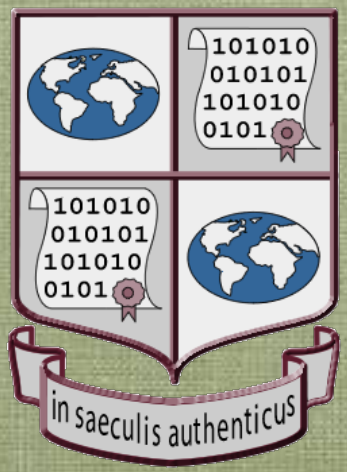
The Future of Our Memory: Lost in Obsolescence? The Memory of Our Future: Lost in Interactivity?



Dr. Luciana Duranti
InterPARES Project Director

Disadvantages of the Digital Medium

- The digital medium does not contain any given record but only bit-strings
- It is not possible to preserve digital records but only the ability to re-produce them
- There is no longer an original (i.e. the first, complete and effective record)
- Authenticity is no longer verifiable on the record itself but must be inferred from information external to the record
- The easiness of reproduction makes it difficult to identify the final version or the official version

…and more

- The Internet makes intellectual property increasingly difficult to protect
- Viruses and technology failures make it easy to lose everything
- Technological obsolescence makes digital materials inaccessible very fast
- Copyright and data privacy laws create serious obstacles to preservation by reproduction or emulation
- The information provided by the materiality of the object no longer exists

…and bad habits make it worse

- Individuals and organizations keep hybrid systems
- They creating records in different applications and leave them there rather that transferring them to a recordkeeping system
- They do not do regular back-up and upgrading of files
- They do not keep magnetic and optical media in the right climatic environments
- They do not refresh the media

and worse…

- Use proprietary or legacy systems and ignore the need for interoperability
- Do not migrate the records to new technology
- Hope that emulation will take care of long-term access
- Do not protect the documents from malicious or accidental tampering—trusting personal or institutional custody
- Use protection systems—encryption or digital signatures—that do not allow for preservation
- Confuse storage with preservation

InterPARES 1 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: 1999-2001

- Major Funding from SSHRC, NHPRC, UBC, NARA, NAC
- 13 countries in 4 continents, 60 researchers
- Public and private sector
- Professionals and academics (ratio 80% to 20%)
- Archival science, diplomatics and records management, plus computer science and engineering, jurisprudence, research methods

What Did We Study?

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

Methodology

- Theory and methods of diplomatics, archival science and law for the definition of concepts and development of requirements and methods
- Grounded theory for selection of case studies
- Comparative analysis for the study of appraisal and preservation reports
- IDEF0 modeling for the representation and definition of the activities involved in appraisal and preservation
- Chemistry and computer engineering for the study of storage media and of digital preservation technology and technological methods of authentication

InterPARES 1 Products

A body of concepts (e.g. record vs. document and data) and principles (e.g. trusted custodian), and a series of analytical instruments (e.g. a template) for studying new types of digital documents and developing new requirements and method as needed

Examples of Concepts

Reliability: the trustworthiness and accuracy of content

Authenticity: the trustworthiness of an entity as such. An authentic record is one that has not been tampered with or otherwise corrupted. Authenticity comprises identity and integrity.

Identity refers to the attributes of a record that uniquely characterize it and distinguish it from other records.

Integrity refers to the wholeness and soundness of a record. A record has integrity if it is intact and uncorrupted

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

Other InterPARES Products

- Authenticity Requirements for those who generate and keep records and for those who preserve them (e.g. metadata for identity and integrity, access privileges, etc.)
- Selection and preservation methods and procedures (e.g. models representing activities and responsibilities)
- A framework for the development of policies, strategies and standards related to the proper creation, maintenance and preservation of digital records that can be proven authentic over time

Most Important Findings

To exist as a record, an electronic entity must have:

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

Most Important Findings (cont.)

With digital records we need to separate formal elements from attributes and digital components:

A record **formal element** is a constituent part of the record's documentary form. It may be either extrinsic, like a seal, or intrinsic, like the subscription

A record **attribute** is a defining characteristic of each given record (i.e. name of author) or of a record element in it (i.e. legend on a seal)

A record **digital component** is a digital object that may contain all or part of a record, and/or the related metadata, or more than one record, and that requires specific methods for preservation.

Most Important Findings (cont.)

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 they are obsolete

Most Important Conclusions

The preservation of authentic electronic 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 trusted record-keeping system and the role of the preserver as a trusted custodian
- must incorporate records appraisal and archival description

Most Important Conclusions (cont.)

The only way of preserving an inactive electronic 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 electronic records throughout their life-cycle

Lessons Learned

- Solutions to the preservation problem are inherently dynamic due to technological change and the increasing complexity of its products

- 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, or…not?

