grounds for valuable results for years to come. For example, the project has moved beyond its foundation in the science of diplomatics to recognize that, in the digital environment, many of the concepts and methods that traditionally were applied to individual documents need to be applied to sets of records. This insight needs to be explored more fully. The work of the Preservation Task Force has focused on defining a comprehensive framework for preserving authentic electronic records. More work is needed to analyze the data and information requirements for executing the processes defined in the preservation model. The model should also be applied to test cases both to validate and enrich it. The model should also be extended to address the application of specific technologies for overcoming technological obsolescence. The accomplishments of the InterPARES Project should be applied to related areas of concern, such as the process of archival description. In sum, this work should not stop when the current project ends. The archival profession, our collaborators, and our stakeholders, have an interest and responsibility to see that further progress is made.