

Named Entities

Encoding Names, Persons, Places and Dates



Content

- Names. Persons ↓ (3-21)
- Names. Places ↓ (22-28)
- Dates ↓ (29-32)
- References ↓ (33)

TEI NAMES. PERSONS

Identifying what are now often referred to as *named entities* from the rest of a sea of words is just as important a task for today's computational linguist building a language understanding system, or today's digital historian tracing social networks within a corpus of 18th century correspondence

– Lou Burnard (2014), *What is the Text Encoding Initiative?*, Marseille: OpenEdition Press.

Persons, Places, Dates

“I was born in the year 1632, in the city of York, of a good family, though not of that country, my father being a foreigner of Bremen, who settled first at Hull. He got a good estate by merchandise, and leaving off his trade, lived afterwards at York, from whence he had married my mother, whose relations were named Robinson, a very good family in that country, and from whom I was called Robinson Kreutznaer”

– Daniel Defoe, *The Life and Adventures of Robinson Crusoe*, 1719

Persons, Places, Dates

“I was born in the year 1632, in the city of York, of a good family, though not of that country, my father being a foreigner of Bremen, who settled first at Hull. He got a good estate by merchandise, and leaving off his trade, lived afterwards at York, from whence he had married my mother, whose relations were named Robinson, a very good family in that country, and from whom I was called Robinson Kreutznaer”

– Daniel Defoe, *The Life and Adventures of Robinson Crusoe*, 1719

Referring string

“I was born in the year 1632, in the city of York, of a good family, though not of that country, my father being a foreigner of Bremen, who settled first at Hull. He got a good estate by merchandise, and leaving off his trade, lived afterwards at York, from whence he had married my mother, whose relations were named Robinson, a very good family in that country, and from whom I was called Robinson Kreutznaer”

– Daniel Defoe, *The Life and Adventures of Robinson Crusoe*, 1719

Names in TEI

The TEI provides several ways of marking up names and nominal expressions.

Core module

- `<rs>` contains a string which refers to a person or place: "my mother", "he".
- `<name>` contains a proper noun: "Robinson Kreutznaer", "Bremen".
- `@type` differentiates persons and places.

More precise (extended module)

- `<persName>`, `<placeName>`, `<orgName>`, `<surname>`, `<forename>`.

Encoding of names using the element <name>

I was born in the year 1632, in the city of
<name>York</name>, of a good family, though not of
that country, my father being a foreigner of
<name>Bremen</name>, who settled first at
<name>Hull</name>. He got a good estate by
merchandise, and leaving off his trade, lived
afterwards at <name>York</name>, from whence he
had married my mother, whose relations were named
Robinson, a very good family in that country, and
from whom I was called <name>Robinson
Kreutznaer</name>

Add the attribute `@type` to differentiate types of names

I was born in the year 1632, in the city of `<name type="place">York</name>`, of a good family, though not of that country, my father being a foreigner of `<name type="place">Bremen</name>`, who settled first at `<name type="place">Hull</name>`. He got a good estate by merchandise, and leaving off his trade, lived afterwards at `<name type="place">York</name>`, from whence he had married my mother, whose relations were named Robinson, a very good family in that country, and from whom I was called `<name type="person">Robinson Kreutznaer</name>`

More precise tagging: *namesdates* module

He got a good estate by merchandise, and leaving off his trade, lived afterwards at
<placeName>York</placeName>, from whence he had married my mother, whose relations were named Robinson, a very good family in that country, and from whom I was called

<persName>

<forename>Robinson</forename>

<surname>Kreutznaer</surname>

</persName>

Components of personal names

- `<forename>`, a forename or given name.
- `<surname>`, a family name.
- `<roleName>`, role or position in society (President, Queen).
- `<addName>`, nickname, pseudonym, alias.
- `<nameLink>`, connecting phrase or link (de, of, van, von).

```
<persName>  
  <forename>Marie-Henri</forename>  
  <surname>Beyle</surname>  
  <addName type="pseudonym">Stendhal</addName>  
</persName>
```

Signs in the text vs. the thing signs refer to

Signifiant (signifier) vs. *Signifié* (signified)

“[...] and from whom I was called **Robinson Kreutznäer**; but, by the usual corruption of words in England, we are now called —nay we call ourselves and write our name— **Crusoe**; and so my companions always called me.”

