InterPARES 3 Project

International Research on Permanent Authentic Records in Electronic Systems

TEAM Canada

The Long-term Preservation of Authentic Digital Records: The Findings of the InterPARES 3 Project

Luciana Duranti

InterPARES Project Director

11 May 2012

Management of Cultural Heritage in a Digital World

& InterPARES 3 – An International Symposium

Istanbul, Turkey

Goal of InterPARES 3 (2007-2012)

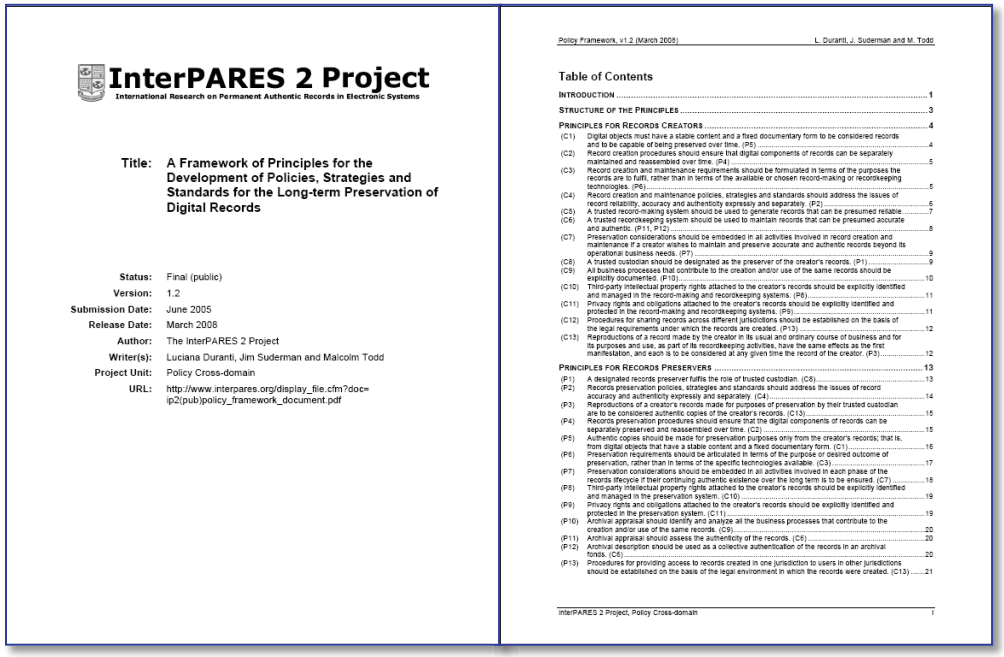
To **enable** public and private **archival organizations and programs** with limited resources **to preserve** over the long term **authentic records** that satisfy the requirements of their stakeholders and society's needs for an adequate record of its past.

It did so by building on the products of the first two phases of InterPARES (1998-2006)

Key IP 1 & 2 Products

Policy Framework

A framework of principles guiding the development of policies for records creating and preserving organizations



InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Title: A Framework of Principles for the Development of Policies, Strategies and Standards for the Long-term Preservation of Digital Records

Status: Final (public)
Version: 1.2
Submission Date: June 2005
Release Date: March 2008
Author: The InterPARES 2 Project
Writer(s): Luciana Duranti, Jim Suderman and Malcolm Todd
Project Unit: Policy Cross-domain
URL: http://www.interpares.org/display_file.cfm?doc=ip2(pub)policy_framework_document.pdf

