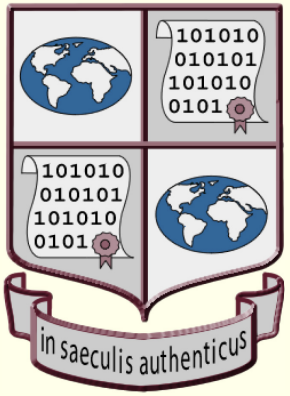
InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems



It is All About Trust

The Reliability, Accuracy and Authenticity of Digital Records

The Digital Records Challenge

- Digital records **do not exist as physical entities**, but are constituted of linked digital components presented in a documentary form (what we see is different from what is stored)
- The **original**, that is, the first complete and effective record, **disappears when closed**: we **cannot maintain or preserve digital records**, either manifested or stored, 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 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

Archival Tradition

Sir Hilary Jenkinson, *Manual of Archival Administration*. London, 1922.

- The archivist's **primary duty is to the records**
- The archivist's **secondary duty is to the user**

By serving the records we serve their users

Serving the Records

It means to maintain intact their characteristics:

1. **Naturalness** (by preserving the sedimentation)
2. **Interrelatedness** (by archival description)
3. **Impartiality** (by planned selection)
4. **Authenticity** (by maintaining a chain of unbroken legitimate custody).

With digital records, these functions are still necessary, but no longer sufficient.

What Else Is Needed

The traditional **concept of preservation** must include the processes necessary to transmit the record through time, including conversion and migration

The **unbroken chain of preservation** must begin at creation and continues from the record-making system to the recordkeeping system and the record preservation system

The new emphasis on accountability allows the archives to fulfill these needs by **presenting itself as the trusted custodian**

Archivist as Trusted Custodian

The trusted custodian is a person who

- 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;
- but, mostly...

The Archivist's New Functions

1. Positions him/herself at the **beginning of the record life-cycle**, taking the role of "designated" trusted custodian
2. Assesses the **authenticity of the records** and **monitors it** throughout their existence
3. Identifies the records to be preserved at the moment of their creation and **monitors their transformation through time**

The Archivist's New Functions (cont.)

4. Determines the **feasibility of preservation** on the basis of the archives technological capacity
5. Determines a **preservation strategy** independently of technological trends (tries to influence the industry through the adoption of standards, but not viceversa) and maintaining the focus on interoperability
6. Controls the **accuracy of the records** after each conversion or migration
7. Develops **procedures** that address issues of **intellectual rights and privacy**

The Archivist's New Functions (cont.)

8. Recognizes to **archival description a primary authentication function**

- The authentication function of archival description is a **collective** attestation of the authenticity of the records in an archive group and of all their interrelationships as made explicit by 1) their administrative, custodial and technological history, 2) the illustration of their scope and content, and 3) the hierarchical representation of the records aggregates
- The unique function of archival description is to provide an **historical view of the records and of their becoming** while presenting them as a whole in which the individuality of each member is subject to the bond of a common provenance and destination

The Archivist's New Functions (cont.)

9. Is constantly **involved in research and development projects** similar to those carried out by the industry, addressing questions like the following:

   - What entity constitutes the record in each dynamic or interactive system
   - What manifestation of such entity can be regarded as the record
   - How to keep such entities authentic through time
   - How to enable users to verify authenticity over time

What Knowledge Shall the Archivist Have to Use?

- **Archival Diplomatics**
  The integration of archival and diplomatic theory about the genesis, inner constitution, and transmission of records; and about their relationship with the facts represented in them, with other records produced in the course of the same function and activities, and with their creators.
- **Archival Methodology**
- **History of Record Making and Recordkeeping**
- **Administrative and Evidence Law**
- **Information Science**

The Uses of Archival Diplomatics

- **Retrospective**: to understand the nature and attributes of existing records and to assess their trustworthiness
- **Prospective**: to design documentary forms and procedures and to develop trusted record-making, recordkeeping and record preservation systems

The InterPARES research project has used archival diplomatics in both ways

InterPARES Project Goal

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 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.

Outcomes of the Use of Archival Diplomatics in InterPARES

- A more nuanced concept of record, although still consistent with and within the boundaries of the traditional definition
- A more nuanced concept of trustworthiness, encompassing reliability, accuracy, authenticity and authentication
- A complex methodology of analysis that keeps into account all the context in which records are created

Digital Record Characteristics (identified in InterPARES 1)

- **Medium**: necessary part of the technological context, not of the record
- **Stable Content and Fixed Form**: the message is unchangeable and so is the documentary presentation
- **Five Necessary Persons**: author, writer, originator, addressee, and creator
- **Act**: an action in which the records participates or which the record supports
- **Archival Bond**: explicit linkages to other records inside or outside the system
- **Five Necessary Contexts**: juridical-administrative, provenancial, procedural, documentary, technological

Digital Record Characteristics (cont.)

- **Formal Elements**: a constituent part of the record documentary form as shown on its face (intrinsic or extrinsic)
- **Metadata**: the attributes of the records that demonstrate its identity and integrity (authenticity)
- **Digital Components**: entities that either contain one or more records or are contained in the record and require a specific preservation measure

Stored and Manifested Records (identified in InterPARES 2)

- **Stored record**: the digital component(s) used in reproducing one or more than one record, which include the data to be processed in order to re-produce the manifested record (content data and form data) and the rules for processing the data, including those enabling variations (composition data)
- **Manifested record**: the visualization or materialization of the record in a form suitable for presentation to a person or system. Sometimes, it does not have a corresponding stored record, but is re-created from fixed content data when a user's action associates them with specific form data and composition data (e.g. a record produced from a relational database)

Types of Digital Entities

Static: They do not provide possibilities for changing their manifest content or form beyond opening, closing and navigating: e-mail, reports, sound recordings, motion video, snapshots of web pages. If the other conditions (medium, archival bond, 5 persons, act, 5 contexts) are satisfied, they are records.

