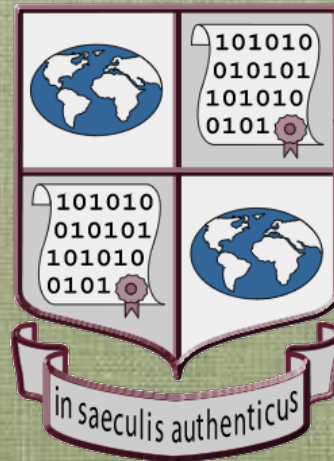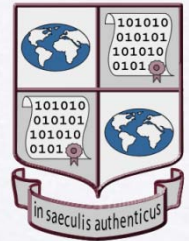# The InterPARES Project

in saeculis authenticus

Dr. Luciana Duranti
InterPARES Project Director

The Goal of InterPARES 1 and 2
1999-2007

To develop the body of theory and methods necessary to ensure that digital records produced in databases and office systems as well as 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 them and of society at large, regardless of technology obsolescence and media fragility.
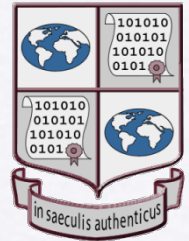
Object of Inquiry: Records

A record is 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** is information affixed to a medium in a determined form

- **information** is a message intended for communication across space or time

- **data** is the smallest meaningful piece of information

A **digital record** is 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**.
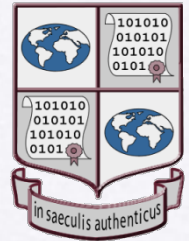
Qualities We Want to Protect

- **Reliability**: the trustworthiness of a record as a statement of fact. A reliable record is <u>complete</u> and generated according to a <u>controlled procedure</u>

- **Accuracy**: the exactness and correctness of a record content, dependent on the <u>competence</u> of the author and the <u>controls on the process by which data are recorded and transmitted</u> 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 a record <u>identity</u> and <u>integrity</u>. It differs from

Authentication: a <u>means of declaring authenticity</u> at a point in time
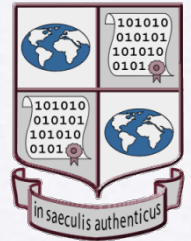
The Digital Records Challenge

- They do not exist as physical entities, but are constituted of linked digital components (the "manifested" record differs from the "stored" record)

- Their original manifestation disappears when they are saved: we cannot maintain or preserve digital records, but only the ability to re-produce or re-create them

- The facility of reproduction and manipulation makes it difficult to identify the final, official, reliable or accurate version

- Technological obsolescence makes digital records inaccessible in a very short time span

- Intellectual property and privacy rights are hard to protect

The Digital Records Challenge (cont.)

Most systems that should contain records do not, because the entities in them **lack fixed form and stable content**.

- In dynamic systems they depend for their content upon data extracted from a variety of other systems which may have variable instantiations.

- In experiential systems they are intended to incorporate the behaviour of the rendering system and the effects of subjective user's interactions.

- In interactive systems, each user intervention or input from another system causes a change of content and/or form.

The systems that do contain records, contain bad records, primarily because of **lack of identifiable contexts and relationships** among themselves and with records outside the system

InterPARES Principles
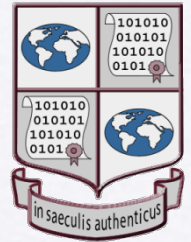
- Technology cannot determine the solution to the reliable and accurate creation of digital records or to their authentic preservation over the long term: **organizational needs** define the problem and **archival principles** must establish the correctness and adequacy of each technical solution

- 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**, till 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

InterPARES Research Activities
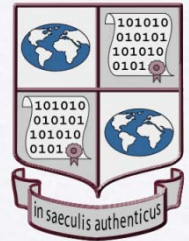
- Identification of **what constitutes a record** in each type of system and in each context, and **what record has the force of an original**, based on archival and diplomatics theory and six thousand years of record making and recordkeeping

- Definition of **what a reliable, accurate and authentic record** is in the arts, science, law and administration on the basis of analysis of literature, surveys, interviews

- Development of the requirements for the design of **a trusted record making system, a trusted recordkeeping system** and **a trusted record preservation system**, on the basis of 42 case studies, and the modeling and diplomatic analysis of the results

- Development of **methods and procedures for the creation, maintenance, appraisal, selection and disposition, and long-term preservation of digital records**, based on archival theory, law, and issues related to organizational culture

InterPARES Final Products

- A **framework of principles guiding the 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

- A **metadata registry** for the registration and analysis of metadata schemas

- Principles and criteria for adoption of **file formats, wrappers, and encoding**

- A **terminology database** including glossary, dictionary and ontologies

Intellectual Framework for Policy Development: Purpose

- To establish a flexible, consistent and solid basis for the development of policies, strategies and standards

