Beyond Preservation to Trust

Toward an Application Profile for Authenticity Metadata

Corinne Rogers

SAA 2011 – Research Day

Research Team: Joe Tennis, U Washington; Corinne Rogers, UBC; Elizabeth Shaffer, UBC
Outline

- Why an application profile for authenticity— theoretical underpinnings
- Functional requirements
- Domain models
- Metadata elements
- Crosswalks to existing schemas
- Next steps
Why an application profile for authenticity?

+ Metadata are human- and machine-readable assertions about information resources, discussed primarily in the context of access and retrieval, and preservation

+ The need for authenticity of those resources and the ability to authenticate them is recognized, but as yet not systematized

**The Goal:** 1) develop an application profile grounded in the findings of the InterPARES that connects theory with functional requirements for authenticity metadata, and 2) test it in the context of cIRcle, the University Institutional Repository at UBC
Context of this research

- International Research on Permanent Authentic Records in Electronic Systems 3 (InterPARES 3)
- University Institutional Repositories: Copyright and Long-Term Preservation
- cIRcle @ UBC
- City of Vancouver Archives & Archivematica
Functional Requirements

- Presumption of authenticity
- Interoperability
- Description
- Retrieval
- Meaningful display
- Parsimony
Authenticity

The trustworthiness of a record as a record; i.e., the quality of a record that is what it purports to be and that is free from tampering or corruption

Authenticity = Identity + Integrity

Identity: The whole of the characteristics of a document or a record that uniquely identify it and distinguish it from any other

Integrity: The quality of being complete and unaltered in all essential respects
Domain models

- Record
  - consists of
    - Attributes
      - comprise
        - Intellectual components
          - Digital components
  - links to
    - Policies & Procedures*
      - may include
        - Access privileges
        - Protective procedures
        - Establishment of documentary forms
        - Authentication methods or policies
        - Identification of authoritative record
        - Removal and transfer of relevant documentation
  - acquires/participates in
    - Controls/actions
      - may govern
        - Transfer
        - Maintenance
        - Modifications
        - Preservation
        - Disposition
        - Archival description

*may also include legislation, standards, codes of practice, and any other documentation that exercises control over records from within or outside the creating organization
Domain models

Attributes

comprise

Intellectual Components

Identity:
- Names of persons:
  - creator
  - author
  - writer
  - originator
  - addressee
  - recipient
  - Name of action or matter
  - Dates of creation and transmission:
    - chronological
    - received
    - archival
    - transmission
  - Expression of archival bond
  - Indication of presence of digital signature
  - Indication of other forms of authentication (corroboration, attestation)
  - Indication of attachments

Integrity:
- Name of handling office
- Name of office of primary responsibility (if different from HO)
- Indications of annotations?
- Indications of technical modifications (all previous digital components)
- Names/dates of actors in maintenance and preservation activities taken on the record

Physical:
- Carrier: bitstream hardware

Logical:
- File format(s)
- Embedded objects
- Linked objects
- Representation information (e.g. Style sheets), Technological means of authentication (e.g. seals, including electronic signatures)

*digital components are the physical manifestations of the current technological context (previous tech context belongs in the indications of modifications - linked to integrity)
What are the sources?
Chain of preservation
Areas of description

**Identity**
- P-persons
- D-date
- S-subject (action or matter)
- B-bond
- A-attachments

**Integrity**
- T-technology
- F-form
- SS-seals & signs
- AU-means of authentication (inc digital sig, attestation etc)
- R-rights and access; H-handling (office)

**DO-external documentation and system metadata** (policy, context, appraisal, transfer, audits of system activity, requests on the records)
## Elements

| A2.2.3.5 | P6 | - *receiver(s) or recipient(s)*—that is, the physical or juridical person(s) to whom the document may be copied or blind copied for information purposes; |
| A2.2.3.6 | S1 | • name of the action or matter—that is, the subject line(s) and/or the title at the top of the document; |
| A2.2.3.7 | AU1 | • indication of the presence of a digital signature; |
| A2.2.3.8 | AU2 | • indication of other forms of authentication, including: |
| A2.2.3.9 | AU3 | • *corroboration*—that is, an explicit mention of the means used to validate the document; |
| A2.2.3.10 | AU4 | • *attestation*—that is, the validation of the document by those who took part in the issuing of it, and by witnesses to the action or to the ‘signing’ of the document; |
| A2.2.3.11 | AU5 | • *subscription*—that is, the name of the author or writer appearing at the bottom of the document; and |
| A2.2.3.12 | A1 | • *qualification of signature*—that is, the mention of the title, capacity and/or address of the person or persons signing the document; |
| A2.2.4 |  | • indication of any attachments—that is, mention of autonomous digital objects linked inextricably to the document. |