Usefulness of Theory

Archival and Diplomatic Theory

- demonstrated when systems were designed to contain fluid data rather than records
- showed what attributes of a record's identity are implicit in the system and need to be made explicit
- supported the identification of the requirements for a presumption of authenticity

However…

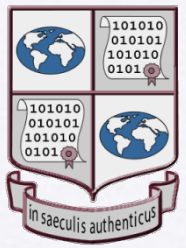
- The classic concept of record limited our capacity to understand electronic systems containing a variety of complex entities that do not correspond to it
- That which is known is not always very useful to understand that which is unknown
- Theory, which decontextualizes the record, is not useful to deal with the variety and complexity of systems: a complementary inductive approach is necessary

Therefore…

We developed a new research project that

- examines all kinds of digital entities in complex systems and then develops a concept of record;
- is concerned with the entire life-cycle of the record and develops a chain of preservation model beginning with the design of the system and capable of ensuring accuracy and reliability in addition to authenticity;
- studies the digital entities created in the course of activities that have used complex systems well before government and business, that is, e-art and e-science; and
- uses the concepts and methodology of all the fields touched by our study

InterPARES 2 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: 2002-2006

- Major funding from SSHRC, NHPRC, NSF, UBC, and UNESCO (for last 18 months)
- 21 countries in 5 continents, 100 researchers
- Public and private sectors
- Academics and professionals (80% to 20%)
- 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 Do We Study?

Dynamic entities: they depend for their content upon data extracted from a variety of systems which may have variable instantiations.

Experiential entities: objects whose essence goes beyond the bits constituting it to incorporate the behaviour of the rendering system and the effects of subjective user's interactions.

Interactive entities: objects to which each user intervention or input from another system causes a change of content and/or form.

Live, active materials, followed from creation to preservation, and obsolete inaccessible materials

Approaches taken by both creators and preservers who have concerned themselves with accuracy, reliability and authenticity

Methodological Principles

- Interdisciplinarity
- Transferability
- Open Inquiry
- Layered Knowledge Environment
- Multi-method design: surveys, case studies, modeling, prototyping, diplomatic and archival analysis, and text analysis, etc. to address domain and cross-domain research questions

Team Structure

	FOCUS 1 Artistic activities	FOCUS 2 Scientific activities	FOCUS 3 Governmental activities
DOMAIN 1 Records creation & maintenance	**Working Group 1.1**	**Working Group 1.2**	**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			

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

Work Accomplished by Focuses

- 23 case studies have been completed, represented in activity and entity models and analysed according to diplomatic and archival principles

- surveys of governments websites, of digital photographers, composers, of the practice of preservation of interactive music, of file formats and encoding languages used for non-textual materials, and the analysis of a prototype for a persistent archives based on data grids

- annotated bibliographies and literature reviews, conceptual analyses of the findings of the reviews, and bibliographic databases for the management of references

Examples of Case Studies

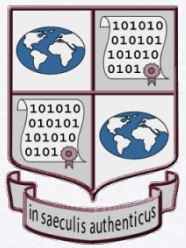
VanMAP. An extensive database of geographic data on the City of Vancouver maintained by the City's Information Technology Department. The data are supplied and updated on a regular basis by the Engineering, Planning, Social Planning, Permits and Licenses, Real Estate Services and other departments, and to a much lesser extent by external agencies such as the provincial government and crown agencies. The fundamental purpose of VanMap is to meet the needs of internal users in providing services to Vancouver's citizens and businesses.

The data, which do not exist elsewhere in similar aggregations, are constantly overwritten. There is no administrative, legal or historical accountability.

Examples of Case Studies (cont.)

The Geographic Information System of the Centre of Desert Archaeology in Tucson, Arizona. This study investigated how the data in the system are created and maintained in a way that they can be considered accurate and authentic over time. The study also investigated how the perception of data is related to the concept of memory of archaeologists, the related (human) landscapes and descendants of the areas that are studied by archaeologists.

Examples of Case Studies (cont.)

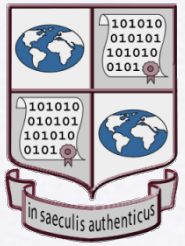


Alsace-Moselle Land Registry. A computerized land registry in Alsace-Moselle, a French, regional administrative entity. Each entry requires the signature of a judge, using a PKI combining biometric access and digital signatures. The focus of this case study was on digital signatures within a dynamic information system designed to improve the efficiency of government-citizen relations in the context of the French civil law evidence system.

Examples of Case Studies (cont.)

The New York State Department of Motor Vehicles. It offers online services to users, who conduct legal and financial transactions within the website, which generates records in a networked and online environment. The DMV's highly interactive online system features a complex set of interwoven electronic activities.

Examples of Case Studies (cont.)



