InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

The Management of Electronic Records for Good Governance

The Power of Archives

InterPARES Project
Dr. Luciana Duranti
Project Director

Archival Legislation

- Archival laws usually say **what** archives do, not **why**, although implicitly regard them as mechanisms for memory and for identity, both collective and personal. This has marginalized archives.
- Mission statements are beginning the fill the gap in the law:
  - "The National Archives and Records Administration serves American democracy by safeguarding and preserving the records of our Government, ensuring that the people can discover, use, and learn from this documentary heritage. We ensure continuing access to the essential documentation of the rights of American citizens and the actions of their government. We support democracy, promote civic education, and facilitate historical understanding of our national experience."
- New emphasis **on protection of accountability**, on **archives as trusted custodian** and warrantor of democracy, due to the use of digital technologies and the challenges presented by digital records.

Digital Records Characteristics

- **Medium**: physical carrier, part of the technological context, but the record must be affixed to it.
- **Stable Content and Fixed Form**
- **Archival Bond**: explicit linkages to other records inside or outside the system
- **Five Necessary Persons**: author, writer, originator, addressee, and creator
- **Act**: an action in which the records participates or which the record supports
- **Five Necessary Contexts**: juridical-administrative, provenancial, procedural, documentary, technological

The Digital Records Challenge

- 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

but mostly

- They do not exist as physical entities, but are constituted of **linked digital components** (the "manifested" record differs from the "stored" record, if there is one)—see slide 5
- 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, or to protect their capacity to instruct or enable the making of records –see slides 6 and 7

Manifested and Stored Record

- **Stored record**: the digital component(s) used in re-producing one or more than one record, which include the data to be processed in order to manifest the 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 Records

- **Legal**: **dispositive**, e.g., contracts; **probative**, e.g., a land registry)
- **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)

Types of Digital Records (cont.)

- **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.

The Digital Records Challenge (cont.)

The systems that contain records, contain bad records, primarily because of **lack of identifiable contexts and relationships** among themselves and with records outside the system (records in applications, e.g. e-mail)

Most systems that should contain records do not, because the entities in them **lack fixed form and stable content**—see slides 9 and 10

- In **dynamic interactive** systems—see slides 11-12—they depend for their content upon data extracted from a variety of other systems which may have variable instantiations (VanMap).
- In **non-dynamic interactive** systems, each user intervention or input from another system causes a change of content and/or form (Alsace-Moselle).

These digital entities, regularly produced by government agencies, only exist in the hands of the government. They are either trusted implicitly, when they should not be, or not trusted, when they are trustworthy (VPD)

Fixed Form and Stable Content

Fixed Form:

- 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)

Fixed Form and Stable Content (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

Digital Entities Behaviour

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—they are records if the other conditions are satisfied.

Interactive: They present variable content, form, or both, and the rules governing the content and form of presentation may be either fixed or variable—they may be records

Interactive Entities Behaviour

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

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 protecting their identity and integrity through 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 **trustworthiness**—see slide 18-22—**of the records** and **monitors it** throughout their existence

Records 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 to be what it purports to be, untampered 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, transmission**
- **the matter or action in which it participates**
- **the expression of its archival bond**
- **documentary form**
- **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)

The analogy is not perfect, because the medieval seal was associated exclusively with a person, while the digital signature is associated with a given person and a specific record, and because the former is an expression of authority, while the latter is only a mathematical expression

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

The Archivist's New Functions (cont.)

3. Identifies the records to be preserved at the moment of their creation and **monitors their transformation through time**—appraisal
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 of a *fonds* 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 instantiation of such entity can be regarded as the record (manifested or stored entity)
   - How to keep such entities accurate and authentic through time
   - How to enable users to verify such authenticity over time

The Old Way

Academics conduct research which very few read and even less try to implement, usually unsuccessfully (as demonstrated by research)

Archival associations establish committees who **issue guidelines**, usually expressing the minimum common denominator shared by the experiences of the members, rather then research findings

ISO issues standards under the pressure of groups who need basic guidance and either develop their own (see OAIS) or are ready to adopt the ways of the most involved parties (see RM standard)

Legislators issue laws too often based on the expertise of IT professionals and without serious consultation with archivists (see Alsace-Moselle, or the European Directive on digital signatures)

Governments make technological choices without consulting with archivists (VanMap)

Archives have to **respect** often unreasonable laws, **implement** far too generic standards, and **preserve** unidentifiable and non preservable material

Why It Does Not Work

Technology changes very rapidly while national and international consensus of any kind is very slow

General standards and laws **need much adaptation** to specific contexts to be implemented

Research results must be translated in concrete terms to be understood by professionals

Research has demonstrated that **solutions to digital records preservation are dynamic and specific**

The financial, technological, and knowledge **resources of archives are very different**

A Better Way

Each **archives becomes a locus of research** by establishing a partnership with academics involved in international research, professionals involved in standards development, experts in law and information technology and, most importantly, with the creators of the records that fall under their jurisdiction.

Each **archival association promotes an environment supportive** of the archives goal by demonstrating to regulatory and auditing bodies, and policy makers that they ought to embed digital records preservation requirements (not rules) in any activity that they regulate, audit or control.

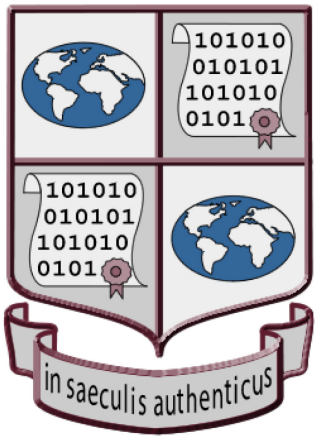
This produces 1) the generation of **new knowledge**, 2) the achievement of **action-oriented outcomes**, 3) the **education** of all participants, 4) results that are **relevant to the local setting**, 5) a **sound** and appropriate research and development **methodology,** and 6) and the **empowerment of the archives.**

The Power of Archives

Each archives will then be able 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 guidance in the creation, maintenance and preservation of its records, a neutral third party to take care of the digital evidence of its activities, or an expert witness who attests the authenticity of digital records presented as evidence in legal proceedings.

InterPARES Web Site



www.interpares.org