Bridging the divide: from theory to practice

Luciana Duranti

Modern office

- Hybrid documentary systems
- Digital environments that support the manipulation of data
- Proprietary and idiosyncratic nature of applications
- Decentralization of records creation
- Repurposing of records
- Obsolescence of systems and media
- Requirements of regulatory agencies (e.g. SEC)

One system: why not?

- Printing out is an impediment to the workflow in the office
- Many digital records are not printable
- Digitization is expensive in the long term
- Courts' decisions have been against routine reproduction

Consequences

- Records are less reliable (manipulability), retrievable (incongruence of classifications), accessible, readable or intelligible (incompatibility and obsolescence)
- It is difficult to prove or verify the authenticity of records (no more originals—the best evidence)

Attempts to solutions

- 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

Attempts to solutions

Research Projects:

- 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)

Results

Good foundation for the development of trusted record-keeping systems, capable of ensuring the reliability and authenticity of the records they contain.

Definition of the fundamental concepts and methods that must be respected to control the trustworthiness of records through their life-cycle, including the concepts of record, reliability and authenticity.

Record

Any document created (i.e., made or received and set aside for further action or reference) by a physical or juridical person in the course of a practical activity as an instrument and by-product of it.

Document

Recorded information

Information

A message intended for communication across space or time

Data

The smallest meaningful piece of information

Electronic Record

A record created (i.e., made or received and set aside for action or reference) in electronic form

A Trustworthy Record

A record that is reliable and authentic

Reliability

The ability of a record to stand for the facts it is about.

In other words, the trustworthiness of the record as a statement of facts.

Authenticity

Refers to the fact that a record is what it purports to be and has not been tampered with or otherwise corrupted.

In other words, the trustworthiness of the record as a record.

Conceptual Framework for Authenticity

- 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

Identity of a Record

- It refers to the attributes of a record that uniquely characterise 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).

Integrity of a Record

- 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

Authentication

- A declaration of authenticity, resulting either by the insertion or the addition of an element or a statement to a record, and the rules governing it are established by legislation.
- A means of proving that a record is what it purports to be at a given moment in time (digital signature), as opposed to a quality of the record.

What about long-term preservation?

What does long-term mean?

- For most businesses and industry and their regulatory bodies, 2-3 years
- For governments and public bodies, anything from 30 years to forever
- For most individuals, their work-life or their lifetime. If they are artists, it may mean forever

Two ways of addressing long-term preservation

- To those who must keep their non current records for 2-3 years in a way that their integrity can be proven, according to the directions of their regulatory bodies, the issue appears to be one of reliable storage.
- To those who must keep their non-current records for an undetermined period, possibly decades and centuries, the issue is much more complex, as it involves a chain of preservation procedures that begin with the design of digital systems and continues through records creation, maintenance, use, and repeated refreshment, reproduction, migration, and use by third parties.

An ancillary issue for both regards the validity of technological means of authenticating the records and the need and/or ability to preserve them across technological obsolescence, as will be discussed by Dr. Blanchette and Mr. Pinkas.

Today's conference

- Mr. Willems will discuss the first way, the storage issue
- Dr. Underwood and Ms. Hackett will discuss the second way, the procedural issues
- I will introduce the whole challenge they are facing by providing an overview of the key theoretical and methodological findings of the only research project that has dealt at international and multidisciplinary level with the long-term preservation of digital records

InterPARES: **INTE**rnational **R**esearch on **P**ermanent **A**uthentic **R**ecords in **E**lectronic **S**ystems

An International Collaborative Research Initiative

InterPARES Project
Dr. Luciana Duranti
Project Director

Research Goal

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.

Identifiable Characteristics of an Electronic Record

- 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)
- Unchangeable content
- Explicit linkages to other records within or outside the digital system through a classification code or other unique identifier
- Identifiable administrative content
- Author
- Addressee
- Writer
- Participant in or supporting an action either procedurally or as part of the decision making process

Other Characteristics

- Elements of form are different from digital components
- 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 preservation
- 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 viceversa, from one digital presentation a variety of record presentations can derive
- It is possible to change the way in which a record is contained in a file without changing the record

Consequences

- It is not possible to preserve an electronic record, it is only possible to preserve the ability to reproduce it
- The first step in reproducing an electronic record is to reassemble all its digital components in the correct order
- The second step is to render the components, individually e 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)
- The third step is to reestablish the relationships between the record in question and all the other records that belong into the same archival unit. This requires, first, to reestablish the structure of the archival unit, and then, to fill it with the records that belong into it.

The Records of the Creator

- The records that exist as created. They are considered authentic because they are the same as they were in the first instantiation
- The records that have undergone some change and therefore cannot be said to exist as first created. They are considered authentic because the creator treats them as such by relying on them for action or reference in the regular conduct of business. However, their authenticity is threatened whenever they are transmitted across space or time. Therefore, an inference of their authenticity must be further supported by evidence that they have been maintained in a way that guarantees their continuing identity and integrity.

Presumption of Authenticity

An inference that is drawn from known facts about the manner in which a record has been created and maintained. The evidence supporting it is enumerated in Benchmark Requirements for authenticity. A presumption of authenticity will be based upon the number of requirements that have been met and the degree to which each has been met.

