

# Metadata

James Cummings

2010-10



# The TEI Header

The TEI header was designed with two goals in mind

- needs of bibliographers and librarians trying to document 'electronic books'
- needs of text analysts trying to document 'coding practices' within digital resources

The result is that discussion of the header tends to be pulled in two directions...

## The Librarian's Header

- Conforms to standard bibliographic model, using similar terminology
- Organized as a single source of information for bibliographic description of a digital resource, with established mappings to other such records (e.g. MARC)
- Emerging code of best practice in its use, endorsed by major digital collections
- Pressure for greater and more exact constraints to improve precision of description: preference for structured data over loose prose

## Everyman's Header

- Gives a polite nod to common bibliographic practice, but has a far wider scope
- Supports a (potentially) huge range of very miscellaneous information, organized in fairly ad hoc ways
- Many different codes of practice in different user communities
- Unpredictable combinations of narrowly encoded documentation systems and loose prose descriptions

## TEI Header Structure

The TEI header has four main components:

- `<fileDesc>` (file description) contains a full bibliographic description of an electronic file.
- `<encodingDesc>` (encoding description) documents the relationship between an electronic text and the source or sources from which it was derived.
- `<profileDesc>` (text-profile description) provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting. (just about everything not covered in the other header elements)
- `<revisionDesc>` (revision description) summarizes the revision history for a file.

Only `<fileDesc>` is required; the others are optional.



## Example Header: Minimal required header

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>A title?</title>
    </titleStmt>
    <publicationStmt>
      <p>Who published?</p>
    </publicationStmt>
    <sourceDesc>
      <p>Where from?</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

## The TEI supports two 'levels' or types of header

- *corpus level* metadata sets default properties for everything in a corpus
- *text level* metadata sets specific properties for one component text of a corpus

# Corpus Header Example

```
<teiCorpus>
<teiHeader type="corpus">
<!-- corpus-level metadata here -->
</teiHeader>
<TEI>
<teiHeader type="text">
<!-- metadata specific to this text here -->
</teiHeader>
<text>
<!-- ... -->
</text>
</TEI>
<TEI>
<teiHeader type="text">
<!-- metadata specific to this text here -->
</teiHeader>
<text>
<!-- ... -->
</text>
</TEI>
</teiCorpus>
```



## Types of content in the TEI header

- free prose
  - prose description: series of paragraphs
  - phrase: character data, interspersed with phrase-level elements, but not paragraphs
- grouping elements: specialised elements recording some structured information
- declarations: Elements whose names end with the suffix Decl (e.g. `subjectDecl`, `refsDecl`) enclose information about specific encoding practices applied in the electronic text.
- descriptions: Elements whose names end with the suffix Desc (e.g. `<settingDesc>`, `<projectDesc>`) contain a prose description, possibly, but not necessarily, organised under some specific headings by suggested sub-elements.



# File Description

- has some mandatory parts:
  - `<titleStmt>`: provides a title for the resource and any associated statements of responsibility
  - `<sourceDesc>`: documents the sources from which the encoded text derives (if any)
  - `<publicationStmt>`: documents how the encoded text is published or distributed
- and some optional ones:
  - `<editionStmt>`: yes, electronic texts have editions too
  - `<seriesStmt>`: and they also fit into "series".
  - `<extent>`: how many floppy disks, gigabits, files?
  - `<notesStmt>`: notes of various types

# The File Description

- `<titleStmt>`: contains a mandatory `<title>` which identifies the electronic file (not its source!)
- optionally followed by additional titles, and by 'statements of responsibility', as appropriate, using `<author>`, `<editor>`, `<sponsor>`, `<funder>`, `<principal>` or the generic `<respStmt>`
- `<publicationStmt>`: may contain
  - plain text (e.g. to say the text is unpublished)
  - one or more `<publisher>`, `<distributor>`, `<authority>`, each followed by `<pubPlace>`, `<address>`, `<availability>`, `<idno>`