Robinson Crusoe

TEI distinction: `<persName>` vs. `<person>`

Encoding of references

I was called `<persName>`Robinson Kreutznaer `</persName>`; but, by the usual corruption of words in `<placeName>`England`</placeName>`, we are now called –nay we call ourselves and write our name– `<persName>`Crusoe`</persName>`

Encoding of referenced entities

```
<person>
  <persName>
    <forename>Robinson</forename>
    <surname xml:lang="en" type="corrupted">Crusoe</surname>
    <surname xml:lang="de" type="birth">Kreutznaer</surname>
  </persName>
  <birth when="1632">1632 <placeName>York</placeName>
    </birth>
</person>
```

Reason for encoding **entities**

- To standardise references to the same thing: “Robinson Kreutznaer”, “Crusoe”, “castaway”.
- To facilitate a more detailed encoding of source documents (historical materials).
- To represent implicit data in a uniform way.
- To enable fine-grained searching of data-centric projects:
 - to set references to authority files and gazeteers (INSI, GeoNames, VIAF);
 - biographical dictionaries;
 - geographical dictionaries;
 - extracted indices.

Organising named entities

- Usually stored in the `<teiHeader>`

```
1 <teiHeader>
2   <fileDesc>...</fileDesc>
3   <profileDesc>
4     <particDesc>
5       <listPerson>
6         <person>...</person>
7         <person>...</person>
8       </listPerson>
9     </particDesc>
10    <settingDesc>
11      <listPlace>
12        <place>...</place>
13      </listPlace>
14    </settingDesc>
15  </profileDesc>
16 </teiHeader>
```

- Create a separate XML file using `<listPerson>` entries.

Linking the association with @xml:id ← @ref



```
1 <profileDesc>
2   <particDesc>
3     <listPerson>
4       <person xml:id="RC">
5         <persName>
6           <forename>Robinson</forename>
7           <surname type="corrupted">Crusoe</surname>
8           <surname type="birth">Kreutznaer</surname>
9         </persName>
10        <birth when="1632">1632</birth>
11      </person>
12    </listPerson>
13  </particDesc>
14 </profileDesc>
```

and from whom I was called `<persName ref="#RC">Robinson Kreutznaer</persName>`; but, by the usual corruption of words in England, we are now called –nay we call ourselves and write our name– `<persName ref="#RC">Crusoe</persName>`; and so my companions always called me.

Attribute @ref as pointing mechanism

Entity defined within the same XML document

```
<persName ref="#RC">Robinson Kreutznaer</persName>
```

In some other place, referred to by means of an URI


```
<persName ref="http://www.example.com/fictionalcharacters.xml#RC">Robinson Kreutznaer</persName>
```


Reference an existing authority file

```
<persName ref="http://www.isni.org/0000000123478267">Daniel Defoe</persName>
```

Authority files: catalogs of names

{BnF Data About Data.bnf.fr | Contact | Français | English | Deutsch

Daniel Defoe (1661?-1731) 



Country : [Grande-Bretagne](#)

Language : [Anglais](#)

Gender : [Masculin](#)

Birth : [Londres, 1661](#)

Death : [Moorfields \(GB\), 26-04-1731](#)

Note : [Journaliste, fondateur du journal "The Review" \(en 1704\) et roi de James Foe, dont il change le nom en Defoe vers 1703](#)

Field : [Littératures](#)