Revenue Commissioners of Ireland's Revenue On-line System. It enables the generation, maintenance, access and preservation of electronic-based tax and other records in a secure and appropriate environment. The research team has examined the metadata models and standards used for information creation and exchange, and issues of accuracy, authenticity and reliability as they relate to ensuring the integrity and confidentiality of information supplied by users.

Examples of Case Studies (cont.)

An engineering firm that manufactures high-assurance, high tolerance machined piece parts for the U.S. government. It produces CAD solid model files to be used in the design and manufacturing of mechanical piece-part assemblies. The firm must maintain these records authentic over time to "enable the production of the pieces as long as the business requires them."

The basis of the study comprises the abstraction of complex information from proprietary CAD formats, expression of this information into enhanced logical forms, rendering it into archival format, sending it across a trusted network to form part of a persistent archive and returning it for verification of authenticity, reliability and usability.

Example of Surveys

The MUSTICA Initiative. A survey of digital, interactive musical works created by composers at the Institute de Recherche et Coordination Acoustique/Musique (IRCAM) and Groupe de Recherches Musicales (GRM) of the Institute National de l'Audiovisuel (INA). This research is partially funded by France's Centre National de la Recherche Scientifique (CNRS).

The study team developed a typology of interactive digital music compositions that supports discussion and analysis of the preservation needs of interactive, digital compositions by identifying the intellectual and physical components of a variety of works

Work Accomplished by Cross-Domains

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

- **Modeling**: Manage the Chain of Preservation model depicting all the activities involved in the management of electronic records throughout their lifecycle, from creation to permanent preservation, and a Business Process Model for organizations that do not have an archival program but keep their records till needed

Work Accomplished by Cross-Domains (cont.)

- **Description**: Metadata Schema Registry, which is a centralized repository of schemas that will aid to identify metadata sets, or the combinations of elements from several sets which are appropriate to serve various recordkeeping needs; identification of the relationship between metadata and archival description; collaborations with modeling and policy cross-domains

- **Policy**: the identification of barriers to preservation which currently exist in laws, regulations, policies and standards concerning copyright and intellectual rights, privacy and freedom of information, authenticity and authentication, open standards and open source, and records and archival management, and the development of a policy framework for creators and one for preservers

Domains' Work in Course

Development of

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

Production of:

guidelines for individual records creators and preservers;

preservation guidelines;

descriptive schemas for digital entities.

Findings to Date: Methodological

Surprisingly successful:

- looking at the arts and sciences for inspiration and guidance
- interaction with the creators both as co-investigators and as the interested party in the case studies and in the surveys
- continuous dissemination for feedback to the various stakeholders (e.g. scientists, arts conservators)

Expected:

The archival and diplomatic analyses of the preliminary results of the various research activities are fruitful only if one keeps in mind that these methods are retroactive by definition while we are examining live entities (not just active entities)

Findings to Date: Conceptual

- Fixed form and stable content are essential to the existence of a document, let alone a record, but these concepts apply to situations that may appear fluid and dynamic

- Some systems (e.g. the VanMap) are conceptually one record at any given moment in time, but they also disappear as such at any given moment in time

- Most entities examined in the case studies exist as potential records: they may become at any time either ‘interim records’, kept records in a recordkeeping system, or records within the system creating them

- Creators want to have records and maintain functionality

- The ‘one size fits all’ solution cannot work

Findings to Date: Some Principles

Only digital entities with a fixed documentary form and a stable content can be considered records and preserved over time

Organizations should establish criteria for determining which digital entities need to be maintained as records and what methods should be employed to fix their form and content. These criteria should be based on legal, administrative and historical accountability in addition to business requirements.

Findings to Date: Some Principles (cont.)

All the business processes that contribute to the making and/or use of the same records should be explicitly documented

The reproduction of a record made by the creator in its usual and ordinary course of conducting business should be regarded as having the force of an original record if it is used in the ordinary course of business for the purposes of such business.

Records creation procedures should ensure that digital components of records can be maintained, transmitted, preserved and reassembled over time.

Findings to Date: Principles (cont.)

Record reliability, accuracy, and authenticity should be expressly and separately provided for by means of proper procedures.

Reliable and accurate records can only be generated in a trusted record-making system.

Inclusion in a trusted recordkeeping system protects the authenticity of the records.

Preservation is to be embedded in the entire lifecycle of the record.

Findings to Date: Some Principles (cont.)

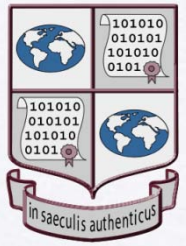
Third-party intellectual property rights and copyright issues should be identified and flagged.

The designated preserver should be invested with the role of trusted custodian.

Access issues surrounding the protection of privacy should be flagged.

Security systems should be adopted that do not hamper preservation over the long term.

InterPARES Web Site



www.interpares.org