Session: Preserving Electronic Records: Research Findings and Practical Approaches

Bridging the divide: from theory to practice

Luciana Duranti
Chair and Professor, Archival Studies
The University of British Columbia
Vancouver, B.C., Canada

Ongoing technological change is causing widespread concern regarding the preservation of the records produced or stored using digital technologies. The obstacles are presented by the fact that such records exist in an environment which is hybrid, because paper and film are still an integral part of the documentary system, favours the manipulation of data, comprises applications that are proprietary and idiosyncratic in nature, tends to support decentralization of records creation and the repurposing of records, and is subject on the one hand to requirements of regulatory agencies based on existing technology and on the other hand to the increasingly frequent obsolescence of systems and media. Consequently, records are less reliable, retrievable, accessible, readable or intelligible than they used to be, and it is very difficult to preserve them over the long term.

Moreover, even if we could overcome media fragility and technological obsolescence and maintain accessibility over time, records are of little value unless we can be sure they are authentic, that is, that they can be trusted as sources. For centuries, our presumption of the authenticity of records has been premised on the presence or absence of visible formal elements such as seals and signatures, of controls on the procedures by which records are generated, transmitted, used and maintained, and on an uninterrupted line of legitimate custody. The use of digital technology to create records has reconfigured the traditional formal elements by which records were recognized as authentic, allowed for the bypassing of procedural controls, and made of physical custody an elusive concept. There have been several attempts to develop solutions, by issuing standards and guidelines or recommendations resulting from research projects. They have produced a good foundation for the development of trusted record-keeping systems, capable of ensuring the reliability and authenticity of the records they contain, and have defined the fundamental concepts and methods that must be respected to control the trustworthiness of records throughout their life-cycle, but none of them has focused on preservation.[1]

---

[1] Standards for record-making/keeping systems: in the USA, Design criteria standard for electronic records management software applications (DOD 5015.2-STD) and, in Europe, Model REQuirements for Electronic Records Management Systems; Guidelines for records preservation: International Council on Archives' Guide for Managing Electronic Records from an Archival Perspective. Among the research project, the best known are the Pittsburgh Project, see "Functional Requirements for Evidence in Recordkeeping." <http://www.web.archive.org/web/19981203042506/www.sis.pitt.edu/~nhprc/> (31March 2003); the Philadelphia Project, in Weinberg, David M., Mark D. Giguere, David S. Miller, and Celia O'Leary. "The Philadelphia Electronic Records Project: Some Clarifications." *Archivaria* 45 (Spring 1998): 1-3; the UBC Project, in Luciana Duranti, Terry Eastwood and Heather MacNeil, The Preservation of the Integrity of Electronic Records (Dordrecht: Kluwer Academic Publishing, 2002).

If electronic records will ever be as trustworthy in the long term as records on traditional media, the practices by which they are created, maintained, and used must be geared to that purpose, and strategies and standards for long-term authentic preservation must be developed. This is the mission of a project known as InterPARES (International research on Permanent Authentic Records in Electronic Systems), whose specific goal is 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.

The InterPARES project decided to start its research from the theory of the records rather than from the observation of best practices, and to bridge the divide between theory and practice by developing a framework made of principles, criteria, methods, procedures, strategies applicable in different cultural, social, juridical and organizational contexts. For this reason, it used the theory and methods of diplomatics and archival science to define concepts and to develop requirements and methods; grounded theory and statistical analysis to carry out and examine case studies; comparative analysis for the study of appraisal and preservation reports from archival institutions which had had experience of these activities; modeling for the representation and definition of the activities involved in appraisal and preservation; computer engineering for the study of storage media and of digital preservation technology and technological methods of authentication; and legal analysis for the study of certification methods.

This paper will discuss the key results of InterPARES to date and will reflect on its success in bridging the divide between theory and practice and providing a theoretical foundation and an intellectual framework useful to the development of sound and consistent practices.

The InterPARES research team determined at the outset the concepts upon which all co-investigators from all countries and disciplines involved were ready to agree. 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). 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.

It is essential to be able to presume the authenticity of the records produced and maintained in live systems. Such presumption of authenticity is an inference that must be drawn from known facts about the manner in which a record has been created and maintained. For the purpose of enabling a preserver to presume the authenticity of the records to be kept over time, InterPARES 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 are eight. While the first identifies the fundamental information about an electronic record that establishes its identity and allows for the demonstration of its integrity, the other seven identify the types of procedural controls over the record's creation, handling and maintenance that support a presumption of integrity. All benchmark requirements are derived from the diplomatic body of knowledge.

