

Preserving Authentic Electronic Records: The InterPARES Project

Dr. Luciana Duranti

Chair and Professor, Master of Archival Studies

University of British Columbia

The InterPARES project was launched in January 1999 and its first phase has reached its conclusion in December 2001. Its goal was to develop the theoretical and methodological knowledge essential to the permanent preservation of authentic records generated and/or maintained electronically, and, on the basis of this knowledge, to formulate model policies, strategies and standards capable of ensuring that preservation. To meet this goal, the research was organized into four domains of inquiry. The objective of Domain I was to formulate the conceptual requirements for preserving authentic electronic records, and to identify the elements of electronic records that are necessary to maintain their authenticity over time. Domain II sought to determine whether or not the evaluation of electronic records for permanent preservation should be based on theoretical criteria different from those applied to traditional records. Domain III aimed at developing procedures for the long-term preservation of authentic electronic records, all the while considering the conceptual requirements for authenticity articulated in Domain I and the conclusions reached in Domain II. Domain IV was devoted to the articulation of an intellectual framework for developing policies, strategies and standards for the long-term preservation of authentic electronic records.

The research was based on concepts agreed upon at the outset. It was established that an electronic record was a record made or received and set aside for reference or action in electronic form, and that its salient characteristics were:

- a fixed form (i.e. its binary content is stored so that it remains complete and unaltered, and its message can be rendered with the same documentary form it had when first set aside);
- an unchangeable content;
- explicit linkages to other records within or outside the digital system through a classification code or other unique identifier
- an identifiable administrative content;
- three persons concurring in its formation, that is, an author, an addressee, and a writer; and
- its participation in or support of an action either procedurally or as part of the decision making process.

It was further agreed that a trustworthy record is a record that is reliable and authentic, where reliability is the ability of a record to stand for the facts it is about, that is, its trustworthiness as a statement of fact, while authenticity refers to the fact that a record is what it purports to be and has not been tampered with or otherwise corrupted, that is, to its trustworthiness as a record. It was emphasized that there is a fundamental difference between authenticity and authentication, the latter being a declaration of authenticity, a means of proving that a record is what it purports to be at a given moment in time.

In archival theory and jurisprudence, records that are relied upon by their creator in the usual and ordinary course of business are presumed authentic. 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 a record, the preserver must be able to establish its identity and demonstrate its integrity. The identity of a record refers to the attributes of a record that uniquely characterize it and distinguish it from other records. These attributes include: the names of the persons concurring in its formation (i.e., author, addressee, writer and originator); its date(s) of creation and transmission; an indication of the matter or action in which it participates; the expression of its archival bond; as well as an indication of any attachment(s). These attributes may be explicitly expressed in an element

of the record, in metadata related to the record, or implicit in its various contexts (documentary, procedural, technological, provenancial, or juridical-administrative). The integrity of a record is its wholeness and soundness. A record has integrity if it is intact and uncorrupted. A record is intact and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered. A record's physical integrity, such as the proper number of bit strings, may be compromised, provided that the articulation of the content and its required elements of form remain the same. Integrity 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 contexts.

The presumption of authenticity is an inference that is drawn from known facts about the manner in which a record has been created and maintained. The Authenticity Task Force, which was responsible for the first domain of investigation, issued Benchmark Requirements, which detail the evidence required for a presumption of authenticity. A presumption of authenticity for the records of a given creator will be based upon the number of requirements that have been met by the creator and the degree to which each has been met. When there is an insufficient basis for a presumption of authenticity, a verification of authenticity is necessary. This verification is the act or process of establishing a correspondence between known facts about the record and the various contexts in which it has been created and maintained, and the proposed fact of the record's authenticity. It involves a detailed examination of the record in all its contexts and of reliable information available from other sources (audit trails, backups, copies preserved elsewhere, textual analysis).