## A minimal header for Punch

```
<fileDesc>
  <titleStmt>
    <title>Punch, or the London Charivari: an electronic
      edition</title>
    <editor>Owen Seaman (1861-1936)</editor>
    <respStmt>
      <resp>TEI version</resp>
      <name>TEI@Oxford team</name>
    </respStmt>
  </titleStmt>
  <publicationStmt>
    <p>Unpublished</p>
  </publicationStmt>
  <sourceDesc>
    <p>Recorded from the Project Gutenberg versions</p>
  </sourceDesc>
</fileDesc>
```

## Title- and Responsibility- statements...

There may be many of them:

```
<title>Artamene</title>  
<title type="alt">Le Grand Cyrus</title>  
<title type="sub">Digital Edition</title>
```

Amongst the guilty parties:

```
<author>Scudery, Madeleine de</author>  
<principal>Geffin, Alexandre</principal>  
<funder>Fonds Nationale Suisse de la Recherche  
Scientifique</funder>  
<respStmt>  
  <resp>Encoding check</resp>  
  <name>Jean Untel</name>  
</respStmt>
```

## <publicationStmt> example

```
<publicationStmt>
  <publisher>TEI Consortium</publisher>
  <distributor>Oxford Text Archive</distributor>
  <idno type="ota">1256</idno>
  <availability>
    <p>Available under the terms of a Creative Commons
      Attribution and Share Alike licence.</p>
  </availability>
</publicationStmt>
```

## <notesStmt> example

<notesStmt> can contain notes on almost any aspect:

```
<notesStmt>  
  <note>Material prepared for the TEI@Oxford Summer  
    School.</note>  
</notesStmt>
```

## The Source Description

All electronic works need to indicate their source, even if it is just to say that it is 'born digital'. There are variety of ways to do this:

- prose description
- `<bibl>` : contains free text or any mixture of bibliographic elements such as `<author>`, `<publisher>` etc.
- `<biblStruct>` contains effectively the same elements but constrained in various ways according to bibliographic standards
- `<biblFull>` special-cases texts which were born TEI by replicating an embedded `<fileDesc>`
- A `<listBibl>` may be used for lists of such descriptions
- Specialised elements for spoken texts (`<recordingStmt>` etc.) and for manuscripts (`<msDesc>`) **Discussed later!**
- Authority lists for e.g people (`<listPerson>`) or places (`<listPlace>`) can be included.





## <sourceDesc> examples

```
<sourceDesc>
  <p>Born digital.</p>
</sourceDesc>
```

```
<sourceDesc>
  <bibl>
    <title level="a">Enigma</title>, <title level="j">Punch:
      or the London Charivari</title>, <date when="1914-07-01">July 1,
    1914</date>, 147, p. 6</bibl>
</sourceDesc>
```

## <bibl> vs. <biblStruct> Example

```
<bibl>  
  <title level="a">Enigma</title>, in <title level="j">Punch:  
    or the London Charivari</title> (July 1, 1914), vol 147,  
pp. 1-20  
</bibl>
```

```
<biblStruct>  
  <analytic>  
    <title level="a">Enigma</title>  
  </analytic>  
  <monogr>  
    <title level="j">Punch: or the London Charivari</title>  
    <imprint>  
      <pubPlace>London</pubPlace>  
      <date when="1914-07-01">July 1, 1914</date>  
      <biblScope type="vol">147</biblScope>  
      <biblScope type="pp">1-20</biblScope>  
    </imprint>  
  </monogr>  
</biblStruct>
```

## Encoding Description

`<encodingDesc>` groups notes about the procedures used when the text was encoded, either summarised in prose or within specific elements such as

- `<projectDesc>`: goals of the project
- `<samplingDecl>`: sampling principles
- `<editorialDecl>`: editorial principals, e.g. `<correction>`, `<normalization>`, `<quotation>`, `<hyphenation>`, `<segmentation>`, `<interpretation>`
- `<classDecl>`: classification system/s used
- `<tagsDecl>`: specifics about usage of particular elements

The `<encodingDesc>` can replace the user manual, or facilitate semi-automatic document management, given agreed codes of practice.