- To provide a core of concepts capable of balancing different cultural, social and juridical perspectives on a variety of issues, such as records trustworthiness, access to information, data privacy, and intellectual property

- To demonstrate the need for a strong continuing relationship between records creators and records preservers throughout the lifecycle of the records

- To show the nature of such relationship and the ways in which it should play out
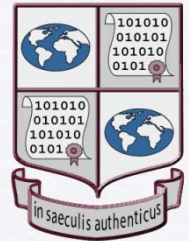
Record: Fixed Form & Stable Content

[C1] Digital entities must have stable content and a fixed documentary form to be considered records, to serve a memorial function, and to be capable of being preserved over time

[P5] Authentic copies should be made for preservation purposes only from the creator's records, that is from digital entities that have a stable content and a fixed documentary form

Stable Content and Fixed Form

- **Stable Content**: the fact that the data and the message in the record are unchanged and unchangeable, meaning that data cannot be overwritten, altered, deleted or added to

- **Fixed Form:** its binary content is stored so that the message it conveys can be rendered with the same presentation it had on the screen when first saved (different digital presentation)
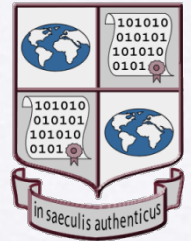
Stable Content and Fixed Form (cont.)

- But, if the same content can be presented on the screen in several different ways in a limited series of possibilities, we may have either a different view of **the same stored record** having stable content and fix form (different documentary presentations e.g. statistical data as a pie chart, a bar chart, or a table) or a **several manifested records** with stable content and fixed form derived from the same stored record

- **Bounded Variability**: when there is no stored record but content data, form data and composition data that are quite separate and only connected by a query, and changes to the form 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

Digital Components

[C2] Records creation procedures should ensure that the digital components of records can be separately maintained and reassembled over time

[P4] Records preservation procedures should ensure that digital components of records can be separately preserved and reassembled over time.

Elements, Attributes, Components

- **Formal Elements**: a constituent part of the record documentary form as shown on its face (intrinsic or extrinsic)—in MoReq 2 called intellectual components

- **Attributes:** the unique characteristics of the records that demonstrate its identity (expressed in identity and integrity metadata)

- **Digital Components**: objects that either contain one or more records or are contained in the record and require a specific preservation measure

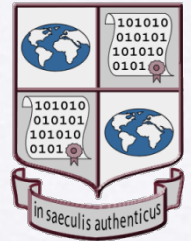# Creation, Maintenance, and Preservation Requirements

[C3] Record creation and maintenance requirements should be formulated in terms of **the purposes the records are to fulfill**, rather than in terms of the available or chosen record-making and record–keeping technologies available

[P6] Preservation requirements should be formulated in terms of **the purpose or desired outcome of preservation**, rather than in terms of the available or chosen technologies available

Records Trustworthiness

[C4] Records creation and maintenance policies, strategies and standards should address the issues of record **reliability, accuracy, and authenticity** expressly and separately

[P2] Records preservation policies, strategies and standards should address the issues of record accuracy and authenticity expressly and separately

Trusted Record-making System

[C5] A trusted record making system should be used to generate records that can be presumed reliable and accurate

No corresponding requirement for the Preserver other than as a Record Creator itself

Trusted Recordkeeping and Record Preservation Systems

[C6] A trusted recordkeeping system should be used to maintain records that can be presumed accurate and authentic

[P11] Archival appraisal should assess the authenticity of the records.

[P12] Archival description should be used as a collective authentication of the records in a fonds
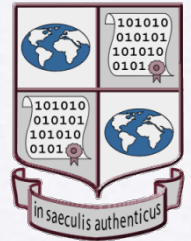
Preservation Begins at Creation

[C7] Preservation considerations should be embedded in
all activities involved in record creation and
maintenance if a creator wishes to maintain and
preserve authentic records beyond its operational
business need

[P7] Preservation considerations should be embedded in
all activities involved in each phase of the records
lifecycle if their continuing authentic existence over
the long term is to be ensured

Trusted Custodian

[C8] A **trusted custodian** should be designated as the preserver of the creator's records

[P1] A **designated record preserver** fulfills the role of trusted custodian

Designated Preserver

Trusted custodian:

- 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

Business Processes

[C9] All business processes that contribute to the creation and/or use of the same records should be explicitly documented

[P10] Archival appraisal should identify and analyze all the business processes that contribute to the creation and/or use of the same records

Intellectual Property Rights

[C10] Third-party intellectual property rights attached to the creator's records should be explicitly identified and managed in the record-making and recordkeeping systems

[P8] Third-party property rights attached to the creator's records should be explicitly identified and managed in the preservation system