The first requirement prescribes that the value of the following attributes[2] are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records:

[2] A record attribute is a defining characteristic of the record or of a record element. A record element is a constituent part of the record's documentary form. An attribute may manifest itself in one or more elements of a record's documentary form (e.g. the name of the author as superscription or as a signature) or in an annotation to the record (e.g. the archival bond as a record identifier) or in metadata in the audit trail, etc.

•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: 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

The attributes listed above may appear as elements of form or as annotations on the face of the record (e.g. the date, the name of the handling office), but they are more likely to be metadata linked to the record. It is essential that these attributes be inextricably linked to the record, and this means that their presence in separate parts of the system, such as the audit trail, is not helpful to guarantee their permanent accessibility in connection with the record and their ongoing existence, in addition to being unpractical, because the preserver would have to maintain a very large amount of unneeded information in order to keep the specific data related to a record. The two primary means of linking these attributes to a record are the record profile and the topic map. A record profile is a form inextricably linked to a record, which includes specific fields for the automatic or manual inclusion of data related to the record, and it is very common in electronic records management systems. A topic map visually expresses the characteristics of a record and the relationships among them. When a record is either removed from the system for external storage, migrated on the occasion of a system upgrade, or transferred to the preserver, the attributes must go with it, remain inextricably linked to it and be accessible to the user.

The second benchmark requirement regards access privileges. It prescribes that a presumption of authenticity be supported by the fact that the creator has defined and effectively implemented access privileges concerning the creation, modification, annotation, relocation, and destruction of records. The assignment of the authority and capacity to carry out administrative action on the records must therefore be accompanied

by the exclusive technical capability to exercise such responsibility. This is usually done by creating inside the system tables of users' profiles that provide differentiated kinds of access depending on the users' administrative competence. However, access control can also be exercised by means of external security systems, such as the exclusive assignment of a key to a location. The effective implementation of access privileges consists in monitoring access through the use of audit trails that record each interaction of a user with a record.

The third requirement prescribes that the creator has established and implemented procedures to prevent, discover, and correct loss or corruption of records. Examples of these procedures are making of regular back-ups of records and their attributes, as well as of the entire system; and ensuring that the backup and recovery procedures will guarantee that, in case of system failure, all complete updates are reflected in the rebuilt files and so is any incomplete operation.

The fourth requirement prescribes that the creator has established and implemented procedures to guarantee the continuing identity and integrity of records against media deterioration and across technological change. In order to counteract media fragility and technological obsolescence, the creator must plan upgrades to the technological infrastructure of its organization, making sure that the ability to retrieve, access and use records when the upgrades occur is maintained. In addition, the creator must plan procedures of refreshment of the records, moving them from a storage medium to another, and of migration of the records from obsolescent to new technologies.

The fifth requirement prescribes that the creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator. This requirement derives from the fact that the authors of electronic records feel much freer in their compilation than the authors of paper records, and tend to let technology rather than administrative procedure determine the form the record. An acceptable compromise is to let the documentary form of a record be determined by workflow control technology, where one can connect each step of a procedure to a documentary form. Also, the creator can customize specific applications for the whole organization, so that all e-mails or all spreadsheets of a certain kind, for example, will present the same form. The control on documentary form must go down to the level of extrinsic and intrinsic elements, because this is the level at which the authenticity of the record is maintained and can be verified.

The sixth requirement prescribes that, 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 what are the means of authentication. This requirement may be met by linking the authentication of specific types of records to the various steps of the administrative procedure, assign responsibility to a given officer or an office for authenticating either individual or all records, and determining either a method of authentication valid for the entire organization or specific authenticating instruments for specific types of records.

The seventh requirement prescribes that, if multiple copies of the same record exist, the creator has established procedures that identify which record is authoritative. One of the greatest problems presented by electronic records is the easiness of reproduction. Innumerable copies of each record may exist everywhere in the organization, each slightly different from the other, as it resides in a different hard drive

of a different computer or because of modifications voluntarily applied to the record by the one or the other person in the organization. It is vital for each organization to know what is its official record, especially because the status of transmission of each instance of the same record is inevitably that of copy, either of an original or of a draft. In fact, the original record, which in electronic systems is the first complete and effective record either received (if transmitted across space) or saved to a file in the system (if transmitted across time), ceases to exist after being stored for the first time. When recalled, the stored entity is a copy, which, in the best of all possible scenarios is a copy in the form of original, but in most cases is simply an imitative copy. Also a draft, while conceptually remaining the sketch or outline of the definitive document, prepared for purposes of correction and meant to be provisional, will only exist as an imitative or simple copy of the draft first stored. Thus, the official copy of each record will have to be subject to strict procedural controls that will serve as a form of authentication, considering that technologically based forms of authentication are useful only when records are transmitted across space, as they usually constitute an obstacle to the maintenance of the record to which they are linked. Of course, when the official record is identified, so is the office of primary responsibility for that record, that is, the office having the formal competence for maintaining the official records that share the same classification and retention period. This will help also reducing duplication in the organization and designating accountability for the records.