## <encodingDesc> Example (1)

```
<encodingDesc>
  <projectDesc>
    <p>The Imaginary Punch Project aims to ....
  </p>
</projectDesc>
<samplingDecl>
  <p>All pages containing editorial text have
    been transcribed in full. Pages containing only
    advertisements or illustrations have been
    omitted.</p>
</samplingDecl>
<editorialDecl>
  <hyphenation>
    <p>Original spelling has been retained, except
      that words hyphenated across line breaks have been
      silently re-assembled. The hyphen has been retained
      only where there exist cases of the same word being
      hyphenated in mid-line position. </p>
  </hyphenation>
<!-- ... -->
</editorialDecl>
<!-- ... -->
</encodingDesc>
```

## <encodingDesc> Example (2)

```
<encodingDesc>
<!-- ... -->
  <classDecl>
    <taxonomy xml:id="size">
      <category xml:id="large">
        <catDesc>story occupies more
          than half a page</catDesc>
      </category>
      <category xml:id="medium">
        <catDesc>story occupies between
          quarter and a half page</catDesc>
      </category>
      <category xml:id="small">
        <catDesc>story occupies less
          than a quarter page</catDesc>
      </category>
    </taxonomy>
    <taxonomy xml:id="topic">
      <category xml:id="politics-domestic">
        <catDesc>Refers to
          domestic political events</catDesc>
      </category>
      <category xml:id="politics-foreign">
        <catDesc>Refers to
          foreign political events</catDesc>
      </category>
    </taxonomy>
  </classDecl>
</encodingDesc>
```

## Profile Description

A collection of descriptions, categorised only as 'non-bibliographic'. Default members of the `model.profileDescPart` class include:

- `<creation>`: information about the origination of the intellectual content of the text, e.g. time and place
- `<langUsage>`: information about languages, registers, writing systems etc used in the text
- `<textDesc>` and `<textClass>`: classifications applied to the text by means of a list of specified criteria or by means of a collection of pointers, respectively
- `<particDesc>` and `<settingDesc>`: information about the 'participants', either real or depicted, in the text
- `<handNotes>`: information about the hands identified in a manuscript



## Language and character set usage

The `<langUsage>` element is provided to document usage of languages in the text. Languages are identified by their ISO codes:

```
<langUsage>
  <language ident="en">English</language>
  <language ident="fr">French</language>
  <language ident="bg-cy">Bulgarian in Cyrillic characters </language>
  <language ident="bg">Romanized Bulgarian</language>
</langUsage>
```

## Classification Methods

`<textClass>` provides a classification (by domain, medium, topic...) for the whole of a text expressed in one or more of the following ways:

using `<catRef>` direct reference to a locally defined (e.g. in the corpus header) category

using `<classCode>` reference to some commonly agreed and externally defined category (e.g. UDC)

using `<keywords>` assign arbitrary descriptive terms taken from a bibliographic controlled vocabulary or a tag cloud



## BNC Example

```
<profileDesc>
  <creation>
    <date when="1962"/>
  </creation>
  <textClass>
    <catRef
      target="#WRI #ALLTIM1 #ALLAVA2 #ALLTYP3 #WRIDOM5 #WRILEV2 #WRIMED1
#WRIPP5 #WRISAM3 #WRISTA2 #WRITAS0"/>
    <classCode scheme="DLEE">W nonAc: humanities
      arts</classCode>
    <keywords scheme="COPAC">
      <term>History, Modern - 19th century</term>
      <term>Capitalism - History - 19th century</term>
      <term>World, 1848-1875</term>
    </keywords>
  </textClass>
</profileDesc>
```

This categorization applies to the whole text. For more fine grained classification, use *@decls* on e.g. a `<div>` element.

