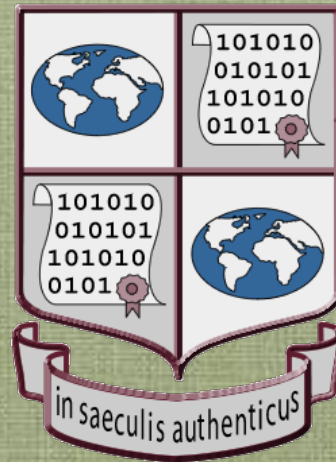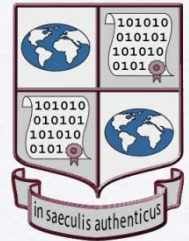# Preserving Authentic Digital Records Over the Long-term:

The InterPARES Project

Luciana Duranti and Terry Eastwood
School of Library, Archival and Information
Studies, University of British Columbia

Advantages of the Digital Medium

- **Digital materials do not fade or become yellow and brittle**

- **It is easy to alter them without leaving a trace for editing or repurposing**

- **They occupy very little storage space**

- **They can be copied an infinite number of times**

- **They can be shared over the internet**

- **They can be sent and received across the world within seconds**

Disadvantages of the Digital Medium
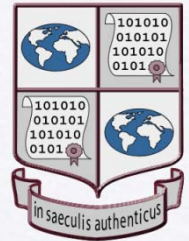
- **A computer is needed to read digital materials: The medium does not contain any given work but only bit-strings**

- **There is no longer an original**

- **Authenticity is no longer verifiable on the work itself**

- **The easiness of reproduction makes it difficult to identify the final version**

- **With interactive and dynamic systems, often we have only views of works, rather than identifiable concrete entities with defined boundaries**

…and more
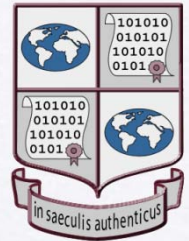
- **The internet makes intellectual property increasingly difficult to protect**

- **Viruses and technology failures make it easy to lose everything**

- **Technological obsolescence makes digital materials inaccessible very fast**

- **Works that comprise text, images, graphics, etc. are broken down and stored in different parts of the memory**

- **The information provided by the materiality of the object no longer exists**

…and bad habits make it worse
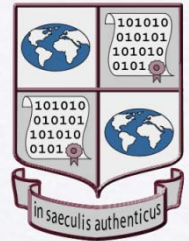
- Hybrid systems

- Creating materials in different applications and leaving them there

- Not doing regular back-up and upgrading of files

- Not keeping media in the right climatic **environments**

- Not refreshing the media

and worse…
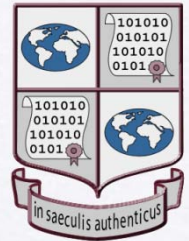
- Using proprietary or legacy systems

- Not migrating the materials to new technology

- Hoping that emulation will take care of long-term access

- Not protecting the documents from malicious or accidental tampering—trusting personal or institutional custody

- Using protection systems—encryption or digital signatures—that do not allow for preservation

InterPARES Goal
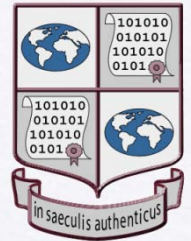
To develop the theoretical and methodological knowledge essential to the permanent preservation of authentic records generated and/or maintained electronically, and, on the basis of this knowledge, to formulate model policies, strategies and standards capable of ensuring that preservation.

InterPARES 2 Goal
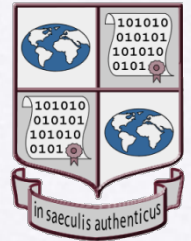
To ensure that the portion of society's recorded memory digitally produced in dynamic, experiential and interactive systems in the course of artistic, scientific and e-government activities can be created in accurate and reliable form and maintained and preserved in authentic form, both in the long and 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 (1999-2006)
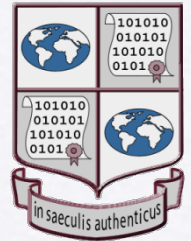
- Major funding from SSHRC, NHPRC, NSF, UBC

- 20 countries in 5 continents, 100 researchers

- Public and private sectors

- Academics and professionals

- Archival science, diplomatics; music theory, composition, performance; film theory, production, description; dance and theatre theory; hard and social sciences methodologies; jurisprudence; computer science and engineering

InterPARES Products

A body of concepts and principles and a series of analytical instruments for studying new types of digital documents and developing new requirements and methods as needed
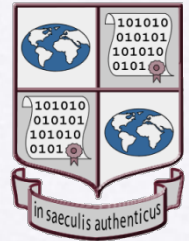
Concepts of Reliability and Authenticity
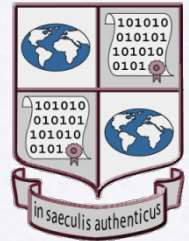
Reliability refers to the trustworthiness and accuracy of content

Authenticity refers to the trustworthiness of an entity as such. An authentic work is one that has not been tampered with or otherwise corrupted

Authenticity comprises identity and integrity. Identity refers to the attributes of a work that uniquely characterize it and distinguish it from other works. Integrity refers to the wholeness and soundness of a work. A work has integrity if it is intact and uncorrupted, that is, the message it was meant to convey is unchanged