Variants of the name : [Daniel De Foë \(1661?-1731\)](#)
[Daniel de Foë \(1661?-1731\)](#)
[Daniel Foe \(1661?-1731\)](#)
[John Avery \(1661-1731\)](#)
[George Carleton \(1661?-1731\)](#)
[Robert Drury \(1661?-1731\)](#)
[Charles Johnson \(1661-1731\)](#)
[Andrew Moreton \(1661?-1731\)](#)
[Alexander Ramkins \(1661?-1731\)](#)
[George Roberts \(1661?-1731\)](#)

ISNI : [ISNI 0000 0004 5875 3523](#)

- VIAF: Virtual International Authority
- ISNI: International Standard Name Identifier (ISO Standard)
- Most national libraries: Bibliothèque nationale de France;
Deutsche Nationalbibliothek
- Perseus Digital Library: Classic authors

(other way of) Referencing authority files

Identifier in the `<idno>` element

```
1 <person xml:id="S1783">
2   <persName>
3     <forename>Marie-Henri</forename>
4     <surname>Beyle</surname>
5     <addName type="pseudonym">Stendhal</addName>
6   </persName>
7   <birth when="1783-01-23">23-01-1783</birth>
8   <death when="1842-03-23">23-03-1842</death>
9   <occupation>writer</occupation>
10  <nationality>French</nationality>
11  <idno type="ISNI">0000000120981326</idno>
12 </person>
```

more info for person?

<birth>, <death>, <event>, <langKnowledge>, <faith>,
<nationality>, <sex>, <age>, <occupation>, <note>, ...

TEI PLACES

TEI distinction: `<placeName>` vs. `<place>`

Encoding of references

I was born in the year 1632, in the city of `<placeName>York</placeName>`, of a good family, though not of that `<rs>country</rs>`

Encoding of referenced entities

```
<profileDesc>
  <particDesc>
    <listPerson>...</listPerson>
  </particDesc>
  <settingDesc>
    <listPlace>
      <place xml:id="york">
        <placeName>York</placeName>
        <location>
          <!--lat, long-->
          <geo>53.95767,-1.082304</geo>
        </location>
        <idno type="GeoNames">2633352</idno>
      </place>
    </listPlace>
  </settingDesc>
</profileDesc>
```


Geo-political place names



En tibi, formosæ sub forma Europa puellæ
Vivida fecundos pandit ut illa sinus.

Ridens Italiam dextra Cimbrosq; sinistra
Obtinet, Hispanum fronte geritq; solum.

Pectore habet Gallos, Germanos corpore gestat
Ac pedibus Græcos, Sauromatasq; fouet.

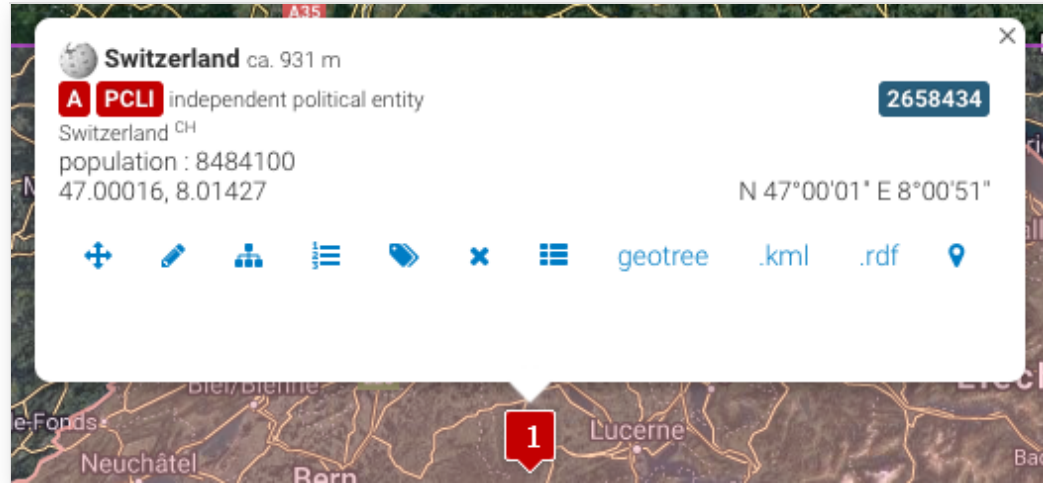
Components of place names

- <bloc> geo-political unit containing two or more nation states: **Europe**.
- <country> geo-political unite: **Italy**.
- <region> administrative unite: **Lazio**.
- <settelment> name of a settlement: **Rome**.
- <district> subdivision of a settlement: **Trastevere**.

