Encoding Names, People, Places and Dates

Magdalena Turska

April 2014

Names, People, and Places

Names and other references to objects appear in most texts. Exactly how this appearance is made can very significantly differ - from text to text, but between references within the same text as well..

> "*My dear Mr. Bennet,*" *said his lady to him one day,* "*have you heard that Netherfield Park is let at last?*"
> *Mr. Bennet replied that he had not.*
> "*But it is,*" *returned she;* "*for Mrs. Long has just been here, and she told me all about it.*"
> *Mr. Bennet made no answer.*

Names in TEI

TEI provides several ways of marking up names and nominal expressions:

- `<rs>` ("referring string") -- any phrase which refers to a person or place, e.g. 'the girl you mentioned', 'my husband'...
- `<name>` - any lexical item recognized as a proper name e.g. 'Siegfried Sassoon' , 'Calais', 'John Doe' ...
- `<persName>`, `<placeName>`, `<orgName>`: 'syntactic sugar' for `<name type="person">` etc.
- A rich set of elements for the *components* of such nominal expressions, e.g. `<surname>`, `<forename>`, `<geogName>`, `<geogFeat>` etc.

Names - example

```
 "Why, <rs>my dear</rs>, you must
know, <persName>Mrs. <surname>Long</surname>
</persName> says that
<placeName>Netherfield</placeName> is taken by a <rs>young
man of large
fortune from the north of England</rs>; that he came down on
Monday in a
chaise and four to see <rs>the place</rs>, and was so much
delighted with it,
that he agreed with <persName>Mr. <surname>Morris</surname>
</persName>
immediately; that he is to take possession before
Michaelmas, and some of his
servants are to be in the house by the end of next week."
"What is his name?"
"<persName>
 <surname>Bingley</surname>
</persName>."
```

Component attributes for <persName>

Handy attributes to categorize or sort them

```xml
<persName ref="tag:projectname.org,2012:pn9">
 <roleName type="honorific" full="abb">Mr</roleName>
 <forename sort="2">Sergei</forename>
 <forename sort="3" type="patronym">Mikhailovic</forename>
 <surname sort="1">Uspensky</surname>
</persName>
```

References may be ambiguous

```
<s>Jean likes <name ref="#NN123">Nancy</name>
</s>
```

Using a more precise element (<persName> or
<placeName>) is one way of resolving the ambiguity;
another is to follow the pointer:

```
<person xml:id="NN123">
 <persName>
  <forename>Nancy</forename>
  <surname>Ide</surname>
 </persName>
<!-- ... -->
</person>
```

or...

```
<place xml:id="N123">
 <placeName notBefore="1400">Nancy</placeName>
 <placeName notAfter="0056">Nantium</placeName>
<!-- ... -->
</place>
```

References are not the entities which they refer to

One entity(person, place, organisation) might be known by many names or might be referred to by some other description entirely.

Recognising the need to distinguish clearly the encoding of references from the encoding of referenced entities (occurrences in the real world) themselves, the TEI provides:

- <person> corresponding with <persName>
- <place> corresponding with <placeName>
- <org> corresponding with <orgName>
- and in addition <state>, <event> and other components for those

Where to store information about named entities?

Information about a person is stored within a `<person>` element. These elements may appear only within a `<listPerson>` element, eg within `<particDesc>` (participant description) element in the `<profileDesc>` element of a TEI header

```xml
<profileDesc>
 <particDesc>
  <listPerson type="historical">
   <person xml:id="ART1">
    <persName>Arthur</persName>
   </person>
   <person xml:id="BERT1">
    <persName>Bertrand</persName>
   </person>
<!-- ... -->
  </listPerson>
 </particDesc>
</profileDesc>
```

Basic <person>

```xml
<person xml:id="WO">
 <persName>
  <forename>Wilfred</forename>
  <forename>Edward</forename>
  <forename>Salter</forename>
  <surname>Owen</surname>
 </persName>
 <birth when="1893-03-18">
  <placeName>Oswestry</placeName>, 18th March
   1893</birth>
 <death when="1918-11-04">
  <placeName>Ors</placeName>, 4th November
   1918</death>
 <bibl type="wikipedia">
  <ptr target="http://en.wikipedia.org/wiki/Wilfred_Owen"/>
 </bibl>
</person>
```

What can we say about named entities?
Potentially, quite a lot...

```xml
<person xml:id="ID1485">
 <persName>Ioannes Dantiscus</persName>
 <persName>Johannes von Höfen</persName>
 <persName>Jan Dantyszek</persName>
 <persName>Johannes Flachsbinder</persName>
 <persName>Ioannes de Curiis</persName>
 <birth notBefore="1485-01-01"
  notAfter="1485-12-31">1485</birth>
 <death when="1548-10-27">†1548-10-27</death>
 <occupation>diplomat, neo-Latin poet and traveller</occupation>
 <occupation from="1504-01-01"
  to="1504-12-31">1504 royal
   scribe</occupation>
 <occupation from="1507-01-01"
  to="1507-12-31">1507 referendary for
   Prussian affairs at the court of Sigismund Jagiellon; </occupation>
 <occupation from="1508" to="1513">1508-1513 royal envoy to Prussian
   towns and to the Prussian assemblies;</occupation>
 <occupation from="1515">1515 secretary of the Polish legation at the
   imperial court; </occupation>
 <occupation from="1516" to="1532">in 1516-1532 envoy in the service of
   the king of Poland Sigismund Jagiellon and emperors Maximilian and
   Charles V of Habsburg; </occupation>
 <event when="1529">Kulm canon; </event>
 <occupation from="1530" to="1537">1530-1537 bishop of Kulm; </occupation>
 <occupation from="1537" to="1548">1537-1548 bishop of
   Ermland</occupation>
</person>
```

iT services

Traits, States, and Events

Inside entities there are generally three *classes* of information:

- <state>: more general-purpose, but usually a time-related property (e.g. <occupation>, <floruit>, <education>)
- <trait>: if you want to a distinguish between time-bound and static, use this for properties that (usually) don't change over time (e.g. <faith>, <langKnowledge>, <nationality>, <sex>, <climate>, <location>, <population>
- <event>: an independent event in the real world which may lead to a change in state or trait (e.g. <birth> or <death> for a person, a war for a place)

Additionally, all these elements are members of the 'datable' class so can have time/dating attributes.

