Long-Term Preservation of Authentic Digital Scientific Data: The InterPARES Approach

Sherry (Li) Xie
PhD Student
University of British Columbia, Canada
2006-10-24
Presentation Outline

• **InterPARES project**
  – covering scientific activities
  – Managing scientific data as records

• **InterPARES products**
  – Theoretical knowledge
  – Methodological knowledge

• **Policy Frameworks**
  – Consistent with theoretical/methodological developments
  – Flexible for application: scientific data records
InterPARES Project

International Research on Permanent Authentic Records in Electronic Systems

www.interPARES.org

Records

• Recorded information
• Practical activities
• Instruments
• Byproducts
InterPARES Project

- Phase 1 (1999-2001): IP1
  - UBC Project
    - Current records
- Phase 2 (2002-2006): IP2
- Long-term preservation entails
  - Identification of digital records
  - Lifecycle management
- Theoretical and methodological knowledge
InterPARES Project 2

• Same theoretical framework
  – Diplomactics & Archival Science
• Interdisciplinary and open inquiry
• Extension of IP 2 scope
  – from document management system to dynamic, experiential, interactive systems
  – government records to artistic and scientific records
## IP2 Intellectual Organization

<table>
<thead>
<tr>
<th>IP Organization</th>
<th>Focus 1 Artistic activities</th>
<th>Focus 2 Scientific activities</th>
<th>Focus 3 Government activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain 1 Records Creation &amp; Maintenance</td>
<td>Working Group 1.1</td>
<td>Working Group 1.2</td>
<td>Working Group 1.3</td>
</tr>
<tr>
<td>Domain 2 Reliability, Accuracy &amp; Authenticity</td>
<td>Working Group 2.1</td>
<td>Working Group 2.2</td>
<td>Working Group 2.3</td>
</tr>
<tr>
<td>Domain 3 Selection &amp; Preservation</td>
<td>Working Group 3.1</td>
<td>Working Group 3.2</td>
<td>Working Group 3.3</td>
</tr>
</tbody>
</table>

**Focus 2 Scientific Activities**

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Policy</th>
<th>Description</th>
<th>Modeling</th>
</tr>
</thead>
</table>

---

*IP2 Intellectual Organization and the InterPARES 2 Project*
Scientific Data Records

- Scientific data records
  - Generated in the course of conducting scientific activities: instruments or byproducts
  - Raw data & processed data
    - SDP: level 0-4
    - Special data products
- Digital scientific data records
  - Stable content and fixed documentary forms
  - Fixed Forms medium
  - Activity
  - Persons
  - Contexts
IP2 SF Case Studies

- CS 06: Cybercartographic Atlas of Antarctica
- CS 08: Mars Global Surveyor Mission
- CS 14: Center for Desert Architecture
- CS 26: MOST Satellite Mission

------- Generation of SD and the activities of keeping, using and preserving SD as records
IP2 SF Case Studies

- CS 01: Persistent Archives
  - SDSC data grid technology
- CS 12: Antarctic Treaty Searchable Database
  - EvREsearch Ltd. Automated granularity
- CS 10: Science Data Archives/Repositories

------ Technological and procedural practices of managing SDR
Theoretical Knowledge:

- Concepts relating to digital preservation
  - Record
  - Reliability
  - Accuracy
  - Authenticity
  - Trusted custodian
  - Trusted recordkeeping system
  - Chain of preservation

Authenticity: the record is what it purports to be; it’s unchanged and unchangeable after creation; it’s not corrupted or tempered
IP2 Products - MK - I

**Methodological Knowledge:**

– Authenticity requirements (IP1)
  • Benchmark requirements: creators
  • Baseline requirements: preservers

– Chain of preservation
  • Reliable and accurate record making
  • Authentic record keeping,
  • Selection for preservation
  • Authentic preservation
IP2 Products - MK - II

Methodological Knowledge:

- MADRAS: Metadata and Archival Description Registry and Analysis System
- 2 models of preservation
  - MCP: record lifecycle
  - MBP: record continuum
IP2 Products - MK - III

Methodological Knowledge:
• 2 sets of guidelines covering production, maintenance, preservation
  • individuals creators, including scientists
  • records preservers
• 2 frameworks basing the development of preservation policies, strategies and standards
  • Organizations creating digital records
  • Digital records preservers
IP Preservation Frameworks

General

- Products of IP Policy Cross-domain Team
- Across 3 types of activities
- Addressing 3 domains
- 2 complementary sets of principles
  - Records Creators: C1 – C13
  - Records Preservers: P1 – P13
- Flexible and consistent
IP Preservation Frameworks

Audiences

- Records creators
  - Policies and strategies makers
  - National and international standards bodies
- Records preservers
  - Archival institutions or programs
  - Any other organizations/persons designated by the organization as preservers of their records

--- Movable responsibilities and close cooperation
IP Preservation Frameworks

Example Principles - PC

C1. Digital entities must have a fixed documentary form and a stable content to be considered as records and to be capable of being preserved over time. [P5]

C2. Records creation procedures should ensure that digital components of records can be separately maintained and reassembled over time [P4]

C4. Records creation and maintenance policies, strategies and standards should address the issues of record reliability, accuracy, and authenticity expressly and separately. [P2]
IP Preservation Frameworks
Example Principles - PP

P1. A designated records preserver fulfils the role of trusted custodian. [C8]

P3. Reproductions of a creator’s records made for purposes of preservation by their trusted custodian are to be considered authentic copies of the creator’s records. [C13]

P6. Preservation requirements should be articulated in terms of the purpose or desired outcome of preservation, rather than in terms of the specific technologies available. [C3]
IP Findings

Will be published in 2008 as two books

- Reports by individual IP research teams
- Articles addressing specific thoughts, issues, and challenges as they emerged in the research process

Accurate
Reliable
Authentic
Selected Bibliography


Questions?
IP Preservation Frameworks
Example Principles - PC

7. Preservation considerations should be embedded in all activities involved in records creation and maintenance if a creator wishes to maintain and preserve authentic records beyond its operational business needs [P7]

10. Third-party intellectual property rights attached to the creator’s records should be explicitly identified and managed in the record-making and recordkeeping systems. [P8]

12. Procedures for sharing records across different jurisdictions should be established on the basis of the legal requirements under which the records are created. [P13]