Privacy Rights

[C11] Privacy rights and obligations attached to the creator's records should be explicitly identified and protected in the record-making and recordkeeping systems

[P9] Privacy rights and obligations attached to the creator's records should be explicitly identified and protected in the preservation system

Records Sharing Across Jurisdictions
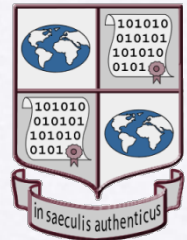
[C12] Procedures for sharing records across different jurisdictions should be established on the basis of the legal requirements under which the records are created

[P13] Procedures for providing access to records created in one jurisdiction to users in other jurisdictions should be established on the basis of the legal environment in which the records were created
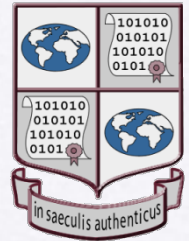
Reproduction of Records

[C13] Reproductions of a record made by the creator in its usual and ordinary course of business and for its purposes and use, as part of its recordkeeping activities, have the same effects of its first created manifestation and each is to be considered at any given time the record of the creator

[P3] Reproductions of a creator's records made for purposes of preservation by their trusted custodian are authentic copies of the creator's records

Towards InterPARES 3

- A study of the effectiveness of workshops and seminars experiences for increasing archivists' skills in digital preservation and their ability to implement these skills in their repositories has shown that very few participants were able to implement the skills once they returned to their work environments (Duff, Wendy M., Amy Marshall, Carrie Limkilde, and Marlene van Ballegooie. "Digital Preservation Education: Educating or Networking?" *The American Archivist*, 69, 1 (2006): 188-212. In the context of ERPANET.)

- Feedback on the outcomes of the two phases of InterPARES from archivists working in institutions smaller than national archives has consistently shown concern about their downward-scalability and their relevance to small and medium sized organizations.

Goal of InterPARES 3

To enable small and medium sized public and private archival organizations and programs (units within records creating organizations), which are responsible for the digital records resulting from government, business, research, art and entertainment, social and/or community activities, to preserve over the long term authentic records that satisfy the requirements of their stakeholders and society's needs for an adequate record of its past.

Expected Products

1. **Policies, strategies and procedures** for small archival organizations or programs, and **guidelines** for the records creators whose records fall under their responsibility.

2. **Action plans** for the specific case studies carried out in the course of the project.

3. **Analysis of the validity, applicability or adaptability of action plans** developed in the specific cases studied to different organizations, contexts or countries.

4. **Comparison among the action plans** developed for the preservation of records at different stages in their lifecycle (i.e. planning, creation, use, maintenance, modification, preservation)

Expected Products (cont.)

5. **Criteria to determine "most-at-risk" materials** - like a checklist of age (date created, date last accessed), physical carrier, operating system, software used, equipment required and its availability, etc.

6. **Guidelines for addressing digital preservation requirements** that apply to specific types of records, but not to other materials.

7. **Evaluation models for assessing the degree of success**, if any, of the chosen preservation action.

8. **Cost-benefit models** for various types of archival organizations or programs and for various kinds of records and/or systems.

9. **Ethical models** that identify and make explicit the consequences for individuals and society of types of preservation measures or lack thereof.
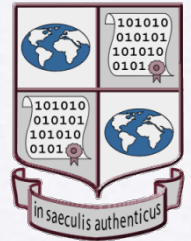
Expected Products (cont.)

10. A dedicated **web site** providing small and medium sized archival organizations world wide with good practical knowledge, internationally developed and shared, concrete examples of successful implementation, models of preservation costs, of programs evaluation, of ethical preservation behaviors, etc.

11. A refined body of **theoretical and methodological knowledge** on digital preservation, communicated in conference papers, symposia, and refereed publications.

12. **Training and education modules** for archival organizations or programs, professional associations and university programs; and awareness and education modules for non archivists, such as IT professionals, vendors, and service providers; human resources and financial managers; doctors, communities of practice, members of the general public, etc.; and a strategy for delivering them.

13. **Position papers** directed to key regulating, auditing and policy making bodies, advocating the vital need of embedding planned digital preservation in the requirements they issue for the activities they regulate, audit or control.
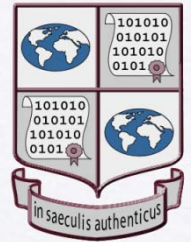
InterPARES 3 International Team

TEAMs (**T**heoretical **E**laboration into **A**rchival **M**anagement):

Canada; Holland and Belgium; Italy; Brazil; Africa; Korea; China; Ireland and England; Norway; Sweden; Singapore and Malaysia; USA; **Mexico**.

Director: Luciana Duranti

Headquarters: University of British Columbia – School of Library, Archival, and Information Studies

InterPARES Web Site

www.interpares.org