TEI and Cultural Heritage Ontologies

Interchange of information?

Øyvind Eide & Christian-Emil Ore

Unit for Digital Documentation,
University of Oslo, Norway
Motivation: Grey literature in Museums 1

Original text (text witness)
  ↓
  Step 1: registration

Bibliographical record
  ↓
  Step 2: reproduction

Facsimile
  ↓
  Step 3: transcription

Text with XML markup
  1) Structural markup
     (2) Lemmatization etc.

Text with XML markup
  ↓
  Step 4: content markup

Museum database
  Artefacts, excavations, referential information

Event/object oriented model
  (CIDOC-CRM compatible)
Motivation: Grey literature in Museums 2

A fragment of an imaginary report:

The excavation in Wasteland in 2005 was performed by Dr. Diggey. He had the misfortune of breaking the beautiful sword (C50435) into 30 pieces.
The excavation in <name type="place" id="n1">Wasteland</name> in <date id="d1">2005</date> was performed by <name type="person" id="n2">Dr. Diggey</name>. He had the misfortune of breaking the beautiful sword identified by Identifier: C50435 into 30 pieces.
Strategies for integration

- Three possible strategies for combining extracted information and the TEI tagged text

  - Store the information as an external XML-document e.g. RDF or CRM-Core. Can be stored together with the TEI document (eg by using METS)

  - Store the information in the TEI-header using an external XML name space, e.g. RDF

  - Store the information in the TEI-header using the existing elements in TEI-P5.
The CIDOC CRM
Top-level Classes relevant for Integration

- E55 Types
  - refer to / refine

- E39 Actors (persons, inst.)
- E28 Conceptual Objects
- E18 Physical Things
- E52 Time-Spans
- E53 Places

- E2 Temporal Entities (Events)
  - participate in
  - affect or refer to
  - within
  - have location
  - at
Corresponding TEI-P5 elements

<person> provides information about an identifiable individual, for example a participant in a language interaction, or a person referred to in a historical source.

<org> (organization) provides information about an identifiable organization such as a business, a tribe, or any other grouping of people.

<place> contains data about a geographic location

<event> contains data relating to any kind of significant event associated with a person, place, or organization.

<relation> (relationship) describes any kind of relationship or linkage amongst a specified group of participants.
Relations between event, place and person
<person xml:id="WM"> <!-- ... -->
<event type="marriage" when="1859-04-26">
<label>Marriage</label>
<desc>
<name type="person" ref="#WM">William Morris</name> and <name type="person" ref="#JBM">Jane Burden</name> were married at <name type="place">St Michael's Church, Ship Street, Oxford</name> on <date when="1859-04-26">26 April 1859</date>. The wedding was conducted by Morris's friend <name type="person" ref="#RWD">R. W. Dixon</name> with <name type="person" ref="#CBF">Charles Faulkner</name> as the best man. The bride was given away by her father, <name type="person" ref="#RB">Robert Burden</name>. According to the account that <name type="person" ref="#EBJ">Burne-Jones</name> gave <name type="person" ref="#JWM">Mackail</name> <quote>M. said to Dixon beforehand <said>Mind you don't call her Mary</said> but he did</quote>. The entry in the Register reads: <quote>William Morris, 25, Bachelor Gentleman, 13 George Street, son of William Morris decd. Gentleman. Jane Burden, minor, spinster,..</quote>
</desc>
<bibl>J. W. Mackail, <title>The Life of William Morris</title>, 1899.</bibl>
</event>
</person>

<relation name="spouse" mutual="#WM #JBM"/>
<relation name="best man" active="#RWD" passive="#WM"/>
<relation name="parent" active="#RB" passive="#JBM"/>
<classDecl> (classification declarations) contains one or more taxonomies defining any classificatory codes used elsewhere in the text.

<taxonomy> defines a typology used to classify texts either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy.

<catRef/> (category reference) specifies one or more defined categories within some taxonomy or text typology.
<strong>state</strong> contains a description of some status or quality attributed to a person, place, or organization at some specific time.

<br/>

<strong>trait</strong> contains a description of some culturally-determined and in principle unchanging characteristic attributed to a person or place.
<table>
<thead>
<tr>
<th>Types, attributes and classification 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Person</strong></td>
</tr>
<tr>
<td>Trait-like</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Faith</td>
</tr>
<tr>
<td>langKnowledge</td>
</tr>
<tr>
<td>Nationality</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>socecStatus</td>
</tr>
<tr>
<td>trait</td>
</tr>
<tr>
<td>Stat-like</td>
</tr>
<tr>
<td>affiliation</td>
</tr>
<tr>
<td>education</td>
</tr>
<tr>
<td>floruit</td>
</tr>
<tr>
<td>occupation</td>
</tr>
<tr>
<td>persName</td>
</tr>
<tr>
<td>residence</td>
</tr>
<tr>
<td>state</td>
</tr>
<tr>
<td>state</td>
</tr>
</tbody>
</table>
Suggestion:
    Introduce a <description> element

where

<description type="trait"> == <trait>
<description type="state"> == <state>
Summing up – conclusions

• TEI-P5 introduces several new useful “ontological” elements.

• Suggested extensions and adjustments:
  – Introduce an element `conceptualObject` for conceptual/abstract objects.
  – Introduce an element `physicalObject` for physical objects.
  – Extend the scope of `relation` to the object elements and to event and add the type attribute.
  – Extend the scope of `taxonomy` to non-textual entities
  – Extend the scope of `desc` to all ontological elements and let desc be a super element of the classification elements, eg. `<age>` will be equal to `<desc type="age">`.
  – Consider to state explicitly other equivalences like `<publisher>` and `<name type="publisher">`.