InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

InterPARES 1 & 2 Overview: Objectives, Methodology and Outcomes (1999-2007)

Randy Preston
InterPARES Project Coordinator

LongRec Team Meeting
Trondheim, Norway
9 May 2007

InterPARES 1 (1999-2001): 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.

InterPARES 1: Funding and Participants

- Major funding from SSHRC, NHPRC, UBC, NARA, NAC
- 13 countries in 4 continents, 60 researchers
- Public and private sector
- Archival science, diplomatics and records management, plus computer science and engineering, jurisprudence, research methods

What Did InterPARES 1 Study?

- Mostly textual records born digital in the course of administrative and legal activities
- Databases and document management systems
- Inactive or non-current records
- Characteristics of e-records and concept of authenticity
- Activities of appraisal and preservation from the preserver point of view

Team Structure

Divided into four **Domain Task Forces**:

1. **Authenticity**: Goal was to identify conceptual requirements for assessing & maintaining authenticity of digitalrecords.
2. **Appraisal**: Goal was to determine whether selection of digital records should be based on the same or different criteria as those for traditional records and how digital technologies affect appraisal methodology.
3. **Preservation**: Goal was to develop preservation methods for authentic digitalrecords.
4. **Strategy**: Goal was to develop an intellectual framework for articulation of international, national and organizational policies, strategies and standards for long-term preservation of authentic digital records.

Key Concepts

- **Record** - any document created (i.e., made or received and set aside for action or reference) by a physical or organizational person in the course of a practical activity as an instrument and by-product of it. A record is more than:
  - **Document** - information affixed to a medium in a determined form
  - **Information** - a message intended for communication across space or time
  - **Data** - the smallest meaningful piece of information
- **Digital record** - a record created (i.e., made or received <u>and</u> set aside for action or reference) in digital form

Characteristics of a Digital Record

A digital record must have:

- **stable content** and **fixed form**;
- **identifiable administrative and documentary contexts**, through explicit linkages to other records within or outside the digital system; and
- **five persons** (author, writer, originator, addressee, creator) involved in its creation.
- It must **participate in or support an action** either procedurally or as part of the decision making process; and
- It presents **formal elements, attributes, and digital components**.

Key Concepts (cont.)

- **Reliability** - the trustworthiness of a record as a statement of fact. A reliable record is complete and generated according to a controlled procedure
- **Accuracy** - the exactness and correctness of record content, dependent on the competence of the author and the controls on the process by which data are recorded and transmitted through space (i.e., between persons, systems or applications) and time (i.e., when stored off line, or when the hardware or software used to process, communicate or maintain it is upgraded or replaced)
- **Authenticity** - the trustworthiness of a record as a record. An authentic record is one that has not been tampered with or otherwise corrupted. Authenticity is maintained by protecting record identity and integrity. It differs from

  Authentication - a means of declaring authenticity at a point in time

Key Concepts (cont.)

Authenticity

- **Identity** refers to the attributes of a record that uniquely characterize it and distinguish it from other records. These include: the names of the persons concurring in its creation (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 relationships with other records; and an indication of any attachment(s).
- **Integrity** is the wholeness and soundness of a record. A record has integrity if it is intact and uncorrupted, that is, if the message that it is meant to communicate to achieve its purpose is unaltered. This means that 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.

Key Instruments

- **Template for Analysis**
  - Documentary Form
    intrinsic and extrinsic elements
  - Annotations
  - Contexts
    administrative, provenancial, procedural, documentary, technological
  - Medium

Key Instruments (cont.)

- **Authenticity Requirements**
  - **Benchmark Requirements**: the requirements that support the presumption of the authenticity of digital records before they are transferred to the custody of the preserver
  - **Baseline Requirements**: the requirements that support the production of authentic copies of digital records that are in the custody of the preserver.

Most Important Findings

- Most systems that should contain records do not, because the entities in them **lack fixed form and stable content**. The systems that do, contain bad records, primarily because of **lack of identifiable contexts and relationships**.
- Inactive records that are no longer kept in active systems often cannot be preserved because either they were not created and/or maintained in **preservable formats** or their formats are **obsolete**.

Most Important Conclusions

The preservation of authentic digital records…

- is a **continuous process** that begins with the records creation and whose purpose is to transmit authentic records across time and space;
- must be predicated on the concepts of a **trusted recordkeeping system** and the role of the preserver as a **trusted custodian**; and
- must incorporate records appraisal and archival description.

Most Important Conclusions (cont.)