Policy Framework, v1.2 (March 2008) L. Duranti, J. Suderman and M. Todd

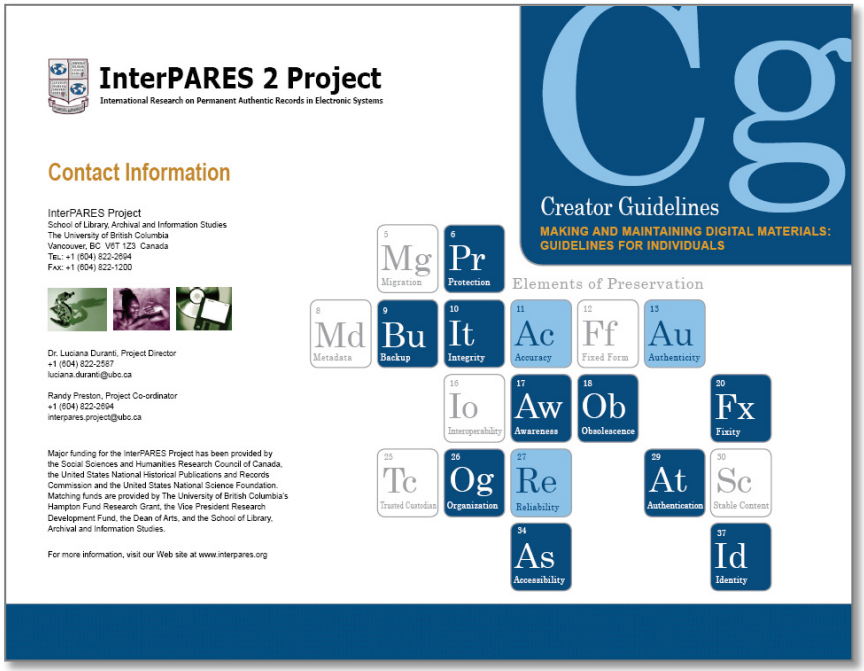
Table of Contents

InterPARES 2 Project, Policy Cross-domain i

IP 1 & 2 Products

Creator Guidelines

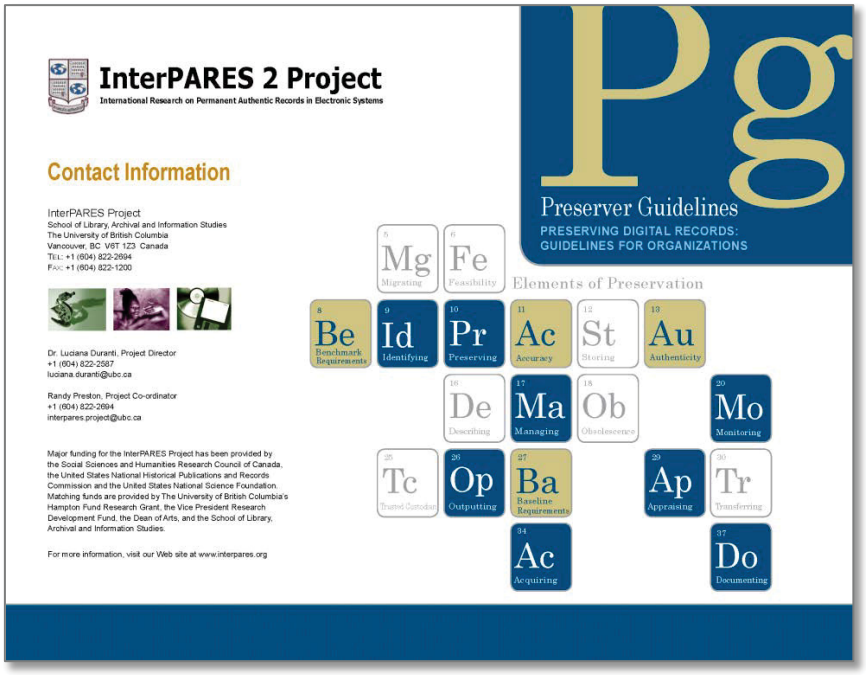
Recommendations for making and maintaining digital materials for individuals and small communities of practice