## Revision Description

- A list of `<change>` elements, each with a `@date` and `@who` attributes, indicating significant stages in the evolution of a document.
- Most recent first.
- Can be maintained manually, but better done by means of a CMS (change management system)

```
<revisionDesc>
  <change>
    <date>$LastChangedDate: 2010-06-28 09:14:36 +0100 (Mon, 28
      Jun 2010) $.</date>
    <name>$LastChangedBy: lou $</name>
    <note>$LastChangedRevision: 10346 $</note>
  </change>
</revisionDesc>
```



# Manuscript Description

## Why are manuscripts special?

- Manuscripts are *unique objects*, often of great cultural or political value.
- Books, by contrast, exist in multiple copies, and can be described adequately by well-established and formalised bibliographic conventions.
- For manuscripts, there are several traditions, often descriptive or *belle lettriste*, and little consensus.

Similar concerns apply to other text-bearing objects.

## Components of a manuscript description

Within the `<msDesc>` element come a required `<msIdentifier>` element, which groups information identifying the manuscript, followed by an optional `<head>`, which can be used to provide in a brief, unstructured way information on the manuscript's contents etc. These are then followed either by one or more paragraphs (`<p>`), or one or more of the following specialised elements:

- `<msContents>`: an itemised list of the intellectual content of the manuscript, with transcriptions of rubrics, incipits, explicits etc, as well as primary bibliographic references
- `<physDesc>`: groups information concerning all physical aspects of the manuscript, its material, size, format, script, decoration, binding, marginalia etc.
- `<history>`: provides information on the history of the manuscript, its origin, provenance and acquisition by its holding institution



## Components of a manuscript description (cont.)

- `<additional>`: groups other information about the manuscript, in particular, administrative information relating to its availability, custodial history, surrogates etc.
- `<msPart>`: contains in essence a nested `<msDesc>`, in cases of composite manuscripts now regarded as constituting a single unit but made up of two or more parts which were originally physically distinct.

Within each of these elements a number of sub-elements is available; `<msContents>`, for example, will normally consist of one or more `<msItem>` elements, each in turn containing specific elements for `<rubric>`, `<incipit>`, `<explicit>` and `<colophon>`, as well as the standard TEI elements `<author>`, `<title>` and `<bibl>` for bibliographic references. As with `<msDescription>` itself, however, the contents of these first-level and second-level elements need not be this structured, since there is also the option of using paragraphs.



# Identification (1)

The `<msIdentifier>`

Traditional three part specification:

- place (`<country>`, `<region>`, `<settlement>`)
- repository (`<institution>`, `<repository>`)
- identifier (`<collection>`, `<idno>`)

```
<msIdentifier>
  <country>Canada</country>
  <settlement>Ottawa</settlement>
  <repository>Library and Archives Canada</repository>
  <collection>E.W.B. Morrison</collection>
  <idno>MG 30 E 81 v. 16</idno>
</msIdentifier>
```

## Identification (2)

Alternative or additional names can also be included:

```
<msIdentifier>
  <country>Danmark</country>
  <settlement>København</settlement>
  <repository> Det ArnamagnæanskeInstitut </repository>
  <idno>AM 45 fol.</idno>
  <msName xml:lang="la">Codex Frisianus</msName>
  <msName xml:lang="is">Fríssbók</msName>
</msIdentifier>
```

## Intellectual Content

- May simply use paragraphs of text...
- ... or a tree of `<msItem>` elements
- ... optionally preceded by a prose summary

We can describe the content in general terms:

```
<msContents>
  <p>An extraordinary charivari of heroic deeds and improving
    tales, including an early version of <title>Guy of
    Warwick</title> and several hymns.</p>
</msContents>
```

or we can provide detail about each distinct item:

```
<msContents>
  <summary>An extraordinary charivari of heroic deeds,
    improving tales, and hymns.</summary>
  <msItem>
<!-- details of Guy of Warwick here -->
  </msItem>
  <msItem>
<!-- other items here -->
  </msItem>
</msContents>
```



## Physical Description

An artificial (but helpful) grouping of many distinct items.