The Benchmark Requirements supporting a presumption of authenticity are the following:

Benchmark Requirement A1:

The value of the following attributes must be 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, writer, originator, and addressee

A.1.a.ii

Name of action or matter

A.1.a.iii

Date(s) of creation and transmission, that is: chronological date, received
date, archival date, transmission date(s)

A.1.a.iv

Expression of archival bond

A.1.a.v

Indication of attachments

A.1.b

Integrity of the record:

A.1.b.i

Name of handling office

A.1.b.ii

Name of office of primary responsibility

A.1.b.iii

Indication of types of annotations added to the record

A.1.b.iv

Indication of technical modifications

Benchmark Requirement A2:

The creator has defined and effectively implemented access privileges
concerning the creation, modification, annotation, relocation, and
destruction of records

Benchmark Requirement A3:

The creator has established and implemented procedures to prevent, discover, and correct loss or corruption of records

Benchmark Requirement A4:

The creator has established and implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change

Benchmark Requirement A5:

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

Benchmark Requirement A6:

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

Benchmark Requirement A7:

If multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative

Benchmark Requirement A8:

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 implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records.

After the records have been presumed or verified authentic in the course of the appraisal process, and have been transferred from the creator to the preserver, their authenticity needs to be maintained by the preserver by reproducing them and authenticating the resulting copies. The production of authentic copies by the preserver is regulated by a second set of requirements developed by the Authenticity Task Force that must all be met and are therefore called “baseline requirements”. They are the following:

Baseline Requirement B1:

The procedures and system(s) used to transfer records to the archival institution or program, maintain them, and reproduce them must embody adequate and effective controls to guarantee the records' identity and integrity, and specifically to ensure that:

- unbroken custody of the records is maintained;
- security and control procedures are implemented and monitored; and
- the content of the record remains unchanged after reproduction.

Baseline Requirement B2

The activity of reproduction has to be documented, and this documentation must include:

- the date of the records' reproduction and the name of the responsible person;
- the relationship between the records acquired from the creator and the copies produced by the preserver;
- the impact of the reproduction process on their form, content, accessibility and use; and
- in those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, information about this fact, which will be readily accessible to the user.

Baseline Requirement B3:

The archival description of the fonds containing the electronic records must include—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.¹

The Appraisal Task Force, which was responsible for the second domain of investigation, began its work by analyzing the existing literature on appraisal of

¹ The Benchmark and Baseline Requirements are a slightly edited version of the requirements included as Appendix 2 in the *Long-Term Preservation of Authentic Electronic Records. Findings of the InterPARES Project*, published on and available at www.interpares.org.

electronic records. Such analysis found that there is general agreement on the fact that electronic records must be selected according to the same theory and criteria used for traditional documents, on the importance of evaluating the entire context of the records, on the necessity of conducting selection very early in the life of the records, and on the importance of having all the documentation related to the technological context of the documents, but it also found that authenticity is noticeably absent among the selection criteria. It therefore developed a model representing the selection function from the point of view of the preserver and including the following activities:

1. Manage the selection function
2. Appraise electronic records
3. Monitor selected electronic records
4. Carrying out the disposition of electronic records.

The key to this model is, however, that it revolves around the assessment of the authenticity of the records. Such assessment has never been part of the responsibility of the appraiser because it is an activity that risks compromising the impartiality of the records by alerting the creators to their inherent value, interpreting the records formal elements and evaluating their processes of creation and maintenance. In addition, archival practice has traditionally rejected the assessment of the authenticity of the records as part of appraisal, on the grounds that it would make appraisal far too laborious and time consuming. However, this common stance of archival theorists and practitioners could be held only because, with traditional records, the documents entering an archival institution or program were the same made or received and set aside by their creator or legitimate successor and evaluated by the archivist in the scheduling process. Thus, the assessment of authenticity could be easily delegated to future researchers, who would be able to analyze the documents under scrutiny in their original instantiation, that is, in the same form and status of transmission they had when first made or received and set aside. This is no longer the case.