IP 1 & 2 Products

Preserver Guidelines

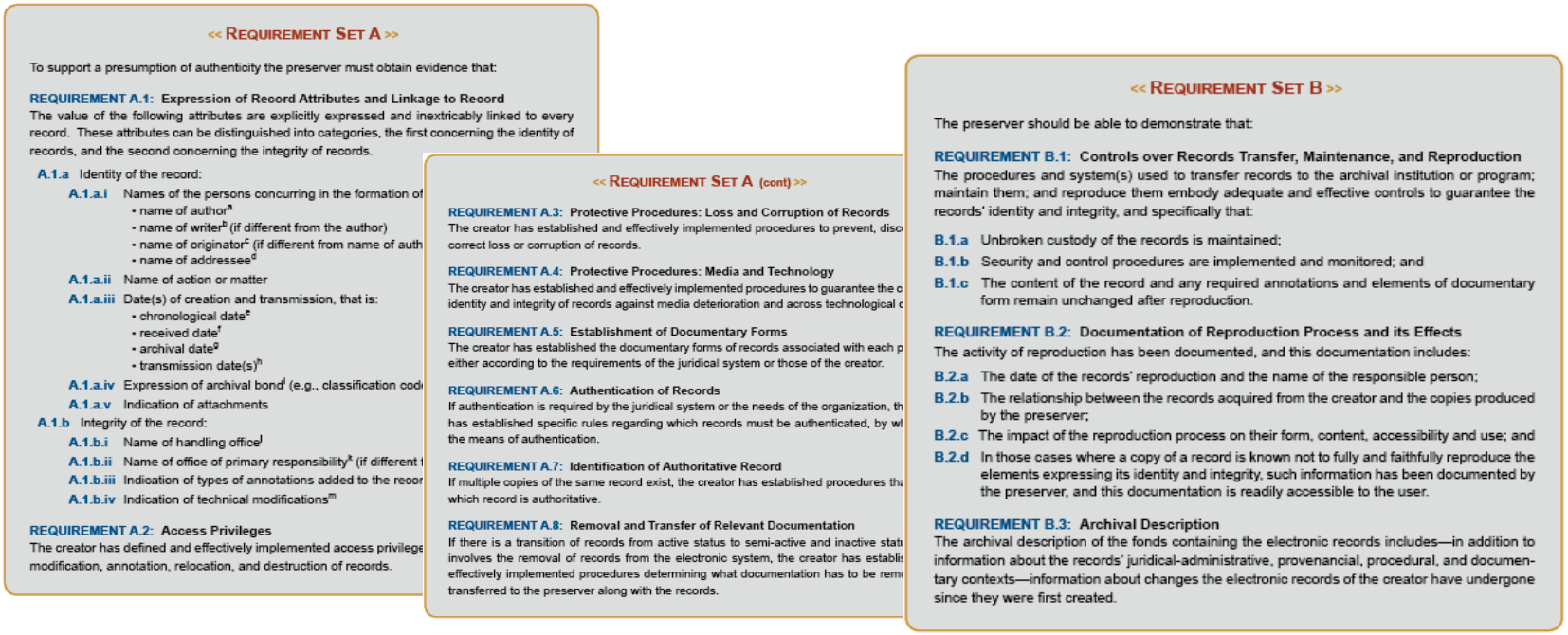
Recommendations for digital preservation for archival institutions



IP 1 & 2 Products

Benchmark and Baseline Requirements

Authenticity requirements for assessing and maintaining the authenticity of digital records



<< REQUIREMENT SET A >>

To support a presumption of authenticity the preserver must obtain evidence that:

REQUIREMENT A.1: Expression of Record Attributes and Linkage to Record
The value of the following attributes are explicitly expressed and inextricably linked to every record. These attributes can be distinguished into categories, the first concerning the identity of records, and the second concerning the integrity of records.

A.1.a Identity of the record:
- A.1.a.i Names of the persons concurring in the formation of
  - name of author[a]
  - name of writer[b] (if different from the author)
  - name of originator[c] (if different from name of auth
  - name of addressee[d]
- A.1.a.ii Name of action or matter
- A.1.a.iii Date(s) of creation and transmission, that is:
  - chronological date[e]
  - received date[f]
  - archival date[g]
  - transmission date(s)[h]
- A.1.a.iv Expression of archival bond[i] (e.g., classification cod
- A.1.a.v Indication of attachments

A.1.b Integrity of the record:
- A.1.b.i Name of handling office[j]
- A.1.b.ii Name of office of primary responsibility[k] (if different
- A.1.b.iii Indication of types of annotations added to the recor
- A.1.b.iv Indication of technical modifications[m]

REQUIREMENT A.2: Access Privileges
The creator has defined and effectively implemented access privilege
modification, annotation, relocation, and destruction of records.

<< REQUIREMENT SET A (cont) >>

REQUIREMENT A.3: Protective Procedures: Loss and Corruption of Records
The creator has established and effectively implemented procedures to prevent, disc
correct loss or corruption of records.

REQUIREMENT A.4: Protective Procedures: Media and Technology
The creator has established and effectively implemented procedures to guarantee the c
identity and integrity of records against media deterioration and across technological c

REQUIREMENT A.5: Establishment of Documentary Forms
The creator has established the documentary forms of records associated with each p
either according to the requirements of the juridical system or those of the creator.

REQUIREMENT A.6: Authentication of Records
If authentication is required by the juridical system or the needs of the organization, th
has established specific rules regarding which records must be authenticated, by wh
the means of authentication.

REQUIREMENT A.7: Identification of Authoritative Record
If multiple copies of the same record exist, the creator has established procedures tha
which record is authoritative.

REQUIREMENT A.8: Removal and Transfer of Relevant Documentation
If there is a transition of records from active status to semi-active and inactive statu
involves the removal of records from the electronic system, the creator has establis
effectively implemented procedures determining what documentation has to be rem
transferred to the preserver along with the records.