- The only way of preserving an inactive digital record is to make an **authentic copy** of its last instantiation as an authentic record of the creator
- The preserver must be concerned with both the assessment and the maintenance of the authenticity of digital records throughout their lifecycle

Lessons Learned: IP1 Principles

- Technology cannot determine the solution to the reliable and accurate creation of digital records or to their authentic preservation over the long term: **organizational needs** define the problem and **archival principles** must establish the correctness and adequacy of each technical solution
- Solutions to the digital records challenges are inherently **dynamic** and **specific** to the cultural, disciplinary, administrative and legal situations
- Preservation is a **continuous process that begins with records creation**
- We must be able to **presume records trustworthiness**, until proof to the contrary is established
- We must be able to **infer authenticity** on the basis of the circumstances of records creation, maintenance and preservation

Based on these findings, conclusions & lessons…

We developed a new research project that:

- examined all kinds of digital entities in complex systems and then developed a concept of record;
- was concerned with the entire record lifecycle and developed a chain of preservation model and a business process model beginning with the design of the system and capable of ensuring accuracy, reliability & authenticity;
- studied digital entities created by artists and scientists, who began using complex digital systems long before most governments and businesses; and
- used the concepts and methodology of all the fields touched by our study ("open inquiry" approach).

InterPARES 2 (2002-2006): Goal

To ensure that the portion of society's recorded memory digitally produced in dynamic, experiential and interactive systems in the course of artistic, scientific and e-government activities 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 it and of society at large, regardless of digital technology obsolescence and media fragility.

InterPARES 2: Funding and Participants

- Major funding from SSHRC, NHPRC, NSF, UBC, and UNESCO (for last 18 months)
- 21 countries in 5 continents, 100 researchers
- Public and private sectors
- Archival science, diplomatics and records management; music theory, composition, performance; film theory, production, description; dance and theatre theory; a variety of hard and social sciences; jurisprudence; computer science and engineering

What Did We Study?

- Live, active materials in interactive, experiential and dynamic electronic systems, followed from creation to preservation
- Obsolete inaccessible materials
- Approaches taken by both creators and preservers who have concerned themselves with accuracy, reliability and authenticity

Team Structure

Matrix depicting Focus and Domain Task Forces, Working Groups and Cross-domain Research Teams

	FOCUS 1 Artistic activities	FOCUS 2 Scientific activities	FOCUS 3 Governmental activities
DOMAIN 1 Records creation & maintenance	Working Group 1.1	Working Group 1.1	Working Group 1.3
DOMAIN 2 Authenticity, accuracy & reliability	Working Group 2.1	Working Group 2.2	Working Group 2.3
DOMAIN 3 Methods of appraisal & preservation	Working Group 3.1	Working Group 3.2	Working Group 3.3

Terminology
Policy
Description
Modeling

Key Work Accomplished: Focuses

Case Studies (22 completed):

- represented in activity and entity models and analysed according to diplomatic and archival principles

General Studies (9 completed):

- surveys of: government Web sites; digital recordkeeping practices of photographers, composers, and GIS archaeologists; the practice of preservation of interactive music; file formats and encoding languages used for non-textual materials; analysis of a prototype for a persistent archives based on data grids

Key Work Accomplished: Domains

Development of:

- conceptual responses to the original research questions, primarily concerning identification of the ‘record’ and of the preservable entities

Production of:

- guidelines for records creation and records preservation for individuals and small organizations
- papers outlining the concepts of reliability, accuracy and authenticity in the areas of activity studied by the project

Key Work Accomplished: Cross-domains

Terminology:

- terminology database including 3 lexicographic instruments:
  - *Register, Dictionary, Glossary*
- Ontology for interrelated terms

Policy:

- identification of barriers to preservation that currently exist in laws, regulations, policies and standards concerning copyright and intellectual rights, privacy and FOI, authenticity and authentication, open standards and open source, and records and archival management
- development of a framework of principles for creators and one for preservers

Key Work Accomplished: Cross-domains (cont.)

Description:

- Metadata Schema Registry (Metadata and Archival Description Registry and Analysis System or MADRAS). Available at:

  http://www.gseis.ucla.edu/us-interpares/madras/

- identification of the relationship between metadata and archival description

Modeling:

- Chain of Preservation (COP) Model depicting all activities involved in the management of digital records throughout their lifecycle, from creation to permanent preservation
- Business-driven Recordkeeping (BDR) Model, based on the records continuum concept

Summary: Key Products

- **Framework of principles guiding development of policies** for records creating and preserving organizations
- **Guidelines for making and maintaining digital records** for individuals and small communities of practice
- **Guidelines for digital preservation** for archival institutions
- **Authenticity requirements** for records systems
- **Template for Analysis** for identifying digital records
- **Metadata registry** for the registration and analysis of metadata schemas
- **Terminology database** including glossary, dictionary and ontologies
- **Records management models**
- Principles and criteria for adoption of **file formats, wrappers, and encoding** (General Study 11)