Electronic records undergo several changes from the moment they are generated to the moment they become inactive and are ready for disposal. Some

of those changes are intentional. Information technology is in a constant state of development. Records creators continually update their systems and the live documents contained in them, at times with minimal consequences for the form, functionality, organization and metadata of the records, other times with dramatic consequences. The latter situation is more likely to occur when records generated in an obsolete system are migrated to a new one. In addition to intentional changes, inadvertent changes occur, simply because of the fact that it is impossible to maintain an electronic document; it is only possible to store its digital components in a way that the document can be reproduced when needed. A digital component is a digital object that contains all or part of the content of an electronic record, and/or data or metadata necessary to order, structure, or manifest the content, and that requires specific methods for storage, maintenance and preservation. In any case, every time an electronic record is reproduced from its digital components, it is slightly different from the previous time. This happens because there are three steps in the reproduction of an electronic record. The first step is to reassemble all the record's digital components in the correct order. The second step is to render the components, individually and collectively, in the correct documentary form. The third step is to reestablish the relationships between the record in question and all the other records that belong in the same archival aggregation (e.g., series, file). This requires, first, to reestablish the structure of the archival aggregation, and then, to fill it with the records that belong to it. Each step involves a margin of error. Considering that the processes of storage and retrieval by re-production imply transformations that are both physical and of presentation, the function of preservation must begin when the records are created respecting established authenticity requirements, and continue with the documentation of all the changes to the records and of the processes of appraisal, transfer, reproduction and preservation. However, the most important consequence of this situation is that the appraisal function must include appropriate activities aiming at ascertaining the authenticity of the records considered for selection, monitoring it, and attesting it.

The appraisal of electronic records must therefore comprise the following activities: compiling information about the records and their contexts, assessing the value of the records, determining the feasibility of preserving them, and making the final appraisal decision. The information that needs to be gathered about the records includes information on the context of creation and on the technological context, which establishes the basis upon which the records are considered authentic. The assessment of the value of the records involves assessing both the continuing value of electronic records and their authenticity. Determining the feasibility of preserving electronic records involves deciding whether the digital components embodying the essential elements that confer identity to and ensure the integrity of the records can be preserved, given the current and anticipated technological capabilities of the archives. This determination process comprises three steps. The appraiser identifies both the record formal elements containing informational content and those formal elements that need to be preserved according to the authenticity requirements enumerated earlier. Then, the appraiser identifies where these crucial formal elements of the record are manifested in its digital components. Finally, the appraiser reconciles these preservation requirements with the preservation capabilities of the institution that is responsible for the continuing preservation of the body of records being appraised. The appraisal decision comprises two parts: a determination of what must be transferred to the archives, including the list of the digital components of each record, and a determination of how and when this should happen, including the identification of acceptable formats and methods of transmission to the archives.

Once appraisal is concluded, the records selected for preservation must be continually monitored till the day of the transfer, especially for identifying changes in their technological context. In some cases, it may be necessary to repeat the appraisal because of changes that can affect the feasibility of preservation. In most cases, however, monitoring produces minor revisions to the documentation on the selection and to the terms and conditions of transfer.

It is important that there be documentation explaining and justifying the appraisal decision. It should be clear why some records were preserved and others were not, both for accountability purposes and so that future users of the records can understand them. In fact, this documentation constitutes a permanent record of the archives that must be accessible to researchers wanting information about appraisal and about records selected for preservation. Information about appraisal decisions is also a crucial mechanism for implementing the monitoring activity described earlier. In addition, it is important that the records selected for preservation be packaged at the moment of transfer with the necessary information for their continuing preservation, including the terms and conditions of transfer, identification of the digital components to be preserved, and associated archival and technical documentation needed for their treatment. This is the information that is compiled and recorded during the various stages of appraisal and monitoring.

