The Preservation of Digital Records: the InterPARES approach (on the basis of its findings)

Mariella Guercio
TEAM Italy Director
SPP/ICA
Dakar, Sengal
22 October 2010
The Goal of InterPARES

To develop the body of theory and methods necessary to ensure that digital records can be created in accurate and reliable form and maintained and preserved in authentic form, both in the long and the short term, for the use of those who created them and of society at large, regardless of technology obsolescence and media fragility; and to develop methods of implementations that can achieve such purposes regardless of the available resources and in respect of organizational culture.
The risk of the technological focus of the national legislators

• The rules approved at national and multi-national levels have generally created more risks and complexities than the technologies themselves,

• The infrastructures based on the bits control and on public key systems are expensive and not always necessary,

• The relevance of qualified record making, record managing and record keeping systems is often ignored even for public sectors

• The preservation of digital archives is considered a question to delegate to the final custodians
The risk of the technological focus of the national legislators (cont)

• The legislators have rarely understood all the implications of the principles of authenticity and integrity and have reduced the solutions to a large use of digital signatures and timestamps.

• Fundamentally, they have handed the maintenance of our memory to a sort of an interrupted chain of technological controls and devices, while the main question is – not so differently than in the past – the capacity to build in new and sustainable forms an uninterrupted chain of verifiable responsibilities and trustworthy custodians.
Policy Framework

A framework of principles guiding the development of policies for records creating and preserving organizations
Key IP 1 & 2 Final Products (cont)

Creator Guidelines

Recommendations for making and maintaining digital materials for individuals and small communities of practice.
Key IP 1 & 2 Final Products (cont)

Preserver Guidelines

Recommendations for digital preservation for archival institutions
Benchmark and Baseline Requirements

Authenticity requirements for assessing and maintaining the authenticity of digital records

**< Requirement Set A >**

To support a presumption of authenticity, the preserver must obtain evidence that:

**Requirement A.1:** Expression of Record Attributes and Linkage to Record
The value of the following attributes are explicitly expressed and statistically linked to every record. These attributes can be distinguished into two categories, the first concerning the identity of the record, and the second concerning the integrity of records.

**A.1.1** Identity of the record
- **A.1.1.1** Names of persons concerned in the formation of the record
  - Name of author(s)
  - Name of editor(s) (if different from the author)
  - Name of original(s) (if different from name of author)
  - Name of addressee(s)
- **A.1.1.2** Name of action or matter
- **A.1.1.3** Dates of creation and transmission, that is:
  - Chronological data
  - Received data
  - Archival data
- **A.1.1.4** Indication of attachments

**Requirement A.2:** Access Privileges
The creator has defined and effectively implemented access privilege modifications, annotation, relocation, and destruction of records.

**< Requirement Set B >**

The preserver should be able to demonstrate that:

**Requirement B.1:** Controls over Records Transfer, Maintenance, and Reproduction
The procedures and system(s) used to transfer records to the archival institution or program, maintain them, and reproduce them embody adequate and effective controls to guarantee the record’s identity and integrity, and specifically that:

- B.1.a Unbroken custody of the records is maintained;
- B.1.b Security and control procedures are implemented and monitored;
- B.1.c The content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.

**Requirement B.2:** Documentation of Reproduction Process and its Effects
The activity of reproduction has been documented, and this documentation includes:

- B.2.a The date and time of the records' reproduction and the name of the responsible person;
- B.2.b The relationship between the records acquired from the creator and the copies produced by the preserver;
- B.2.c The impact of the reproduction process on their form, content, accessibility, and use.

In those cases where a copy of a record is known to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user.

**Requirement B.3:** Archival Description
The archival description of the record containing the electronic records includes—information about the records' administrative, provenance, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created.
Key IP 1 & 2 Final Products (cont)

File Format Selection Guidelines

Principles and criteria for adoption of file formats, wrappers and encoding schemes
Key IP 1 & 2 Final Products (cont)

Terminology Database
Including a glossary, a dictionary and ontologies
Key IP 1 & 2 Final Products (cont)

Two Records Management Models

Chain of Preservation (COP) Model (lifecycle)
Business-driven Recordkeeping (BDR) Model (continuum)
Key IP 1 & 2 Final Products (cont)

Two books:


Most Important Findings

• Conceptual
  – The Concept of Record
  – The Concept of Chain of Preservation
  – The Concept of Trustworthiness

• Methodological
  – Appraisal
  – Preservation

• Strategic
  – Relationship Creator-Preserver
  – The Role of the Archivist
The Concept of Record