The eight and last requirement prescribes that, if there is a transition of records from active status to semi-active and inactive status that 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. This documentation includes all the information necessary to access the records, to establish their identity and to demonstrate their integrity. If the system generates records profiles, it will be sufficient to ensure that all records are accompanied by their profile. Otherwise, it may be necessary to transfer with the records audit trails, indexes, data directories and data dictionaries, etc.

The requirements listed above are thus intended to allow the preserver to assess, in the course of the process of appraisal, the authenticity of the electronic records of a creator before they are transferred to his/her custody. As it regards appraisal, our investigations 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 they also found that authenticity is noticeably absent among the selection criteria.

InterPARES proposes an appraisal procedure that revolves around the authenticity of records. 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 physically intact an electronic document. The most important consequence 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 appraiser must gather information on the context of creation and on the technological context, which establishes the basis upon which the records are considered authentic; must determine the feasibility of preserving the records on the basis of the current and anticipated technological capabilities of the archives; and must decide what should be transferred for long term preservation, and 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.

After the records have been presumed or verified authentic and transferred to the custody of the preserver, their authenticity must be maintained by the preserver. To do so, the preserver must produce authentic copies of the records in question, because the production of authentic copies is the only way of ensuring their preservation. This fact derives from the nature of electronic records.

In electronic records, the physical and intellectual parts do not necessarily coincide, and the concept of digital component (physical part) accompanies that of element of documentary form (intellectual part). A digital component is a digital object that contains all or part of the content of an electronic record, and/or the data or metadata necessary to order, structure, or manifest the content, and that requires specific methods for preservation. In addition, these other conditions exist: the relation between a record and a computer 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 computer file without changing the record. Thus, the risks of corruption and loss are very high, and become

very complex when records go across technological boundaries. To minimize these risks, controls of two types are implemented: 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. These controls are technological in nature but are determined on the basis of the theoretical understanding of the structure of the record, because it is impossible to maintain literally unaltered an electronic record. What is possible to do is to protect those components of the record that include the elements of form conveying its meaning.

In other words, it is not possible to maintain an electronic record, but only to maintain the ability to reproduce it. To make possible the reproduction of an electronic record, it is necessary first to store its digital components; second, to reassemble all its digital components in the correct order; third, to render the components, individually and collectively, in the correct documentary form (i.e. the elements of the record that constitute its external appearance and convey the action in which it participates and the immediate context in which it was created must appear on the face of the record and in its profile as they were originally); and, fourth, to reestablish the relationships between the record in question and all the other records that belong into the same archival unit. This requires reestablishing the structure of the archival unit and filling it with the records that belong into it. However, to prove that a record so rendered is authentic requires either its inclusion in the set of procedures prescribed by the Benchmark Requirements, added to the fact that the creator is still relying on the record in the usual and ordinary course of business, or a declaration produced by the preserver in a certificate of authenticity. While certifying authenticity can be easily done at the time when a presumption of authenticity is established on the basis of the Benchmark Requirements, or when a verification of authenticity occurs, after the transfer of the authentic records to their custodian, it is only possible if the preserver's procedures are also controlled by strict requirements. In fact, while an electronic copy of an authentic electronic record is authentic if attested to be so by the official preserver, such attestation must be supported by the preserver's ability to demonstrate that it has satisfied all the baseline requirements for the production of authentic copies. Only by virtue of this attestation, the copy is deemed to conform to the record it reproduces until proof to the contrary is shown. For this reason, the second set of requirements developed by InterPARES, the Baseline Requirements, directed exclusively to the preserver, must all be implemented at the highest degree.

The Baseline Requirements are as follows. The first requirement prescribes that the procedures and system(s) used to transfer records to the preserving institution or program, maintain them, and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically 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

As part of the transfer process, the assessment of the authenticity of the records, which had occurred during the appraisal process, should be verified by ensuring that the attributes relating to the records' identity and integrity have been carried forward with the records themselves (Benchmark Requirement 1), along with any relevant documentation (Benchmark Requirement 8). Once the records have been transferred, the preserver must establish many of the controls that were described in the Benchmark Requirements, that

is, must establish access privileges concerning the access, use and reproduction of the records within the archives, implement and monitor them; must establish procedures to prevent, discover, and correct loss or corruption of records, as well as procedures to guarantee the continuing identity and integrity of the records against media deterioration and across technological changes; and, if authentication is required, must establish rules determining responsibility for and means of authentication.