<< REQUIREMENT SET B >>

The preserver should be able to demonstrate that:

REQUIREMENT B.1: Controls over Records Transfer, Maintenance, and Reproduction
The procedures and system(s) used to transfer records to the archival institution or program; maintain them; and reproduce them embody adequate and effective controls to guarantee the records' identity and integrity, and specifically that:

- B.1.a Unbroken custody of the records is maintained;
- B.1.b Security and control procedures are implemented and monitored; and
- B.1.c The content of the record and any required annotations and elements of documentary form remain unchanged after reproduction.

REQUIREMENT B.2: Documentation of Reproduction Process and its Effects
The activity of reproduction has been documented, and this documentation includes:

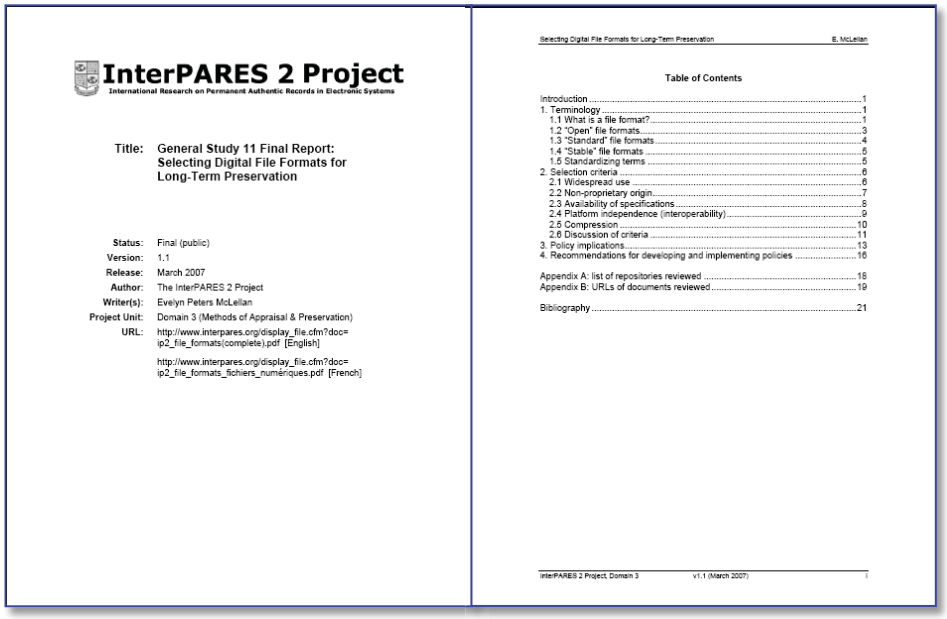
- B.2.a The date of the records' reproduction and the name of the responsible person;
- B.2.b The relationship between the records acquired from the creator and the copies produced by the preserver;
- B.2.c The impact of the reproduction process on their form, content, accessibility and use; and
- B.2.d In those cases where a copy of a record is known not to fully and faithfully reproduce the elements expressing its identity and integrity, such information has been documented by the preserver, and this documentation is readily accessible to the user.

REQUIREMENT B.3: Archival Description
The archival description of the fonds containing the electronic records includes—in addition to information about the records' juridical-administrative, provenancial, procedural, and documentary contexts—information about changes the electronic records of the creator have undergone since they were first created.

IP 1 & 2 Products

File Format Selection Guidelines

Principles and criteria for adoption of file formats, wrappers and encoding schemes



InterPARES 2 Project
International Research on Permanent Authentic Records in Electronic Systems

Title: General Study 11 Final Report: Selecting Digital File Formats for Long-Term Preservation

Status: Final (public)
Version: 1.1
Release: March 2007
Author: The InterPARES 2 Project
Writer(s): Evelyn Peters McLellan
Project Unit: Domain 3 (Methods of Appraisal & Preservation)
URL: http://www.interpares.org/display_file.cfm?doc=ip2_file_formats(complete).pdf [English]
http://www.interpares.org/display_file.cfm?doc=ip2_file_formats_fichiers_numériques.pdf [French]