- **Record**: any document created (i.e., made or received and set aside for action or reference) by a physical or juridical person in the course of activity as an instrument and by-product of it
- **Document**: recorded information (i.e., information affixed to a medium in an objectified and syntactic form)
- **Information**: a message intended for communication across time and space
- **Data**: the smallest meaningful piece of information
Digital Record Characteristics

- **Medium**: necessary part of the technological context, not of the record
- **Stable Content and Fixed Form**
- **Archival Bond**: explicit linkages to other records inside or outside the system
- **Five Necessary Persons**: author, writer, originator, addressee, and creator
- **Act**: an action in which the records participates or which the record supports
- **Five Necessary Contexts**: juridical-administrative, provenancial, procedural, documentary, technological
Fixed Form

- An entity has fixed form if its binary content is stored so that the message it conveys can be rendered with the same documentary presentation it had on the screen when first saved (different digital presentation: Word to .pdf).
- An entity has fixed form also if the same content can be presented on the screen in several different ways in a limited series of possibilities: we have a different documentary presentation of the same stored record having stable content and fixed form (e.g. statistical data viewed as a pie chart, a bar chart, or a table).
Stable Content

- An entity has stable content if the data and the message it conveys are *unchanged and unchangeable*, meaning that data cannot be overwritten, altered, deleted or added to.

- **Bounded Variability**: when changes to the documentary presentation of a determined stable content are limited and controlled by fixed rules, so that the same query or interaction always generates the same result, and we have different views of different subsets of content, due to the intention of the author or to different operating systems or applications.
The Parts of a Digital Record

- **Formal Elements**: constituent parts of the record documentary form as shown on its face, e.g. address, salutation, preamble, complimentary close
- **Metadata**: the attributes of the records that demonstrate its identity and integrity
- **Digital Components**: stored digital entities that either contain one or more records or are contained in the record and require a specific preservation measure
**Stored and Manifested Record**

- **Stored record**: it is constituted of the digital component(s) used in re-producing it, which comprise the data to be processed in order to manifest the record (content data and form data) and the rules for processing the data, including those enabling variations (composition data).

- **Manifested record**: the visualization of the record in a form suitable for presentation to a person or a system. Sometimes, it does not have a corresponding stored record, but it is re-created from fixed content data when a user’s action associates them with specific form data and composition data (e.g. a record produced from a relational database).
Static and Interactive Records

- **Static Records**: They do not provide possibilities for changing their manifest content or form beyond opening, closing and navigating: e-mail, reports, sound recordings, motion video, snapshots of web pages.

- **Interactive Records**: They present variable content, form, or both, but the rules governing the content and form of presentation are fixed. Ex. Interactive web pages, online catalogs, records enabling performances.
Interactive Entities

• **Not-dynamic**: the rules governing the presentation of content and form do not vary, and the content presented each time is selected from a fixed store of data. Ex. Interactive web pages, online catalogs, records enabling performances—they are records

• **Dynamic**: the rules governing the presentation of content and form may vary—they are potential records
New Concept of Records Life-Cycle

Based on the recognition that we cannot maintain or preserve digital records, but only the ability to re-produce or re-create them, and that re-productions and re-creations of digital records

• if made by the creator in the course of and for the purposes of its business, are records of the creator, while

• if made by the preserver in the course and for the purposes of archival functions, are authentic copies of the records of the creator
As a Consequence...

• The preserver can only preserve what it receives from the creator by making an authentic copy of it, and has no right to stabilize it or alter its documentary form—only its digital presentation, or format.

• Whether the stabilized record of the creator and its authentic copy made by the preserver are to be considered trustworthy depends on the context in which they are created and used but also on the capacity to document the preservation processes.
Trustworthiness

Reliability
The trustworthiness of a record as a statement of fact, based on:
• the competence of its author
• the controls on its creation

Accuracy
The correctness and precision of a record’s content, based on:
• the competence of its author
• the controls on content recording and transmission

Authenticity
The trustworthiness of a record that is what it purports to be, untampered with and uncorrupted, based on:
• identity
• integrity
**Authenticity: Identity**

The whole of the attributes of a record that characterize it as unique, and that distinguish it from other records.

**Identity metadata:**

- names of the 5 persons concurring in its creation
- date(s) and time(s) of issuing, creation and transmission
  - the matter or action in which it participates
    - the expression of its archival bond
      - documentary form
      - digital presentation
    - the indication of any attachment(s)
      - digital signature
- name of the person responsible for the business matter
Authenticity: Integrity

A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered.

**Integrity metadata:**

- name(s) of handling persons over time
- name of person responsible for keeping the record
  - indication of annotations
  - indication of technical changes
- indication of presence or removal of digital signature
  - time of planned removal from the system
  - time of transfer to a custodian
  - time of planned deletion
- existence and location of duplicates outside the system
Authentication

A means of declaring the authenticity of a record at one particular moment in time -- possibly without regard to other evidence of identity and integrity.