The second requirement prescribes that the activity of reproduction be documented, and this documentation includes:

- 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, the details of this information made readily accessible to the user.

The documentation of the reproduction process is essential for the preserver to fulfil the role of trusted custodian of the record, for the user to have access to the history of reproduction, which becomes an integral part of the history of the record, and for future generations to be able to verify the authenticity of the records.

The third requirement prescribes that the archival description of the archival body, or fonds, containing the electronic records includes–in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts– information about changes the electronic records of the creator have undergone since they were first created.

It has always been the function, either explicit or implicit, of archival description to authenticate the records in context and to perpetuate their administrative and documentary relationships, but with electronic records, this function has become indispensable. In fact, as original electronic records disappear and an interminable chain of non-identical reproductions follows them, the researchers looking at the last of those reproductions cannot find in it any information regarding provenance, authority, context, or authenticity. The authentication function of archival description is different from that of a certificate of authenticity, because it is a collective attestation of the authenticity of the records and of all their interrelationships as made explicit in the description rather than being simply an attestation of the authenticity of individual records. One could say that, given the mandatory documentation of each reproduction process carried out by the preserver, for the purposes of demonstrating the authenticity of the records copies themselves archival description is superfluous. However, if archival description summarizes the history of all reproductions, it obviates the need to preserve permanently all the documentation of each reproduction and acts as a certificate of authenticity for the fonds.

It was quite clear to the InterPARES team that, in light of all the above, the traditional archival concept of "unbroken chain of custody," which used to ensure the authenticity of records over time, must be extended to include the processes necessary to ensure the unaltered transmission of the record through time and become an "unbroken chain of preservation," which 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. It also appeared evident that technology cannot determine the solution to the permanent preservation of electronic records; on the contrary, archival needs must define the problems and archival principles must establish the correctness and adequacy of each technical solution. Finally, it was a sober realization that solutions to the preservation problem are inherently dynamic, thus ongoing research is vital to deal with the challenges presented by new information technologies.

But, what sort of research? Has the InterPARES approach yielded the expected results, at least up to this point? Would a less theoretical approach have served the profession better and faster?

The use of clear theoretical concepts and principles has allowed us to see very easily in the course of the analysis of case studies which systems were designed to contain data rather than records. It has showed us what attributes of a record's identity are implicit in the system and need to be made explicit and linked to the record to ensure that they are not lost when the record is removed from the system. It has revealed the fundamental indifference of the records creator to the issue of authenticity, due to unfunded confidence in technology, and it has supported the identification of the requirements for a presumption of authenticity and for its preservation over time.

However, approaching practical challenges starting from theory may be problematic. For example, the classic concept of record has limited our capacity to understand electronic systems containing a variety of complex entities that do not correspond to it, because that which is known is not always very useful to understand the unknown. Thus, now that, in the second phase of InterPARES, we are dealing with interactive, dynamic and experiential records, we start wondering whether giving them a fix form would not falsify their nature, whether would not be more important to fix the traces of changes or even to describe the interactions, for example, than to fix the record at a given moment in time. One could also think of a record as existing in two separate modes, as an entity to be consulted and as an entity in continuing becoming, and manage them separately. This issue is presently under investigation, but certainly at least an expansion of the existing theory of the record is needed. The other problem that we encountered by starting with theory is that theory tends to decontextualize the record and is therefore general, while the variety and complexity of systems requires attention to details: a complementary inductive approach may be necessary.

Nonetheless, when everything we are confronted with is new and difficult to handle, theory remains our only reference point. Take, for example, the digital signature. Almost immediately we saw the enormous preservation problem that it presented. The question we had to answer was whether we could eliminate it from the documents that we wished to preserve over the long-term without falsifying the document. Thus, we examined the function of the digital signature and realised that it is not a signature but a seal, being the document finished and complete before its attachment. As a consequence, as long as we maintained the information in form of metadata that the document originally had a digital signature, we could eliminate it, just as we used to do with the wax seals on the closed letters of a century ago.

In conclusion, I believe that it is possible to bridge the divide between theory and practice, as long as we recognize that it is necessary to approach the challenges presented

to us by the new technologies using an inter/multi-disciplinary perspective and a direct analysis of our objects of inquiry, and that we must test our theoretical hypotheses, and later our findings in real settings. Whether such belief will borne out will be revealed by the progress of the second phase of InterPARES which can be followed on the InterPARES website at <www.interpares.org>.