Selecting Digital File Formats for Long-Term Preservation E. McLellan

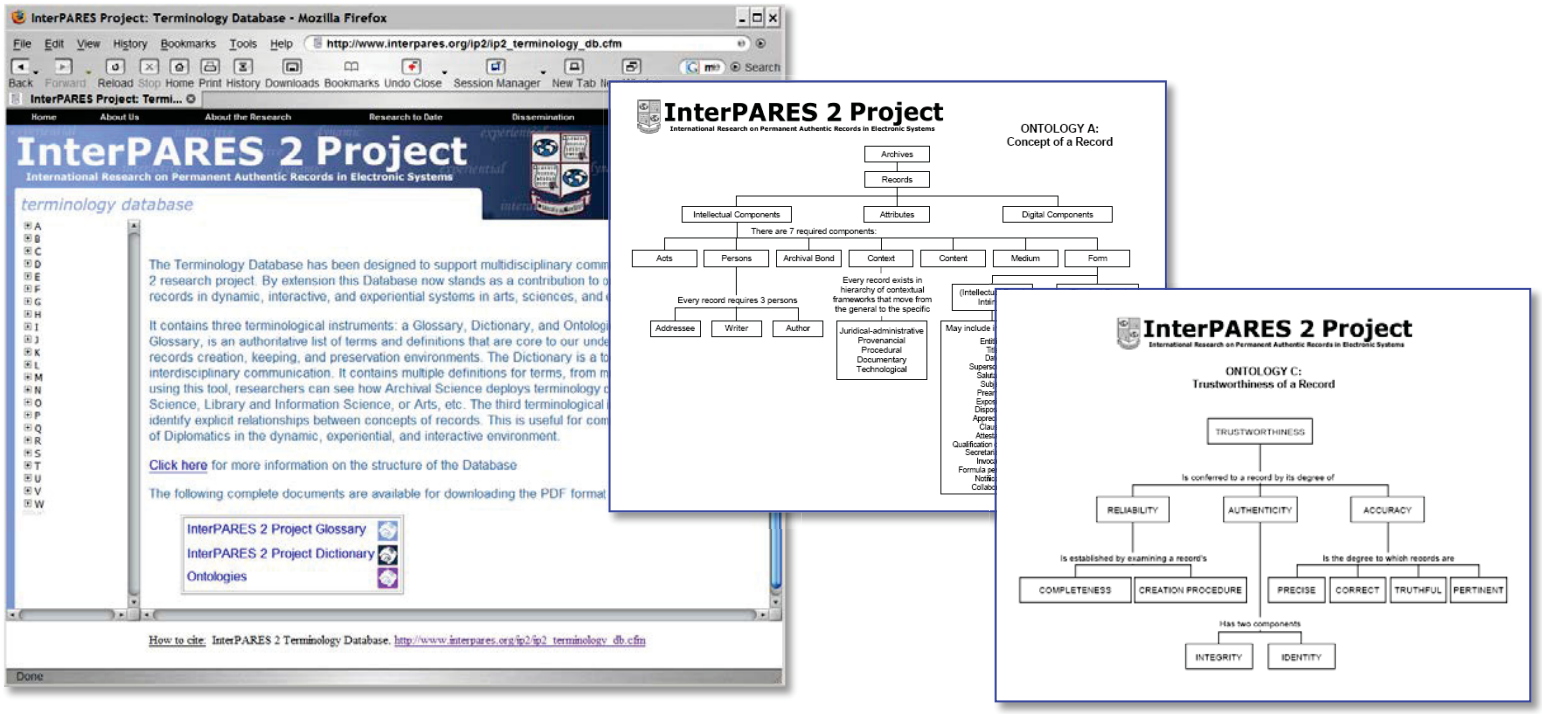
Table of Contents

InterPARES 2 Project, Domain 3 v1.1 (March 2007) i

IP 1 & 2 Products

Terminology Database

Including a glossary, a dictionary and ontologies

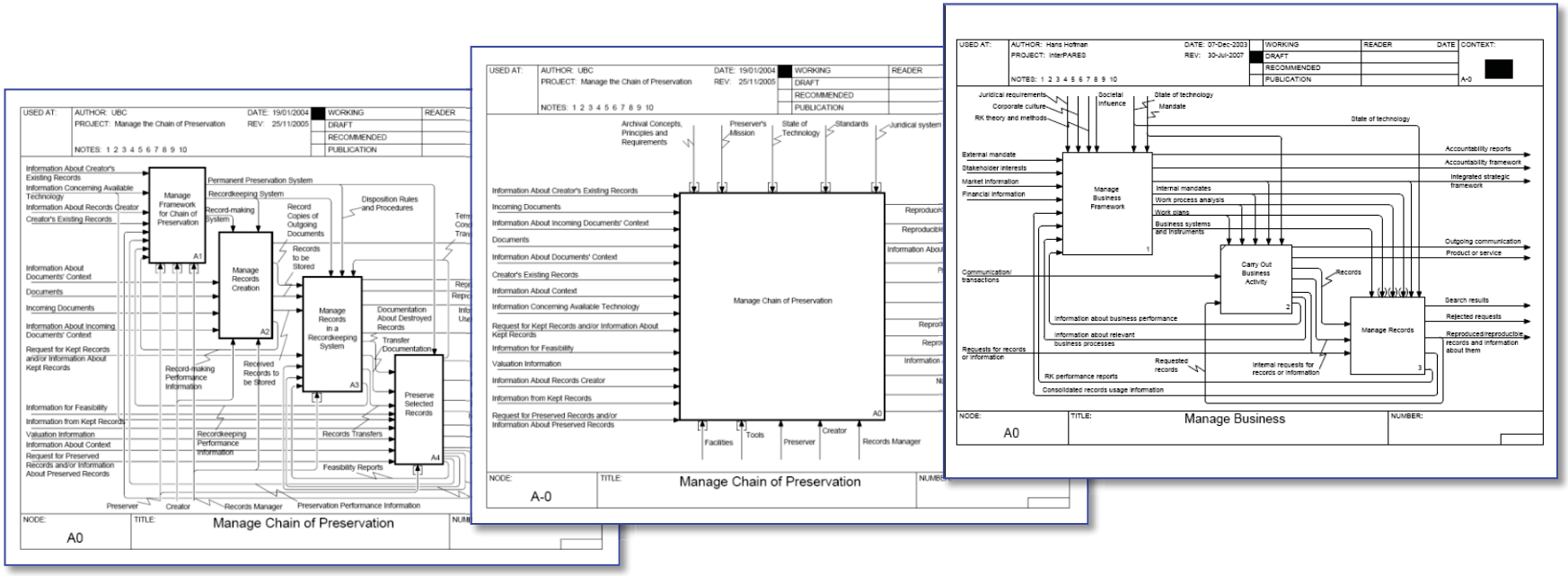


IP 1 & 2 Products

Two Records Management Models

Chain of Preservation (COP) Model (lifecycle)

Business-driven Recordkeeping (BDR) Model (continuum)



IP 1 & 2 Final Products

Two books:

Luciana Duranti, ed. *The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project* (San Miniato: Archilab, 2005). Available on line at http://www.interpares.org/book/index.cfm

Luciana Duranti and Randy Preston, eds. *InterPARES 2: Interactive, Dynamic and Experiential Records* (Padova, Italy: Associazione Nazionale Archivistica Italiana, 2008). Available on line at http://www.interpares.org/ip2/book.cfm

InterPARES 3 International Alliance

- **Teams:** TEAM (**T**heoretical **E**laboration into **A**rchival **M**anagement) Canada (including US); Brazil; Catalonia; China; Colombia; Italy; Korea; Malaysia; Mexico; Norway (till 2009); Singapore (till 2009); and Turkey.
- **Director:** Luciana Duranti
- **Headquarters:** UBC - SLAIS (facilities provided by UBC)
- **Funding:** SSHRC, and various sources from each country

3 Primary Components

1. **Research component**
   (short-term and long-term projects, including case studies and general studies)
2. **Education and training component**
   (in the context of research projects, apprenticeships, activities credited as part of coursework, etc.)
3. **Knowledge-mobilization component** (workshops, seminars, colloquia, policy manuals and other publications, public lectures, etc., that meet the needs of both academic and community partners)

Methodology

Action Research

- Practical, collaborative, pragmatic research directed toward producing solutions that were **directly useful** to a group of people
- Research **subjects were co-participants** and stakeholders in the process
- We **jointly** defined research objectives and goals, co-constructed research questions, pooled knowledge and **developed solutions** and performance tests that implemented specific strategies

Methodology (cont.)

Action Research

Two distinct methods of research:

1. **Prototype development research**
2. **Ethnographic research**

Methodology (cont.)

Prototype development research

- **User-centered, collaborative** prototyping approach that explores the interplay between theory and practice
- **Proof-by-demonstration**
- Comprises three major **iterative** stages:
  1. concept building
  2. system building
  3. system evaluation

Methodology (cont.)