You can simply supply paragraphs of prose, covering such topics as

- `<objectDesc>`: the physical carrier
- `<handDesc>`: what is carried on it
- `<musicNotation>`, `<decoDesc>`, `<additions>`
- `<bindingDesc>` and `<sealDesc>`
- `<accMat>`: accompanying material

Or, group your discussion within the specific elements mentioned above.

Similarly, within the specific elements, you can supply paragraphs of prose, or further specific elements.

# The carrier 1

The `<objectDesc>` can contain just paragraphs, or `<supportDesc>` and `<layoutDesc>`

```
<objectDesc form="codex">
  <supportDesc material="mixed">
    <p>Early modern <material>parchment</material> and
    <material>paper</material>.</p>
  </supportDesc>
  <layoutDesc>
    <layout columns="1" ruledLines="25 32"/>
  </layoutDesc>
</objectDesc>
```

## The carrier 2

A more complex substructure with specific elements for `<support>`, `<extent>`, `<foliation>`, `<collation>`, `<condition>`. Multiple layouts may also be specified:

```
<layoutDesc>
  <layout ruledLines="25" columns="1">
    <p>
      <locus from="1r" to="202v"/>
      <locus from="210r" to="212v"/> Between 25 and 32 ruled
        lines.</p>
    </layout>
  <layout ruledLines="34 50" columns="1">
    <p>
      <locus from="203r" to="209v"/>Between 34 and 50 ruled
        lines.</p>
    </layout>
</layoutDesc>
```

## <handDesc> and <decoDesc>

- <handNote> (note on hand) describes a particular style or hand distinguished within a manuscript.
- <decoNote> contains a note describing either a decorative component of a manuscript or a fairly homogenous class of such components.

## <additions>

The `<additions>` element can be used to list or describe any additions to the manuscript, such as marginalia, scribblings, doodles, etc., which are considered to be of interest or importance.

```
<additions>
```

```
<p>The text of this manuscript is not interpolated with sentences from Royal decrees promulgated in 1294, 1305 and 1314. In the margins, however, another somewhat later scribe has added the relevant paragraphs of these decrees, see pp. 8, 24, 44, 47 etc.</p>
```

```
<p>As a humorous gesture the scribe in one opening of the manuscript, pp. 36 and 37, has prolonged the lower stems of one letter f and five letters þ and has them drizzle down the margin.</p>
```

```
</additions>
```

## <accMat>

<accMat> (accompanying material) contains details of any significant additional material which may be closely associated with the manuscript being described, such as non-contemporaneous documents or fragments bound in with the manuscript at some earlier historical period.

```
<accMat> A copy of a tax form from 1947 is included in the envelope with the letter. It is not catalogued separately.
</accMat>
```

## <history>

- <origin>: where it all began
- <provenance>: everything in between
- <acquisition>: how you acquired it

<origin> is *datable* element and thus has attributes *@notBefore* and *@notAfter*, *@when* etc.

## <history> Example

```

<history>
  <origin>
    <p>Written in <origPlace>England</origPlace> in the
    <origDate notAfter="1300" notBefore="1200">13th cent.
    </origDate>
  </p>
</origin>
<provenance>
  <p>On fol. 54v very faint is <q>Iste liber est fratris
    guillelmi de buria de <gap reason="illegible"/>
    Roberti ordinis fratrum Pred<ex>icatorum</ex>
  </q>,
    14th cent. (?): <q>hanauilla</q> is written at the foot
    of the page (15th cent.).</p>
</provenance>
<acquisition>
  <p>Bought from the Rev. <name type="person">W. D.
    Macray</name> on <date when="1863-03-17">March 17,
    1863</date>, for 1 pound 10s.</p>
</acquisition>
</history>