# Concepts of Dynamic, Experiential and Interactive Digital Entities

A dynamic entity depends for its content upon data extracted from databases which may have variable instantiations. The main challenge for its reliability and authenticity is the lack of a fixed form and resides in the ability to preserve track of change

An experiential entity is an object whose essence goes beyond the bits constituting it to incorporate the behaviour of the rendering system. Its reliability and authenticity depends on the ability to preserve the environment in which the system was experienced

An interactive entity is one in which each user intervention causes a response and/or a change in its substance. Its reliability and authenticity depends on the tracking and preservation of the changes, and perhaps of the functionality of the creating system

Other InterPARES Products
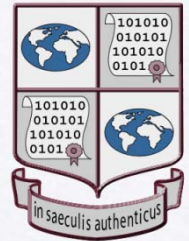
- Authenticity Requirements for those who generate and keep entities and for those who preserve them (e.g. metadata for identity and integrity, access privileges, etc.)

- Selection and preservation methods and procedures (e.g. models representing procedures and responsibilities)

- A framework for the development of policies, strategies and standards related to the proper creation, maintenance and preservation of digital entities that are reliable and accurate, and that can be proven authentic over time

Digital preservation

- The transmission of digital objects across technological boundaries (computer platforms, operating systems, applications) to overcome technological obsolescence

- A digital object possesses:

  - A physical dimension, as an inscription on a physical carrier (punch card, magnetic tape, optical disc)

  - A logical dimension, as this inscription must be recognized and processed by software

  - A conceptual dimension, as an object produced and to be understood within a specific context

Digital Preservation (cont.)
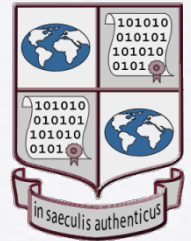
- In order to preserve a digital object, we must be able to identify and retrieve all of its logical and physical parts and **reconstitute** the conceptual object

- That is, to access any digital object, **stored** bit sequences must be **interpreted** as logical objects and **presented** as conceptual objects

- In the paper-and-ink world, the basis of preservation is the caring for the integrity of the physical carrier itself, but…

…in the digital world

- Digital preservation is the process of **preserving the ability to reproduce a digital object,** and this process is complete only when the object is successfully output

- However, the reproduced object does not have to be identical to its first instantiation, provided that its identity and integrity are ensured (bits and bit streams may change)

Migration
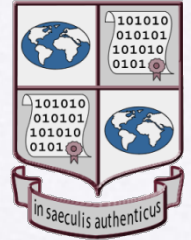
- The most prevalent preservation strategy is currently that of **migration** of

  - **media**, using state-of-the-art technology for **storage** and **access**

  - **data formats**, using state-of-the-art technology for **output**

- Migrating data formats involves changes to the object's bitstream

Migration
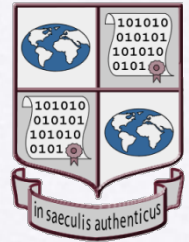
- The problem becomes, "Which changes are permissible and/or beneficial?"

- Given that a digital object is something that can only be **re-constructed** by using software, it is necessary to have an **explicit model or standard** that provides **criteria for assessing the authenticity** of the re-constructed object

The Purpose of the Model
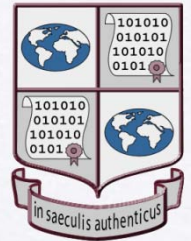
Ensuring that the reproduction is carried out according to a prescribed procedure and is the responsibility of a trusted custodian having no interest in the content of the objects, and having the authority and the capacity of documenting such procedure thoroughly
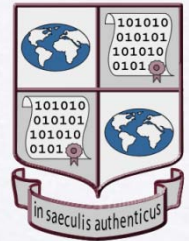
Trusted Custodian

The procedures and system(s) used for maintaining the records in the long term and reproducing them must 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;

- the content of the record remains unchanged after reproduction

Trusted Custodian

The activity of reproduction must be documented, and this documentation should include:

- the date of the objects' reproduction and the name of the responsible person;

- the relationship between the objects acquired from the author and the copies produced by the custodian;

- the impact of the reproduction process on their form, content, accessibility and use; and

- in those cases where a copy of an object is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the custodian, and this documentation is readily accessible to the user

Key points concerning preservation
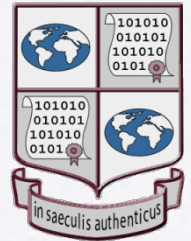
- Technology alone cannot solve the long-term preservation of digital objects

- Authors' and institutions' 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

Research In Progress

- An understanding of the concepts of accuracy, reliability and authenticity in the visual and performing arts and in the sciences and their implementation

- A guide to encoding formats that can be preserved over the long-term

- Guidelines for individuals who are not part of government or large organizations to help them to create works than can be preserved

- Registers of metadata schemata for different types of work

- Methods for selecting works for permanent preservation on the basis of their legal, administrative, social or cultural value

Research In Progress
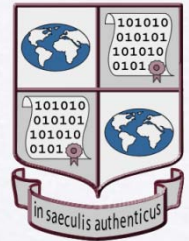
- Methods and strategies for keeping dynamic, experiential and interactive works destined for permanent preservation in authentic form over the long term

- Processes for analyzing and criteria for evaluating advanced technologies for the implementation of the above methods in ways that respect cultural diversity and pluralism

- A framework for policies and strategies dealing with intellectual property and copyright issues in the dynamic, experiential and interactive digital environment

InterPARES Web Site

www.interpares.org