Ethnographic research

- Creators of records, their users and archivists form a community of practice—**the archival environment**—for which social interaction creates meaning and defines values
- Researchers place themselves within an archival environment to **gain the cultural perspective** of those responsible for records
- **Observation** of the environment with detailed description, extensive **interviewing and analysis** of the documents

Case Studies

Case studies (3 types):

1. Dealing with specific groups or types of **records**.
2. Dealing with record/information **systems**.
3. Dealing with **policies**.

- The entire process was guided by **a case study flowchart**, which ensured that all steps were followed in the correct order, and was concluded by **a final report**.
- Each document produced in the course of the case study was structured as to form and content on the basis of **a template used for all case studies**.

General Studies

General studies (3 types):

1. Studies carried out by **one TEAM or a group of TEAMs for the benefit of all** TEAMs (e.g., Annotated Bibliography of International Standards, E-mail Preservation)
2. Studies in which **all TEAMs** take part (e.g., Terminology Database).
3. Studies conducted by a TEAM that will prove **useful for that TEAM only** (e.g., Annotated Canadian Standards)
4. Studies conducted by **a group of TEAMS for its own benefit** (e.g. Protocol Register)

General Studies (collaborations)

- Web 2.0/Social Media (Canada and Turkey)
- **Terminology (International Alliance)**
- Digital Preservation Projects (International Alliance)
- International Standards Relevant to IP3 (International Alliance)
- Bibliographic Database
- E-mail Preservation (Italy)
- Protocol Registry (Catalonia, Italy, Brazil)

General Studies (TEAM Canada)

- Canadian Standards Relevant to IP3
- Community Archives e-Records Assessment
- Public Sector Audit Report for Digital Recordkeeping
- Records Management Policies and Procedures Template
- Cost-benefit Models
- Ethical Models
- File Viewers
- **Education Modules**
- Open Source Records Management Software
- Metadata Applications Profiles
- Organizational Culture & Risk Assessment

Findings

- **Conceptual**
  - The Concept of Record
  - The Concept of Trustworthiness
  - The Concept of Life Cycle
- **Methodological**
  - Appraisal
  - Preservation Concept and Procedure
- **Strategic**
  - Relationship Creator-Preserver
  - The Role of the Archivist

The Concept of Record

- **Record**: any document made or received by a physical or juridical person in the course of activity as an instrument and by-product of it, and kept for action or reference
- **Document**: recorded information (i.e., information affixed to a medium in an objectified and syntactic form)
- **Information**: "intelligence given," or a message intended for communication across time and space
- **Data**: the smallest meaningful piece of information

Digital Record Characteristics

- **Act**: an action in which the records participates or which the record supports (naturalness and impartiality)
- **Persons Concurring to Its Creation**: author, writer, originator, addressee, and creator
- **Archival Bond**: explicit linkages to other records inside or outside the system
- **Identifiable Contexts**: juridical-administrative, provenancial, procedural, documentary, technological
- **Medium**: necessary part of the technological context, not of the record
- **Fixed Form and Stable Content**

Fixed Form

- An entity has 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)
- An entity has fixed form also 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)

Stable Content

- An entity has stable content if the data and the message it conveys are **unchanged and unchangeable**, meaning that data cannot be overwritten, altered, deleted or added to

- **Bounded Variability**: when changes to the documentary presentation of a determined stable content 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 Record Characteristics (cont.)

- **Formal Elements**: constituent parts of the record documentary form as shown on its face, e.g. address, salutation, preamble, complimentary close
- **Metadata**: the attributes of the records that demonstrate its identity and integrity
- **Digital Components**: stored digital entities that either contain one or more records or are contained in the record and require a specific preservation measure

Stored and Manifested Records

- **Stored record**: it is constituted of the digital component(s) used in re-producing it, which comprise 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)—e.g. instructive and enabling records

- **Manifested record**: the visualization or instantiation of the record in a form suitable for presentation to a person or a system. Sometimes, it does not have a corresponding stored record, but it 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

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

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

Interactive Entities

- **Non-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 either information systems or potential 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 that is what it purports to be, untampered with and uncorrupted

based on:

- identity
- Integrity
- reliability of the system

Authenticity: Identity

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

<u>Identity metadata</u>:

- names of the 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
- digital presentation
- the indication of any attachment(s)
- digital signature
- name of the person responsible for the business matter

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

New Concept of Records Life-Cycle

We cannot maintain or preserve digital records, but only the ability to re-produce or re-create them, therefore

- re-productions of digital records, if made by the creator in the course of and for the purposes of its business, are **records of the creator**, while
- if made by the preserver in the course and for the purposes of archival functions, are **authentic copies of the records of the creator**

Digital Records Life-Cycle

It comprises **two phases** related to who reproduces the record: whether the creator's or the preserver's

Phase 1: **Records of the creator**: the re-productions made by the creator in the usual and ordinary course of its activities behave and have to be treated as **originals** every time they are used and acted upon

Phase 2: **Authentic copies of the records of the creator**: the re-productions made by the preserver cannot be treated as originals because the creator has never used or acted upon them after re-production, which is made for preservation purposes rather than for the purposes of the creator's activities.

