



The Long-term Preservation of Accurate and Authentic Digital Data: The InterPARES Project

Luciana Duranti

InterPARES Project &
School of Library, Archival and Information Studies,
University of British Columbia



InterPARES Project
Dr. Luciana Duranti
Project Director

Modern Working Environment

- Hybrid documentary systems
- Digital environments that support the manipulation and repurposing of data
- Obsolescence of systems and media
- Proprietary and idiosyncratic nature of applications



Consequences

- Data are less
 - reliable (manipulability)
 - retrievable (incongruence of classifications/taxonomies)
 - accessible, readable or intelligible (incompatibility and obsolescence)
- It is difficult to prove their accuracy and authenticity while current as we no longer have originals
- It is difficult to maintain accountability
- It is difficult to provide for their long-term authentic preservation



How to deal with this situation

- Developing a policy, a strategy and a procedure that address separately scientific data and the other digital objects
- Focusing any such policy and strategy on the continuing accuracy and authenticity of records
- Recognizing that preservation of accurate and authentic digital data is a continuous process that begins at the moment of creation and whose purpose is to transmit trustworthy information across time and space



Identifiable Characteristics of Preservable Digital Data

- Fixed form (i.e. its binary content is stored so that it remains complete and unaltered, and its content can be rendered with the same form it had when first set aside)
- Unchangeable content
- Explicit linkages to other data within or outside the digital system through metadata and a unique identifier
- Identifiable context
- Identifiable author and creator
- Participation in or support of an activity either procedurally or as part of a process



Accuracy

The trustworthiness of data as facts.

The degree to which data are precise, correct, truthful, free of error or distortion



Authenticity

Refers to the fact that data is what it purports to be and has not been tampered with or otherwise corrupted.

Authenticity is the trustworthiness of the data as a recorded entity. To verify it, one must verify the identity and integrity of the record made of it.



Authentication

A declaration of authenticity, resulting either by the insertion or the addition of an element or a statement to a record, and the rules governing it are established by legislation.

A means of proving that a record is what it purports to be at a given moment in time.



Threats to accuracy and authenticity

Accuracy and authenticity are most at risk when data are transmitted **across space** (that is, when sent between persons, systems, or applications) **or time** (that is, either when they are stored offline, or when the hardware or software used to process, communicate, or maintain them is upgraded or replaced)



Conceptual Framework for Authenticity

- In archival theory and jurisprudence, recorded data that are relied upon by their creator in the usual and ordinary course of business are presumed accurate and authentic
- In electronic systems, this presumption must be supported by **evidence** that a data is what it purports to be and has not been modified or corrupted in essential respects.
- To assess the authenticity of recorded data, the preserver must be able to **establish its identity and demonstrate its integrity**



Identity of Recorded Data

- It refers to the attributes that uniquely characterize it and distinguish it from other entities. These attributes include: the names of the person concurring to its formation (e.g., author, writer, creator, originator); its date(s) of creation and transmission; an indication of the matter or activity in which it participates; classification code or other unique identifier; as well as an indication of related data.
- These attributes should be explicitly expressed in metadata related to the recorded data, or in its technological environment (e.g. audit trail)



Integrity of Recorded Data

- Its wholeness and soundness. Data exhibiting integrity are intact and uncorrupted
- They are intact and uncorrupted if the content that they are meant to communicate is unaltered
- Recorded data physical integrity, such as the proper number of bit strings, may be compromised, provided that the content and identifying attributes remain the same
- Integrity may be demonstrated by metadata related to the actions carried out on recorded data, or its technological environment



How to preserve the accuracy and authenticity of digital data?

- Maintaining them in a **trusted record keeping system**
- Understanding that it is **not** possible to keep digital data as stored physical objects: it is only possible to maintain the ability to reproduce them
- Ensuring that the reproduction process is the responsibility of a **trusted preserver** having the authority and the capacity of documenting it thoroughly



Trusted Recordkeeping System

A trusted record-keeping system comprises the whole of the rules that control the creation, maintenance, and use of the records of the creator and that provide a circumstantial probability of the authenticity of the records within the system.



Trusted Preserver

To be considered a trusted custodian, the person responsible for keeping the data must demonstrate that he/she has no reason to alter them or allow others to alter them, and is capable of implementing all of the requirements for a trusted preservation system





InterPARES: **INT**ernational **R**esearch on **P**ermanent **A**uthentic **R**ecords in **E**lectronic **S**ystems

An International Collaborative Research Initiative



InterPARES Project

Dr. Luciana Duranti
Project Director

Research Goal

To develop the theoretical and methodological knowledge essential to ensuring 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.



Methodology

- Theory and methods of diplomatics and archival science the definition of concepts and development of requirements and methods
- Literature reviews and surveys of types of data creators
- Grounded theory for case studies
- Comparative analysis for the study of appraisal and preservation reports from archival institution
- IDEF0 modeling for the representation and definition of the activities involved in data creation, appraisal and preservation
- Computer engineering for the study of storage media and of digital preservation technology and technological methods of authentication
- Legal analysis for the study of certification methods



InterPARES Products to Date

A body of concepts and principles and a series of analytical instruments for studying new types of digital entities and developing new requirements and method as needed



InterPARES Products (cont.)

- Authenticity Requirements for those who generate and keep data 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



Research In Progress Relevant to You

- An understanding of the concepts of accuracy, reliability and authenticity in the context of a variety of sciences and their implementation
- A guide to encoding formats that can be preserved
- Guidelines for individuals who are not part of government or large organizations to help them to create data that can be preserved
- Registers of metadata schemata for different types of data
- Methods for selecting scientific data for permanent preservation



Research In Progress Relevant to You

- Methods and strategies for keeping dynamic, experiential and interactive entities destined to permanent preservation in accurate and 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



Examples of Case Studies

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



InterPARES Project

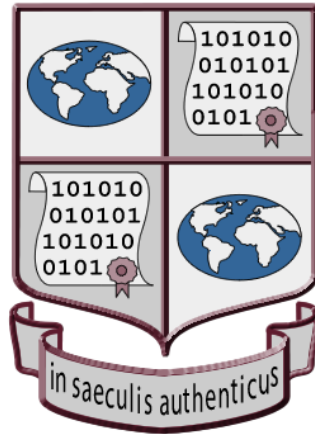
Dr. Luciana Duranti
Project Director

Examples of Case Studies

The electronic engineering and manufacturing records which have been created in computer-assisted engineering, computer-assisted design and industrial automation systems. The focus of this case study is on examining the ability of complex engineering records to stand for the solid objects modeled in the records, and the ability of the manufacturing records to represent the processes required to produce such solid objects.



InterPARES Web Site



www.interpares.org



InterPARES Project

Dr. Luciana Duranti
Project Director