```



## And finally

A `<msDesc>` can contain `<msPart>`, essentially a nested `<msDesc>`, where originally distinct manuscripts or parts of a manuscripts have been brought together to form a composite manuscript.

```
<msDesc>
  <msIdentifier>
    <settlement>Amiens</settlement>
    <repository>Bibliothèque Municipale</repository>
    <idno>MS 3</idno>
    <msName>Maurdramus Bible</msName>
  </msIdentifier>
  <!-- other elements here -->
  <msPart>
    <altIdentifier>
      <idno>MS 6</idno>
    </altIdentifier>
    <!-- other information specific to this part here -->
  </msPart>
  <!-- other msParts here -->
</msDesc>
```

# Annotating names of people and places

# Names

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. 'Budleigh Salterton', 'Bouallebec', 'John Doe' ...
- `<persName>`, `<placeName>`, `<orgName>`: 'syntactic sugar' for `<name type="person">` etc.
- A rich set of proposals for the *components* of such nominal expressions, e.g. `<surname>`, `<forename>`, `<geogName>`, `<geogFeat>` etc.

# Entities

Before P5.1.0, the TEI was reluctant to stray into database territory, while recognising the need to distinguish clearly the encoding of references from the encoding of referenced entities. It now provides:

- `<person>` corresponding with `<persName>`
- `<place>` corresponding with `<placeName>`
- `<org>` corresponding with `<orgName>`
- and in addition `<relation>`, `<event>` ...

## Why?

- to facilitate a more detailed and explicit encoding source documents (historical materials for example) which are primarily of interest because they concern objects in the real world
- to support the encoding of "data-centric" documents, such as authority files, biographical or geographical dictionaries and gazeteers etc.
- to represent and model in a uniform way data which is only implicit in readings of many different documents

## Reference theory

*Reference* is a fundamental semiotic concept

- We can talk about the real world using natural languages because we know that some types of word are closely associated with real, specific, objects
- Proper names and technical terms are canonical examples of this kind of word
- ‘Martin Luther King’ refers to a single real world entity; ‘Lyon’ and ‘River Thames’ to others: a specific place, a specific river respectively
- When we translate between natural languages, usually the proper names don’t change, or are conventionally equivalent

## How do we represent this association?

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.

as well as the attributes

*@role* may be used to specify further information about the entity referenced by this name, for example the occupation of a person, or the status of a place.

*@nymRef* provides a means of locating the canonical form (*nym*) of the names associated with the object named by the element bearing it.

Arguably, *@key* is redundant, since *@ref* is defined as anyURI

# Examples

```
<p>... <name ref="#jsbach" type="person">Johann Sebastian Bach
</name> the German composer was born in 1685... </p>
```

```
<p>... <name ref="grove:jsbach" type="person">Johann Sebastian
  Bach </name>the German composer was born in 1685...
</p>
```

```
<p>... <name role="composer">Engelbert Humperdink</name> was
born in 1854... </p>
<p>... <name role="singer">Engelbert Humperdink</name> was born
in 1936... </p>
```



## References take many forms

Even within a single language, in a single document, there may be many ways of referencing the same person:

```
...<persName>Clara  
Schumann</persName>.... <persName>Clara</persName>  
....  
<persName>Frau Schumann</persName>
```

The *@key* or *@ref* can be used simply to combine all references to a specified person:

```
.... <persName ref="#CS">Clara Schumann</persName>....  
<persName ref="#CS">Clara</persName> ....  
<persName ref="#CS">Frau  
Schumann</persName>  
<!-- ... elsewhere -->  
<person xml:id="CS">  
  <persName xml:lang="de">  
<!-- everything we want to say about this lady -->  
    </persName>  
</person>
```

## References are also 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="NN123">  
  <placeName notBefore="1400">Nancy</placeName>  
  <placeName notAfter="0056">Nantium</placeName>  
<!-- ... -->  
</place>
```

## Components of <persName> elements