### A2.2.4

**Declaring Captured and Identified Documents as Records (A2.2.4)**

Identity metadata captured for this activity could include:

- **New metadata**
- A2.2.4.1 | B1 | • classification code; and |
- A2.2.4.2 | B2 | • registration number.
## Crosswalks

- CoP model
- PREMIS
- Dublin Core
- MoReq 2010

<table>
<thead>
<tr>
<th>Function</th>
<th>MoReq 2010</th>
<th>Dublin Core</th>
<th>CoP</th>
<th>AuthAP</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>creation</td>
<td></td>
<td></td>
<td>A2.2.5.2</td>
<td>D04</td>
<td>transmission date, time and/or place;</td>
</tr>
<tr>
<td>creation</td>
<td></td>
<td></td>
<td>A2.2.5.4</td>
<td>D05</td>
<td>dates and times of further action or transmission; and</td>
</tr>
<tr>
<td>creation</td>
<td>M14.4.1 (M14.4)</td>
<td></td>
<td>A2.3.2.1</td>
<td>D06</td>
<td>archival or filing date—that is, the date on which a record is officially</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>incorporated into the creator’s records;</td>
</tr>
<tr>
<td>creation</td>
<td>M14.4.28</td>
<td></td>
<td>A2.3.3.4</td>
<td>D07</td>
<td>date/time of the transfer.</td>
</tr>
<tr>
<td>management</td>
<td></td>
<td>modified</td>
<td>A3.2.3.2.5</td>
<td>D08</td>
<td>date/time of any modification(s);</td>
</tr>
<tr>
<td>management</td>
<td>M14.4.28</td>
<td></td>
<td>A3.2.3.2.8</td>
<td>D09</td>
<td>date/time the record(s) was/were placed in storage; and</td>
</tr>
<tr>
<td>management</td>
<td></td>
<td>modified</td>
<td>A3.2.3.3.2.5</td>
<td>D10</td>
<td>date/time of the backup;</td>
</tr>
<tr>
<td>management</td>
<td></td>
<td>modified</td>
<td>A3.2.3.3.6</td>
<td>D11</td>
<td>date/time of the correction(s);</td>
</tr>
<tr>
<td>management</td>
<td></td>
<td>modified</td>
<td>A3.2.3.3.4.6</td>
<td>D12</td>
<td>date/time of the update(s);</td>
</tr>
<tr>
<td>management</td>
<td></td>
<td>modified</td>
<td>A3.2.3.3.5.6</td>
<td>D13</td>
<td>date/time of the refreshing(s);</td>
</tr>
</tbody>
</table>
## Term Records

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property URI</td>
<td><a href="http://interpares.org/aapip/author">http://interpares.org/aapip/author</a></td>
</tr>
<tr>
<td>Qualified name for property</td>
<td>aapip:author</td>
</tr>
<tr>
<td>Defined by</td>
<td><a href="http://interpares.org/aapip">http://interpares.org/aapip</a></td>
</tr>
<tr>
<td>Type of term</td>
<td>property</td>
</tr>
<tr>
<td>Subproperty of</td>
<td>NA</td>
</tr>
<tr>
<td>Source label</td>
<td>author</td>
</tr>
<tr>
<td>Source definition</td>
<td>(IP definition of author)</td>
</tr>
<tr>
<td>Usage in this AP</td>
<td>(definition as necessary)</td>
</tr>
<tr>
<td>Comments for this AP</td>
<td></td>
</tr>
<tr>
<td>Uses vocabulary encoding scheme</td>
<td>No</td>
</tr>
<tr>
<td>Obligation</td>
<td>Mandatory (Recommended or Optional)</td>
</tr>
<tr>
<td>Condition</td>
<td>Mandatory if...</td>
</tr>
<tr>
<td>Minimum occurrences</td>
<td>0</td>
</tr>
<tr>
<td>Maximum occurrences</td>
<td>No limit</td>
</tr>
</tbody>
</table>
Next steps

+ Refine – finish creating term records
+ Test in cIRcle, the University of British Columbia’s institutional repository
  + Test case: student theses and dissertations
  + Currently using enhanced Dublin Core, primarily for access
  + These are university records and as such require a more robust metadata application to ensure their authenticity over time
+ Refine – *parsimony*

+ Make available
Thank you!

- Corinne Rogers cmrogers@interchange.ubc.ca
- Joe Tennis jtennis@uw.edu
- Elizabeth Shaffer eshaffer@interchange.ubc.ca