Example: the digital signature. Functionally equivalent to medieval seals (not signatures):

- verifies origin (identity)
- certifies intactness (integrity)
- makes record indisputable and incontestable (non-repudiation)

The analogy is not perfect, because the medieval seal was associated exclusively with a person, while the digital signature is associated with a given person and a specific record, and because the former is an expression of authority, while the latter is only a mathematical expression.
Digital Records Preservation

1. Establish a digital preservation program
2. Appraise digital records
3. Transfer digital records to the preserver’s custody
4. Process digital records
5. Create archival description
6. Store digital records
7. Convert and migrate digital records
Establish a Digital Preservation Program

- Establish scope and objectives
- Acquire resources
- Offer advice (for creators)
- Set a good example:
  - preservers must establish, within their own organization, a record-making and a recordkeeping environment; ensure hands-on training to archivists; develop a test-bed where upgrades and innovations can be introduced and evaluated and a working prototype that can be used in demonstrations
Establish a Digital Preservation Program (cont.)

• Develop procedures: controls over records transfer, maintenance and reproduction to ensure that
  – unbroken custody of the records is maintained;
  – security and control procedures are implemented and monitored;
  – the content of the records and the required information, including metadata, remain unchanged after reproduction
Establish a Digital Preservation Program (cont.)

• Implement maintenance strategies, including
  – A1. Clear allocation of responsibilities
  – A2. Provision of appropriate technical infrastructure
  – A3. Implementation of a plan for system maintenance, support and replacement
  – A4. Implementation of a plan for the transfer of records to new storage media on a regular basis
  – A5. Adherence to appropriate storage and handling conditions for storage media
  – A6. Redundancy and regular backup of the digital objects
  – A7. Establishment of system security
  – A8. Disaster planning
Appraise Digital Records

1. As early as possible in the life of the records, assess their continuing value to the creator (after having identified multiple owners) and/or the designated preserver

2. Assess the authenticity of the records considered for continuing preservation, and document the process

3. Determine the feasibility of their preservation (after having identified all digital components) by the designated preserver (creator and/or delegated archives); and,

4. Constantly monitor all the records of the creator and, if warranted by the changes that they have undergone through time, revise the appraisal decision
Document the Authenticity Assessment

• When appraising records created in a digital environment, the assessment of the authenticity of records must become a more overt, visible process performed and documented by the preserver with reference to:
  – unbroken chain of custody,
  – knowledge of recordkeeping practices
• The appraisal report should document the controls put in place by the creator to guarantee the identity and integrity of the records and thus the presumption of their authenticity.
The controls documented in the report include:

• A.1 Expression of Record Attributes and Linkage to Record (e.g., identity and integrity metadata)
• A.2 Access Privileges
• A.3 Protective Procedures against Loss and Corruption of Records
• A.4 Protective Procedures against Media Deterioration and Technological Change
• A.5 Establishment of Documentary Forms
• A.6 Authentication of Records
• A.7 Identification of Authoritative Record
• A.8 Removal and Transfer of Relevant Documentation
Determine the Feasibility of Preservation

- It means deciding whether the digital components (to be carefully identified) 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: a careful investigation of the technical preservation requirements for preservation is required with reference to the costs analysis.

- A digital component is a digital object that contains all or part of the digital record, and/or data or metadata necessary to order, structure, or manifest its content, and that requires specific methods for preservation.

- Digital components are to the stored record what elements of form are to the manifested record.
Determine the Feasibility of Preservation: an Example

• One common digital component is the library of fonts, any number of which can be selected by the creator to be used in the presentation of a word-processed document.

• In Windows, the fonts are stored in ‘.dll’ (or dynamic link library) files.

• For the preserver to be able to reproduce this record to reflect the creator’s original intentions, both the digital component containing the text and the digital component containing the font must have been preserved, as well as the link between them established in such a way that the software attempting to display the content of the text file can find the appropriate font library.
Monitor the Records

• All the records (not only those appraised for continuing preservation) and all the digital components must be checked on a regular basis.

• This involves looking for changes both in their technological context and in their use.

• In some cases it may be necessary to repeat the appraisal because of changes that can affect the feasibility of preservation or because the records result from other functions or present different characteristics.

• In most cases, monitoring produces minor revisions to the documentation on the selection and to the terms and conditions of transfer.
Acquire Selected Digital Records

• Develop shared plan for transfer:
  – a plan agreed upon by both parties is required
    (including – if applicable – a common decision on
    the physical and logical formats for transfer)

• Enforce standardised procedures:
  – See the ICA draft standard for digital records transfer
    (Record Exchange Standard BRS (Business
    Requirements Specification))
Acquire Selected Digital Records (cont.)