Consequences

- The **creator** can decide at any time to stabilize its own fluid entities and to give them the most useful, accessible, interoperable form, or the form that best serves its present and projected needs
- The **preserver** can only preserve what it receives from the creator by making an authentic copy of it
- Whether the stabilized record of the creator and its authentic copy made by the preserver are to be considered **trustworthy depends on the context** in which they are created and used

Preservation

Involves the creation by the preserver of authentic copies of the records of the creator. Their authenticity is guaranteed by:

- a controlled process of **migration** of the acquired records to the archives technological environment (always keeping the records also in the format in which they were acquired)
- the accurate **documentation** of any change that the records undergo during such process and every time that the archives technological environment is upgraded
- the implementation and **monitoring** of privileges concerning the access, use and reproduction of the records within the archives

Preservation (cont.)

- the establishment of **procedures** to prevent, discover, and correct loss or corruption of records, as well as
- procedures to guarantee the continuing identity and integrity (i.e. **authenticity**) of the records against media deterioration and across technological changes; and
- if **authentication** of individual records is required, by the existence of rules determining responsibility for and means of authentication.

Archival Description

Archival description acquires 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 their administrative, custodial and technological history, the illustration of their scope and content, and 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 transformations** while presenting them as a universe in which each part is subject to the bond of a common provenance and destination

Methodological & Strategic Findings

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 continue 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 Role

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**
4. Determines the **feasibility of preservation** on the basis of the archives technological capacity

The Archivist's New Role (cont.)

5. Determines a **preservation strategy** independently of technological trends (tries to influence the industry through the adoption of standards, but not vice versa) and maintains a 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**
8. Recognizes to **archival description a primary authentication function**

The Archivist's New Role (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
   - If this entity has several instantiations, which can be regarded as the record (manifested or stored entity; if the former, which)
   - How to keep such entities accurate and authentic through time
   - How to enable users to verify such authenticity over time

Other Key Findings

- Developing, learning and teaching **how to use** the structural features of an application or a system **is important**
- More important is **learning the spirit behind those features**
- **Users who are not acquainted with archival principles and methodologies may** – intentionally or unintentionally – **appropriate an application or a system "unfaithfully"** more easily than records professionals
- **With digital tools**, which are **mostly developed by IT experts** outside the organization that will use them and often without consulting archival professionals, **unfaithful appropriations are likely to happen more frequently**
- **Interpretive flexibility** or lack of it may determine acceptance or rejection

Other Key Findings (cont.)

- **Training** for users of applications and systems **"emphasizes details of use rather than general philosophy."**
- **Time is important**: the moment of the launch of a new system is very critical for its success
- **Managers favor** the **explicit knowledge** that is incorporated in organizational artifacts like processes, structures, documents, and technology
- Thus, it has been common to design **systems primarily focused on the codified, explicit organizational knowledge**
- Management reporting systems, decision support systems, and ERMS, are all focused on the identification, collection, and dissemination of this knowledge type

Other Key Findings (cont.)

- Rather, we should pay more attention to **knowledge management** literature
- **A core competency for implementers requires *know-how***, i.e. "the particular ability to put know-what into practice"
- **Fostering this more complex form of organizational capital should be the focus of our case studies**
- The outcome of our efforts will be successful only if
  - we are able to make the archival environment **understand the spirit of what we recommend**
  - we will be able **to incorporate into our recommendations the outlook and way of working of those whom it intends to serve.**

InterPARES Website: Products

- The 3rd phase of InterPARES was completed on 31 March 2012.
- The research findings and products are being uploaded on the public area of the InterPARES website.
- From the home page (www.interpares.org) the material can be accessed by clicking on Products and selecting InterPARES 3.
- Then, you can either click again on Products, and select the material by type (Case Studies, General Studies, Teams Reports) , or you can search by keywords, or select the theme you are interested in from the list on the page.
- It will take another couple of months to finalise all products and complete the uploading.

So, keep checking and stay tuned for the next phase.