A practical way to get shared URLs into Cultural Heritage data

Richard Light
Richard Light Consultancy
Open Data on the Web, 24 April 2013
Cultural heritage data sharing

- IRGMA multidisciplinary data standard (1960s)
- MDA data standard (1980s)
- CIDOC Reconciliation of standards WG (1980s)
- CIDOC relational data model (1990s)
- SPECTRUM (1990s-) incl. XML Schema
- CIDOC Conceptual Reference Model (1990s-)

*All structural frameworks*
Terminology control

• Vocabularies/thesauri
  – Getty vocabs. (AAT, ULAN, TGN)
  – ICOM Costume vocabulary (1981)
  – Herts. Simple Name termlist

• Classifications
  – ICONCLASS
  – SHIC

  *Wide use; local customization; ‘locked into’ systems*
Cultural Linked Data (1)

• Some frameworks: notably ICONCLASS
Cultural Linked Data (2)

• Some resources, notably BM:

```xml
- <rdf:RDF>
  - <rdf:Description rdf:about="http://collection.britishmuseum.org/id/object/EAF119772">
    <ns1:P24B.changed_ownership_through rdf:resource="http://collection.britishmuseum.org/id/object/EAF119772/acquisition"/>
    <ns1:P2F.has_type rdf:resource="http://collection.britishmuseum.org/id/thesauri/x5373"/>
    <ns1:P2F.has_type rdf:resource="http://collection.britishmuseum.org/id/thesauri/x7812"/>
    <ns1:P3F.has_note>Acquisition date :: 1994 ::</ns1:P3F.has_note>
    <ns1:P3F.has_note>Object type :: basket ::</ns1:P3F.has_note>
    <ns1:P3F.has_note>Object type :: marriage equipment ::</ns1:P3F.has_note>
    <ns1:P48F.is_composed_of rdf:resource="http://collection.britishmuseum.org/id/object/EAF119772/1"/>
    <ns1:P50F.has_current_keeper rdf:resource="http://collection.britishmuseum.org/id/department/E"/>
    <ns1:P52F.has_current_owner rdf:resource="http://collection.britishmuseum.org/id/the-british-museum"/>
    <ns1:P57F.has_number_of_parts>1</ns1:P57F.has_number_of_parts>
    <ns2:codex_id>3380108</ns2:codex_id>
    <ns3:sameAs rdf:resource="http://collection.britishmuseum.org/id/codex/3380108"/>
  </rdf:Description>
</rdf:RDF>
```
Current state

• Cross-sectoral discussions (libraries, archives, museums)
• Design patterns e.g. uncertainty, attribution
• Potential use of CIDOC CRM as ‘mortar’ (awaiting BM guidelines)
• Few shared ontologies expressed as Linked Data
• Linked Data resources full of strings, or silos
Desired state

• Cultural heritage resources routinely published as Linked Data
• Intra-domain structural compatibility, e.g. using CIDOC CRM
• Shared use of central authorities for places, people, events, leading to ...
• Distributed database of cultural heritage/history
Modes and data control

• Modes: native XML database, widely used for cataloguing U.K. museums
• Supports “termlists”: primary keys in one file control data values in another
• Publishes a standard set of “termlists”
• Extended termlist concept to “web termlists”: use a web service to control data
  – a means of getting LD URLs into Modes data
<table>
<thead>
<tr>
<th>BCRTM:</th>
<th>M1930.1</th>
<th>watch</th>
<th>Pocket watch, silver</th>
<th>purchase</th>
<th>Windermere, 40 Staithe, 1841930</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCRTM:</td>
<td>M1930.1a</td>
<td>postcard</td>
<td>Pocket watch, silver</td>
<td>purchase</td>
<td>Windermere, 40 Staithe, 1841930</td>
</tr>
<tr>
<td>BCRTM:</td>
<td>M1930.1b</td>
<td>postcard</td>
<td>Pocket watch, silver</td>
<td>purchase</td>
<td>Windermere, 40 Staithe, 1841930</td>
</tr>
<tr>
<td>BCRTM:</td>
<td>M1930.1d</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Select from termlist**

**Term: bequest**

bequest is the preferred term for bequeathed, bequested and gift by executor.
“Web termlist” (Geonames)
SPARQL end-point as web term list
Control record for web termlist
Modes Linked Data Framework

- Generic framework for publishing *any* Modes resource as Linked Data
- Based on a 404 Not Found handler (js)
- Control records state what can be published
- Uses XSLT to transform and deliver content
- Uses Linked Data API-like approach to URL structure
- Simple search; no SPARQL support
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```xml
<rdf:RDF>
  <skos:ConceptScheme rdf:about="http://data.modes.org.uk/Termist/SFIC_termlist/1">
    <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
    <skos:prefLabel xml:lang="en">SFIC</skos:prefLabel>
    <skos:scopeNote xml:lang="en">SFIC: Standard Family Information Code</skos:scopeNote>
  </skos:ConceptScheme>
</rdf:RDF>
```
Shakespeare’s plays
Attack problem from both ends

• Publish as Linked Data
  – Shared ontologies, vocabularies, classifications
  – Resources (e.g. Shakespeare; museum catalogues)

• Consume Linked Data
  – Use “web termlists” to grab URLs for use alongside (or instead of) string-value data

• Source-agnostic: there is nothing Modes-specific about interface or data

• Adlib, CALM also developing web termlists
Dog food! Modes LD as web termlist