Verification of Authenticity

- 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, when there is an insufficient basis for a presumption of authenticity.
- It involves a detailed examination of the record in all their contexts and of reliable information available from other sources (audit trails, backups, copies preserved elsewhere, textual analysis)

Benchmark Requirement A1:
Expression of Record Attributes & Linkage to Record

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

Benchmark Requirement A1:

- *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

Benchmark Requirement A1:

- *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:
Access Privileges

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

Benchmark Requirement A3:
Protective Procedures: Loss and Corruption of Records

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

Benchmark Requirement A4:
Protective Procedures: Media and Technology

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:
Establishment of Documentary Forms

The creator has established the documentary forms of records associated with each procedure either according to the requirements of the juridical system or those of the creator

Benchmark Requirement A6:
Authentication of Records

If authentication is required by the juridical system or the needs of the organization, the creator has established specific rules regarding which records must be authenticated, by whom, and the means of authentication

Benchmark Requirement A7:
Identification of Authoritative Record

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

Benchmark Requirement A8:
Removal and Transfer of Relevant Documentation

If there is a transition of records from active status to semi-active and inactive status, which involves the removal of records from the electronic system, the creator has established and implemented procedures determining what documentation has to be removed and transferred to the preserver along with the records

Selection of electronic records

- compiling information on the context of creation and on the technological context: it allows to establish the basis upon which the records are considered authentic and to determine the value of the records
- assessing the continuing value of the records and their authenticity, and determining their overall value

Selection of electronic records

- deciding whether the digital components embodying the essential elements that confer identity and ensure the integrity of the records can be preserved, given the preserver's current and anticipated capabilities: the appraiser determines both the record elements containing informational content and those elements that need to be preserved according to requirements for authenticity, identifies where these crucial record elements are manifested in the digital components of the electronic record, and 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.

- making the appraisal decision, deciding what must be either transferred to the archives or destroyed (including the list of the digital components) and how and when this should happen

Monitor selected records

The records selected for preservation must be monitored, especially for 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, monitoring produces minor revisions to the documentation on the selection and to the terms and conditions of transfer

Preparation of the records for disposition and transfer

Copying and, if necessary, formatting the records selected for preservation so as to prepare them physically for transfer, and to package them 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

Key points related to selection

- It is essential for the appraiser to assess the authenticity of electronic records
- It is important to determine the feasibility of preservation
- It is vital that electronic records be appraised early in their life cycle
- It is important to monitoring the appraisal decision

Baseline Requirements for the Production of Authentic Copies

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

Baseline Requirement B1:

Controls over Records Transfer, Maintenance, and Reproduction

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

- 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:
Documentation of Reproduction Process and its Effects

The activity of reproduction has been 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, such information has been documented by the preserver, and this documentation is readily accessible to the user

Baseline Requirement B3:
Archival Description

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

Fundamental concept: chain of preservation

- 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
- 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

Manage the preservation function

On the basis of the information accumulated during the selection function, it is necessary to develop a strategy of preservation, action plans (each connected to a specific body of records selected for preservation), a technologic infrastructure and the methods necessary to implement the action plans.

Maintain electronic records

To apply 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 produce again (re-produce) the records when requested, to certify their authenticity, and to make possible their comprehension. To produce information used to evaluate the execution of this function and satisfy the baseline requirements for authenticity

Provide access to electronic records

To apply the re-production method established for the body of records in question and to implement the action plan to produce an authentic copy of the record in response to a request of access. If specified in the request, to generate a certificate that attests the record's authenticity. Alternatively, if requested, to give the user a reproducible electronic records, that is, the record's digital components with the instructions for rendering it as an authentic copy and with the information necessary to comprehend it.

Key points concerning preservation

- Technology cannot determine the solution to the long-term preservation of electronic records
- Archival needs define the problem and archival principles must establish the correctness and adequacy of each technical solution
- Solutions to the preservation problem are inherently dynamic

Intellectual framework: principles

What every organization must develop is a records preservation policy and a strategy that

- **address records specifically rather than digital objects generally**
- **focus on authentic electronic records**
- **recognize and provide for the fact that authenticity is most at risk when records are transmitted across space (that is, when sent between persons, systems, or applications) or time (that is, either when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced)**
- **recognize that preservation of authentic electronic records is a continuous process that begins with the process of records creation and whose purpose is to transmit authentic records across time and space**

Intellectual framework: principles

- be based on the concept of trust in records keeping and record preservation and specifically on the concepts of a trusted record-keeping system and the role of the preserver as a trusted custodian

- be predicated on the understanding that it is not possible to preserve an electronic record as a stored physical object: it is only possible to preserve the ability to reproduce the record

- recognize that the physical and intellectual components of an electronic record do not necessarily coincide and that the concept of digital component is distinct from the concept of element of documentary form

- specify the requirements a copy of a record should satisfy to be considered equivalent to an original

Intellectual framework: principles

- integrate records appraisal and archival description in the continuous process of preservation
- explicitly state that the entire process of preservation must be thoroughly documented as a primary means for protecting and assessing authenticity over the long term
- explicitly recognize that the traditional principle that records relied upon in the usual and ordinary course of business can be presumed to be authentic needs to be supplemented in the case of electronic records by evidence that the records have not been inappropriately altered

Intellectual framework: principles

- recognize that the preserver is concerned with both the assessment and the maintenance of the authenticity of electronic records. The assessment of the authenticity of electronic records takes place before records are transferred to the custody of the preserver as part of the process of appraisal, while the maintenance of the authenticity of copies of electronic records takes place once they have been transferred to the preserver's custody as part of the process of long-term preservation

- draw a clear distinction between the preservation of the authenticity of records and the authentication of a record

Reference

For the findings of InterPARES see the project's web site:

http://www.interpares.org