• Keep the oldest available logical format:
  – the original format should, whenever feasible, be maintained by the preserver, in addition to any preservation or reference copies generated after the transfer.

• Avoid duplicates
Acquire Selected Digital Records (cont.)

• **Document** all processing:
  – *why* certain processes were applied to the records;
  – *what* records were processed;
  – the *date* when the process was performed;
  – the *names* of persons performing and documenting the various steps of the process(es);
  – the *impact* of the process performed on the records’ form, content, accessibility and use; and
  – the *description* of any damage, loss or other problems encountered as a result of the processing, including any effect on the elements expressing the records’ identity and integrity
Preservation

It involves the creation of authentic copies of the records of the creator. Their authenticity is guaranteed by:

- a **controlled process** of migration of the acquired records to the archives technological environment (always keeping the records in the format in which they were acquired)
- the **accurate documentation** of any change that the records undergo during such process and every time that the archives technological environment is upgraded
- the **implementation and monitoring of privileges** concerning the access, use and reproduction of the records within the archives
Preservation (cont.)

• the establishment of *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 (ongoing conversion and migration) also by maintaining proper storage; and
• if authentication of individual records is required, by the existence of *rules determining responsibility for and means of authentication*.
Archival Description

Archival description acquires a primary authentication function

• The function of archival description is to provide an historical view of the records and of their transformations while maintaining the bond of their common provenance and destination

• It should include the information about the records (and the relationships among digital components) and their contexts collected during the appraisal and processing stages, including the documentation related to the transfer and all the preservation processes
What Else Is Needed?

- The unbroken chain of preservation must begin at creation and continue from the record-making system to the recordkeeping system and the record preservation system.

- Keeping in mind that solutions to digital preservation are always specific, respect the organizational culture: the collection of values and norms that are shared by people and groups in an organization and that control the way they interact with each other and with stakeholders outside the organization.

- The new emphasis on accountability allows the archives to fulfill these needs by presenting itself as the trusted custodian.
Archivist as Trusted Custodian

The trusted custodian is a person who

- acts as a **neutral third party**, i.e., demonstrates that he/she has no stake in the content of the records and no reason to alter records under his/her custody, and that he/she will not allow anybody to alter the records either accidentally or on purpose,

- is equipped with the **knowledge and skills** necessary to fulfil its responsibilities, which should be acquired through formal education, and

- establishes a **trusted preservation system** that is capable of ensuring that accurate and authentic copies of the creator’s records are acquired and preserved
InterPARES 3 Web Site

www.interpares.org

InterPARES 3 Project
International Research on Permanent Authentic Records in Electronic Systems

Welcome to the InterPARES 3 Project
The International Research on Permanent Authentic Records in Electronic Systems (InterPARES) Project is an international collaborative endeavour, composed of several national and multinational teams. The direction of the entire project is funded by Canada’s Social Sciences and Humanities Research Council Community-University Research Alliances grant, which also supports TEAM Canada, the name of which derives from the specific title given to this third phase of the project. Theoretical Elaborations into Archival Management (TEAM): Implementing the theory of preservation of authentic records in digital systems in small and medium-sized archival organizations. The members of the InterPARES 3 Project International Alliance are:

- TEAM Africa
- TEAM Brazil
- TEAM Canada
- TEAM Catalonia
- TEAM China
- TEAM Netherlands and Belgium
- TEAM United Kingdom and Ireland

Objectives

Objetivo geral
O objetivo geral do InterPARES 3 é capacitar as organizações arquivísticas públicas e privadas de médio e pequeno porte, bem como programas, que são responsáveis por documentos arquivísticos digitais resultantes de atividades de governo, de negócios, de pesquisa, da arte e do entretenimento, sociais e outros comunidades, a preservar ao longo do tempo documentos arquivísticos autênticos, que satisfazem as necessidades da instituição e da sociedade de ter um registro confiável de seus passados.

Objetivos específicos

1. propiciar um ambiente que apoie o objetivo geral da pesquisa, demonstrando aos organismos reguladores e de auditoria e aos responsáveis pela elaboração de políticas que é essencial integrar os requisitos de preservação digital em todas as atividades que são por eles reguladas, auditiadas ou controladas;

2. colaborar com arquivos de pequeno e médio porte no desenvolvimento de políticas, estratégias, procedimentos e/ou planos de ação escalonáveis, que possam ser implementados com o objetivo de preservar o material digital sob sua guarda, ou que serão recebidos, por eles, utilizando recomendações