The third domain of investigation was entrusted to the Preservation Task Force, which based its work on the following ascertained facts:

- it is not possible to preserve an electronic record, it is only possible to preserve the ability to reproduce it;
- the relation between a record and a file can be one-to-one, one-to-many, many-to-one, or many to many;
- the same presentation of a record can be created by a variety of digital presentations and, vice versa, from one digital presentation a variety of record presentations can derive; and
- it is possible to change the way in which a record is contained in a file without changing the record.

In addition, the Task Force articulated several fundamental concepts, some of which have already been mentioned but merit repeating, as follows:

- to make possible the reproduction of electronic records it is necessary to preserve their digital components;
- the risks of corruption and loss of records are more frequent and complex when records cross technological boundaries. Thus, the

controls are divided in two types: those inside the system, which ensure that the records remain unaltered within it, and the dynamic ones, which ensure that the records remain unaltered when they cross technological boundaries;

- the controls may be technological in nature but must be determined on the basis of archival principles and criteria;
- it is impossible to maintain literally unaltered an electronic record;
- the only way to prove that an electronic record is authentic is to produce an authentic copy of it;
- considering that the processes of storage and retrieval imply transformations both physical and of presentation, the traditional concept of unbroken chain of custody must be extended to include the processes necessary to ensure the unaltered transmission of the record through time; and
- the unbroken chain of preservation begins when the records are created respecting the benchmark requirements, and continues with the documentation of all the changes to the records and of the processes of selection, transfer, reproduction and preservation.

Consistently with these facts and concepts, the Task Force developed a model representing the preservation function from the point of view of the preserver, and including the following activities:

- Manage the preservation function
- Bring in electronic records
- Maintain electronic records
- Output electronic records

Accordingly, on the basis of the information accumulated during the selection function, the preserver develops a strategy of preservation, action plans that are each connected to a specific body of records selected for preservation, a technologic infrastructure and the methods necessary to implement the action plans. This activity produces also reports on what has been preserved. Then, the preserver ascertains that the records which were transferred to the archives are

those selected for preservation, that it is possible to preserve them, and that there are no other obstacles to preservation. In the process, the preserver produces information on each transfer for the purpose of confirming or revising the strategy of preservation and the applicable action plans, and produces information useful to evaluate the execution of the acquisition activity and to satisfy the baseline requirements for authenticity.

Once the records have been formally acquired, the preserver applies specific methods of preservation for the body of records in question on the basis of the action plan for those records, maintaining the digital components together with the information necessary to re-produce the records when requested, to certify their authenticity, and to make possible their comprehension. In the process, the preserver produces information useful to evaluate the execution of this function and satisfy the baseline requirements for authenticity. The following activity consists of applying the re-production method established for the body of records in question and implementing the action plan to produce an authentic copy of the record in response to a request of access. If specified in the request, the preserver generates a certificate that attests the record's authenticity. Alternatively, if requested, the preserver can give the user a reproducible electronic records, that is, the digital components of the record with the instructions for rendering it as an authentic copy and with the information necessary to comprehend it.

The key points resulting from the work of the Preservation Task Force can be considered an outcome of the work of all InterPARES Task Forces. First, technology cannot determine the solution to the permanent preservation of electronic records; second, archival needs must define the problems and archival principles must establish the correctness and adequacy of each technical solution; and, third, solutions to the preservation problem are inherently dynamic, thus ongoing research is vital to deal with the challenges presented by the new information technologies.

As a consequence of this realization, the InterPARES Project has moved into a second phase, which, on the basis of the findings of the first, addresses

issues of reliability and accuracy, in addition to issues of authenticity, throughout the records' life-cycle; and focuses on records produced in new digital environments, experiential, dynamic, and interactive, and resulting from artistic, scientific and e-government activities. The results of this second phase are expected for 2007, so...stay tuned.²

² The findings of the first phase of the InterPARES Project are electronically published as the *Long-Term Preservation of Authentic Electronic Records. Findings of the InterPARES Project*, and available on the project's web site www.interpoares.org. Basic information on the second phase of InterPARES can be found at the same URL.