Interactive: They present variable content, form, or both, and the rules governing the context and form of presentation may be either fixed or variable. They may or may not be records.

Interactive Entities

- **Not-dynamic**: the rules governing the presentation of content do not vary, the content presented each time is selected from a fixed store of data, and the form of its presentation is fixed. Ex. Interactive web pages, online catalogs, records enabling performances—if the other conditions are satisfied, **they are records**

- **Dynamic**: the rules governing the presentation of content and form vary—**they are potential records till such time when the rules are fixed**

Interactive Record

Fixed Form (in a not-dynamic entity):

- if its binary content is stored so that the message it conveys can be rendered with the same documentary presentation it had on the screen when first saved (different digital presentation: Word to .pdf)
- if the same content can be presented on the screen in several different ways in a limited series of possibilities: we have a different documentary presentation of the same stored record having stable content and fixed form (e.g. statistical data viewed as a pie chart, a bar chart, or a table)

Interactive Record (cont.)

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

Records Functions

- ***Ad substantiam*** and ***ad probationem*** (dispositive and probative=legal records)
- **Supporting**: generated to be used in the course of multiple activities as a source of information (e.g., GIS)
- **Narrative**: generated as an instrument of communication but not required by the juridical system (e.g., most e-mails, reports, web sites)

Records Functions

- **Instructive**: delineate the form in which external data are to be presented (e.g., scores, scripts, regulations, manuals of procedure, instructions for filling out forms)
- **Enabling**: enable performance of artworks (software patches), execution of business transactions (interacting business applications), conduct of experiments (a workflow generated and used to carry out an experiment of which it is instrument, byproduct and residue), analysis of observational data (interpreting software), etc.

Trustworthiness

Reliability

The trustworthiness of a record as a statement of fact,

based on:

- the competence of its author
- the controls on its creation

Accuracy

The correctness and precision of a record's content

based on:

- the competence of its author
- the controls on content recording and transmission

Authenticity

The trustworthiness of a record that is what it purports to be, not tampered with and uncorrupted

based on:

- identity
- integrity

Authenticity: Identity

The attributes of a record that characterize it as unique, and that distinguish it from other records.

Identity metadata:

- names of the 5 persons concurring in its creation
- date(s) and time(s) of issuing, creation and transmission
- the matter or action in which it participates
- the expression of its archival bond
- documentary form and digital presentation
- the indication of any attachment(s)
- digital signature
- name of the person responsible for the record

Authenticity: Integrity

A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered.

Integrity metadata:

- name(s) of handling persons over time
- name of person responsible for keeping the record
- indication of annotations
- indication of technical changes
- indication of presence or removal of digital signature
- time of planned removal from the system
- time of transfer to a custodian
- time of planned deletion
- existence and location of duplicates outside the system

Authentication

A means of declaring the authenticity of a record at one particular moment in time -- possibly without regard to other evidence of identity and integrity.

Example: the **digital signature**. Functionally equivalent to medieval seals (not signatures):

- verifies origin (identity)
- certifies intactness (integrity)
- makes record indisputable and incontestable (non-repudiation)

Trusted Systems

Rules, and tools and methods to implement rules, for

Making reliable and accurate records

- record-identity metadata schemes
- business and documentary procedures integrated in a workflow structure linked to classification schemes and filing plans
- specifications of record forms
- record-making access privileges

Maintaining and keeping authentic records

- record-integrity metadata schemes
- classification schemes and filing plans
- linked retention schedule
- registration system
- retrieval system
- record-keeping access privileges

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

Reasons for the InterPARES Success

- **Multidisciplinary** and **multicultural**

- It does not break with the past, thereby **taking away the fear of the unknown**: it is based on the hypothesis that there is no record, record element or record making process we see today that we have not seen in the past six thousand years (e.g. blogs, GIS)

- It puts forward the figure of the records and information manager or archivist as the **trusted custodian**, defines its role as the neutral third party, and establishes its body of knowledge and qualifications consistently with every country's historical tradition (this is also consistent with oral traditions, entrusted to remembrancers)

Why the Success? (cont.)

- It makes clear that, without **immediate and continuative action**, there will be no data, records or information we and our descendants will be able to rely upon as sources

- InterPARES ultimate purpose is to ensure that digital records and digital communication can be trusted, and **trust** is the most rare, yet, the most shared, understood and valued commodity in every civilization, and one that cannot be bought, but needs to be built, protected, nurtured, and served

The Power of Archives

It derives from this need for trust. To fulfill it, each archives will have to establish a **policy** for the institution, **strategies** for implementing it, **plans of action** for specific aggregations or types of records, and **detailed procedures**, and to **update** all of the above continuously according to changes in available technologies, records produced, and resource availability

This is what we are beginning to do in **InterPARES 3 (2007-2012)** in an effort **to place the archives at the center of society** as **an instrument of accountability** and **a point of reference** for any institution, organization, community or person who needs a)**guidance** in the creation, maintenance and preservation of its records, b)**a trusted third party** to take care of the digital evidence of its activities, or c)**an expert witness** attesting the authenticity of digital records presented as evidence in legal proceedings.

InterPARES Findings and Products

All findings and products will be made available this spring on the InterPARES Web site:

www.interpares.org