How do we identify the entity being named?

Every element which is a member of the att.naming class inherits two attributes from the att.canonical class:

@key provides an externally-defined means of identifying the entity (or entities) being named, using a coded value of some kind.

@ref provides an explicit means of locating a full definition for the entity being named by means of one or more URIs.

Arguably, *@key* is redundant, since *@ref* is defined as anyURI, this can point from the name instance to the *@xml:id* of metadata about the entity, prefixing it with a '#' if in the same file, or use a private URI syntax.

Pointing Mechanisms

The ref attribute can take any kind of pointer.
Entity defined within the same XML document

```
That silly man<name ref="#DPB1" type="person">David Paul
Brown</name> has suffered ...
```

or in some other place, refered to by means of a URI

```
That silly man
<name ref="http://www.example.com/personography.xml#DPB1"
 type="person">David
Paul Brown</name> has suffered ...
```

Multiple pointers: reference to 'the Browns' might be
encoded

```
That wretched pair <name ref="#DPB1
#EBB1" type="person">the Browns</name> came to dine ...
```

Organizational names

An organizations is any named collection of people regarded as a single unit. An <orgName> can point back to an <org> in the header.

```
<p>On <date when="1915-10-21">21
    October 1915</date> Owen enlisted in the
<orgName ref="#AROTC">Artists' Rifles Officers' Training
Corps</orgName>.</p>
```

```
<org xml:id="AROTC">
<!-- Information about the organization -->
</org>
```

Components of place names

- <placeName> (names can be made up of other names)
- <geogName> a name associated with some geographical feature such as a mountain or river
- <geogFeat> a term for some particular kind of geographical feature e.g. 'Mount', 'Lake'

```
<placeName>
 <geogFeat>Mont</geogFeat>
 <geogName>Blanc</geogName>
</placeName>
```

A place is defined by its <location>

The <location> element can contain

- a more or less well-structured description using the hierarchy of place name components (a politico-geographical location)
- a set of geographical co-ordinates

```
<place xml:id="craiglockhart">
 <placeName>Craiglockhart War Hospital</placeName>
 <settlement>Edinburgh</settlement>
 <region>Scotland</region>
 <country key="UK">United Kingdom</country>
 <location>
  <geo>55.91812 -3.24019</geo>
 </location>
</place>
```

Another <location>

```xml
<place type="building">
 <placeName>Brasserie Georges</placeName>
 <location>
  <country key="FR"/>
  <settlement type="city">Lyon</settlement>
  <district type="arrondissement">Perrache</district>
  <placeName type="street">cours de Verdun</placeName>
 </location>
 <location>
  <geo>45.748 4.828</geo>
 </location>
</place>
```

\<listPlace\> in context of \<settingDesc\>

```xml
<settingDesc>
 <listPlace>
  <place xml:id="west01">
   <placeName>West Copice</placeName>
   <region>Shropshire</region>
   <note>'Westcopice' was approximately three-quarters of a
mile
        east of Sheinton, on the south bank of the Severn
opposite
        Buildwas, near the abbey ruins. Probably Henry
Wood's manor
        or estate is named in this reference.</note>
  </place>
  <place xml:id="shei01">
   <placeName>Sheinton</placeName>
   <region>Shropshire</region>
  </place>
  <place xml:id="shro01">
   <placeName>Shropshire</placeName>
  </place>
 </listPlace>
</settingDesc>
```

Shared personography across multiple TEI files

Most probably in real-life project there will be many TEI documents that should not repeat all the shared metadata about people and place. In such cases you might want to have one 'master' file for all people etc. What to do then? Instead of putting <listPerson> &co in the teiHeader just put in <include> element in its place, that points to a file that has <listPerson> as a root.

```
<profileDesc>
 <particDesc> <include
xmlns="http://www.w3.org/2001/XInclude" href="people.xml"/>
 </particDesc>
</profileDesc>
```

Linking with external authority files

Why?

- identification across projects
- linking
- reuse

It makes sense to provide links to external authority files, wherever appropriate records exist.

VIAF: The Virtual International Authority File http://viaf.org might be a sensible choice but there's no reason to limit yourself to just one.

How? By means of <idno> element.

Idno element example

```xml
<person>
 <persName>
  <forename>Edward</forename>
  <forename full="init">R.</forename>
  <surname>Tufte</surname>
 </persName>
 <idno type="scopus">6506403994</idno>
 <idno type="lcaf">http://id.loc.gov/authorites/names/n50012763.h
</person>
```

W3C Date Formats

All these events are 'datable' and so can be associated with a more or less exact date or date range using any combination of the following attributes:

@when supplies the value of a date or time in a standard form

@notBefore specifies the earliest possible date for the event in standard form

@notAfter specifies the latest possible date for the event in standard form

@from indicates the starting point of the period in standard form

@to indicates the ending point of the period in standard form

The 'standard form' is that defined by W3C. All dates are normalised to the Gregorian calendar.

The most commonly-encountered format for the date

part of the when attribute is yyyy-mm-dd, but yyyy

Thank You!

Any Questions?