```
<persName ref="#jsbach" xml:lang="de">
  <forename type="first">Johann</forename>
  <forename type="middle">Sebastian</forename>
  <surname>Bach</surname>
</persName>
<persName ref="#jsbach" xml:lang="fr">
  <forename type="composé">Jean-Sébastien</forename>
  <surname>Bach</surname>
</persName>
```

Not to mention... <roleName> (e.g. 'Emperor'), <genName> (eg 'the Elder') <addName> (e.g. 'Hammer of the Scots'), <nameLink> a link between components (e.g. 'van der') ...

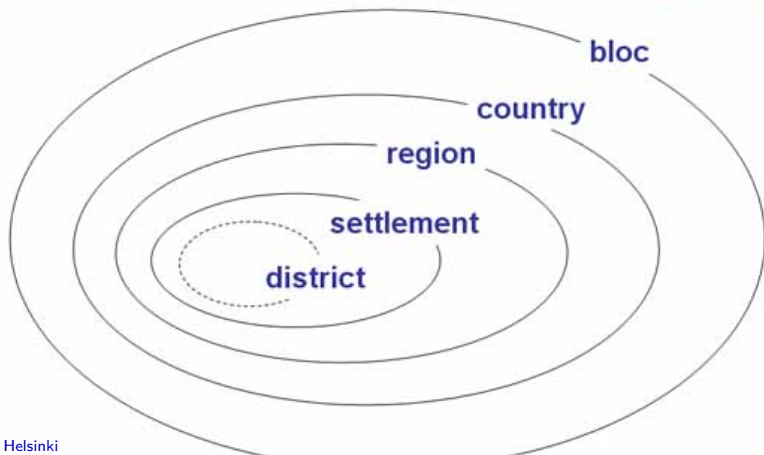
## Components of place names

- `<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>
```

Place names generally fall into a kind of hierarchy

## Geo-political Place names



# What can we say about named entities?

Potentially, quite a lot...

```
<person xml:id="VM1893">
  <persName xml:lang="ru">Владимир Владимирович
    Маяковский</persName>
  <persName xml:lang="fr">Wladimir Maïakowski</persName>
  <birth when="1893-07-19">7 July (OS) 1893, <placeName ref="#BGDT" xml:lang="en">Baghdati,
    Georgia</placeName>
  </birth>
  <death when="1930-04-14"/>
  <occupation>Poet and playwright, among the foremost
    representatives of early-20th century Russian
    Futurism.</occupation>
</person>
```

What elements should the TEI provide for such a purposes?

## Traits, States, and Events

As elsewhere in the TEI, we resolve this question by adding a layer of abstraction. We distinguish three *classes* of information:

- **<trait>**: a more or less intrinsic property of the entity, which (usually) does not change over time (e.g. eye colour for a person, location for a place)
- **<state>**: a property which applies to the entity for some specific extent of time (e.g. occupation for a person, population for a place)
- **<event>**: an independent event in the real world which may lead to a change in state or trait (e.g. birth for a person, a war for a place)

We define elements for these prototypes, together with more specific elements, make up the associated element classes. The more specific elements aim to meet most common needs

Additionally, all these elements are members of the 'datable' class (see below)



## Traits

Some typical traits of a person

- `<faith>`: faith, belief system, religion etc. of a person
- `<langKnowledge>`: linguistic knowledge of a person
- `<nationality>`: nationality (socio-politico status)
- `<sex>`: sex
- `<socecStatus>`: socio-economic status

Some typical traits of a place:

- `<climate>`: describes the climate
- `<location>`: describes where a place is (see later)
- `<population>`: describes its population
- `<terrain>`: describes its terrain

Some of these (e.g. `sex`) have normalised attributes, but mostly they contain free text descriptions.