Components of place names

- <location>
 - geographical coordinates
 - <geo>
 - place name components
 - an address
 - <address>

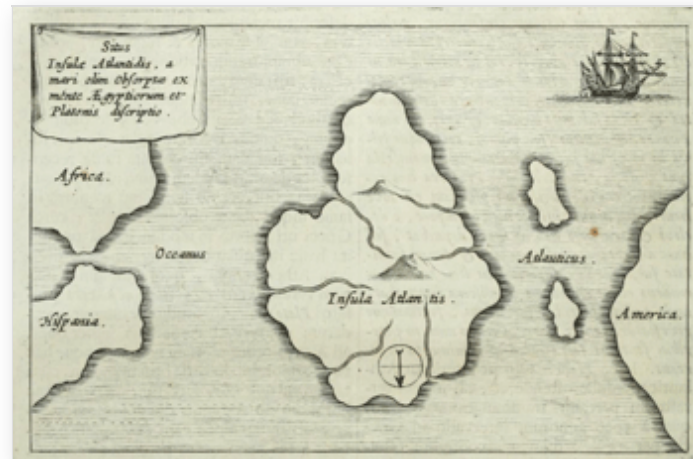
Gazetteers: catalogs of places



- GeoNames
- Pleiades (ancient places)
- Peripleo (hub of gazetteers)
- European Comision (batch geocoding)
- Geolocation via gazetteers APIs

Fictional places

```
<place type="fictional">  
<placeName>Atlantis</placeName>  
  <location>  
    <offset>fifty leagues beyond</offset>  
    <placeName>The Pillars of <persName>Hercules</persName></placeName>  
  </location>  
</place>
```



– Kircher, Mundus Subterraneus, 1665

TEI DATES

Core module: 3.5.4 Dates and Times

- `<date>` a date in any format
- `<time>` a phrase defining a time of day in any format

```
I was born in the year <date when="1632">1632</date>, in the city of  
<placeName ref="#york">York</placeName>, of a good family, though no  
t of that country
```

Attributes for date

To normalize elements that contain dates, times, etc. **in standard form (yyyy-mm-dd)**:

- **@when** value of the date or time.
- **@notBefore** the earliest possible date.
- **@notAfter** the latest possible date.
- **@from** indicates the starting point of a period.
- **@to** indicates the ending point of a period.

@when a useful way of normalizing or disambiguating

```
<birth when="1788-02-22">22 February 1788</birth>  
<birth when="1788-02-22">22 de febrero de 1788</birth>  
<birth when="1788-02-22">22 février 1788</birth>  
<birth when="1788-02-22">22/02/1788</birth>
```


References

- *A Gentle Introduction to XML, Guidelines for Electronic Text Encoding and Interchange*, TEI P5: Version 3.5.0.
- *digital.humanities @ Oxford*, *TEI @ Oxford Teaching Pages*.
- Lou Burnard (2014), *What is the Text Encoding Initiative? How to add intelligent markup to digital resources*, Marseille: OpenEdition Press.
- Christof Schöch (2014), *Digitale Textedition mit TEI*, DARIAH TEI-Tutorial.
- Martina Scholger (2017), *Names, dates, people and places*.

José Luis Losada Palenzuela, *Named Entities. Encoding Names,
Persons, Places and Dates*,
<<http://editio.github.io/presentations/named-entities.html>>,
2019.