# States

Some typical states for a person

- `<occupation>` an informal description of a person's trade, profession or occupation
- `<residence>` (residence) a person's present or past places of residence
- `<affiliation>` an informal description of a person's present or past affiliation with some organization
- `<education>` a description of the educational experience of a person
- `<floruit>` contains information about a person's period of activity

## 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 mentioned earlier (a politico-geographical location)
- a set of geographical co-ordinates

```
<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>
```



# A place can be fictional

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

## Places can self-nest

```
<place xml:id="LT">
  <country>Lithuania</country>
  <country xml:lang="lt">Lietuva</country>
  <place xml:id="LT-VN">
    <settlement>Vilnius</settlement>
  </place>
  <place xml:id="LT-KA">
    <settlement>Kaunas</settlement>
  </place>
</place>
```

## Events

For persons, only two specific event elements are defined: `<birth>` and `<death>`. Anything else must be defined using the generic `<event>` element and its `@type` attribute.

```
<person xml:id="rwagner">
  <persName>
    <forename>Richard</forename>
    <surname>Wagner</surname>
  </persName>
  <birth when="1813-05-22"/>
  <event type="marriage" when="1836-11-24">
    <desc>On 24 November 1836, Wagner and <persName ref="#MINPLAN">Minna
Planer</persName> were married.
    </desc>
  </event>
  <event type="move" notBefore="1836-11-24">
    <desc>They moved to the town of
    <placeName>Riga</placeName>, at that time part of
    <bloc>the Russian empire</bloc>.</desc>
  </event>
</person>
```



## W3C Date Formats

Any datable element 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' by default is that defined by W3C; a parallel set of attributes is provided for ISO format dates.

All dates are normalised to the Gregorian calendar.

## Personal Relationships

The `<relation>` (relationship) element describes any kind of relationship or linkage amongst other entities

We distinguish 'mutual' relationships (e.g. sibling) from non-mutual or directed relationships (e.g. parent-of).

The following attributes are available:

- `@name` supplies a name for the kind of relationship of which this is an instance
- `@active` identifies the 'active' participants in a non-mutual relationship, or all the participants in a mutual one
- `@mutual` supplies a list of participants amongst all of whom the relationship holds equally
- `@passive` identifies the 'passive' participants in a non-mutual relationship

# Example

```
<person xml:id="jsbach">
  <persName>Johann Sebastian Bach</persName>
</person>
<person xml:id="cdbach">
  <persName>Catharina Dorothea Bach</persName>
</person>
<person xml:id="ghbach">
  <persName>Gottfried Heinrich Bach</persName>
</person>
<!--...-->
<relationGrp type="children" subtype="first-marriage">
  <relation name="parent" active="#jsbach" passive="#cdbach"/>
<!--...-->
</relationGrp>
<relationGrp type="children" subtype="second-marriage">
  <relation name="parent" active="#jsbach" passive="#ghbach"/>
<!--...-->
</relationGrp>
```



# Nyms

The elements `<listNym>` and `<nym>` are used to document the canonical form of a name or name-component.

- `<nym>`
  - can contain `model.entryParts` (e.g. `<form>`, `<orth>`, `<etym>`) and may also include a number of other `<nym>`s
  - in addition to global attributes and `att.typed`, it includes the attribute `@parts` to point to constituent `<nym>`s
- `<listNym>` a list of canonical names
- `@nymRef` has been added to the attribute class `att.naming` to refer to the canonical name

# Example

```
<nym xml:id="J45">
  <form xml:lang="la">Iohannes</form>
  <nym xml:id="J450">
    <form xml:lang="en">John</form>
    <nym xml:id="J4501">
      <form>Johnny</form>
    </nym>
    <nym xml:id="J4502">
      <form>Jon</form>
    </nym>
  </nym>
  <nym xml:id="J455">
    <form xml:lang="ru">Ivan</form>
  </nym>
  <nym xml:id="J453">
    <form xml:lang="fr">Jean</form>
  </nym>
</nym>
```

## Optional Exercise: Adding metadata

The second exercise instructs you on how to add a complete manuscript description to the transcription we completed in the first exercise. If you didn't finish it the first exercise you can continue it now. If you have no interest in manuscript description, perhaps just give the exercise a quick read through. If we've run out of time because I've talked too much, then you can do this exercise in your own time if you wish.