

## Talk 3: The teiHeader; Names/People/Places

James Cummings

19 September 2013

## The <teiHeader>

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...

(全てのTEI準拠テキストが伴う、電子版のタイトルページを構成する、記述的・宣言的情報を示す)

## The Librarian's Header

- Conforms to standard bibliographic models, 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, EAD, etc. or embedding in containers like METS)
- General consensus available in the 'Best Practice for TEI for Libraries' from the TEI-LIB SIG
- Pressure for greater and more exact constraints to improve precision of description: preference for structured data over loose prose

## Everyone else'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 or individualistic 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> (電子ファイルに関する完全な書誌情報を示す) contains a full bibliographic description of an electronic file.
- <encodingDesc> (電子テキストとその元資料との関係を示す) documents the relationship between an electronic text and the source or sources from which it was derived.
- <profileDesc> (書誌情動的ではない詳細な解説を示す。例えば、言語や特殊言語、生成されたときの状況、参加者など) provides a detailed description of non-bibliographic aspects of a text. (just about everything not covered in the other header elements)
- <revisionDesc> (ファイルの改訂履歴を示す) 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 xmlns="http://www.tei-c.org/ns/1.0">
<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.

# A minimal header for 'Strange Meeting'

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>My 'Strange Meeting' document</title>
    </titleStmt>
    <publicationStmt>
      <p>An exercise for learning TEI.</p>
    </publicationStmt>
    <sourceDesc>
      <p>The primary resource of this file is
        <ref
          target="http://www.oucs.ox.ac.uk/ww1lit/collections/item/3350">
          Strange Meeting</ref> from Jon Stallworthy's edition, available
on the First World War Poetry Digital Archive. </p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

# File Description <fileDesc>

- has some **mandatory** parts:
  - <titleStmt> (作品や知的内容に責任のあるもののタイトルに関する情報をまとめる)
  - <sourceDesc> (電子テキストが作られた元テキストの情報を示す)
  - <publicationStmt> (電子テキストなどの出版や頒布に関する情報をまとめる)
- and some **optional** ones:
  - <editionStmt> (版に関する情報をまとめる)
  - <seriesStmt> (出版されたシリーズの情報をまとめる)
  - <extent> (電子・非電子テキストのおよその大きさを任意の単位で示す)
  - <notesStmt>  
(当該書誌情報の他の場所に記録されているテキストに関する注釈をまとめる)

## 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
  - paragraph(s) of prose
  - one or more <publisher>, <distributor>, <authority>, each followed by <pubPlace>, <address>, <availability>, <idno>

## Title and Responsibility statements...

You can duplicate any of these elements as necessary, and document additional responsibilities with a generic <respStmt>

(著者や編集者など特定の役割を示す要素が充分ではない場合に、テキスト、版、記録などの知的内容に関する責任を示す):

```
<titleStmt>
  <title>Letter to Leslie Gunston</title>
  <author>Wilfred Owen</author>
  <editor>Renée van Baalen</editor>
  <principal>James Cummings</principal>
  <meeting>Introduction to TEI Course</meeting>
  <respStmt>
    <resp>Improved encoding</resp>
    <name>James Cummings</name>
  </respStmt>
</titleStmt>
```

## <editionStmt> and <extent> example

<editionStmt> can be used to document the details of this particular edition:

```
<editionStmt>  
  <edition>First Edition</edition>  
</editionStmt>  
<extent>6.5 kb</extent>
```

<extent> is sometimes used to document number of words in a corpus.

## &lt;publicationStmt&gt; example

```
<publicationStmt>
  <publisher>TEI @ Oxford</publisher>
  <distributor>IT Services, University of Oxford</distributor>
  <authority>James Cummings</authority>
  <pubPlace>
    <address>
      <orgName>IT Services, University of Oxford</orgName>
      <street>13 Banbury Road</street>
      <settlement>Oxford</settlement>
      <postCode>OX2 6NN</postCode>
      <country>United Kingdom</country>
    </address>
  </pubPlace>
  <date when="2013-09-19">19 September 2013</date>
  <idno>dhoxxs-tei-ex03</idno>
  <availability>
    <licence>Licensed with a <ref
      target="http://creativecommons.org/licenses/by/3.0/">Creative
      Commons Attribution</ref> licence.</licence>
  </availability>
</publicationStmt>
```

## <notesStmt> example

<notesStmt> can contain notes on almost any aspect of the file or its contents:

```
<notesStmt>  
  <note>Transcribed for TEI Workshop</note>  
</notesStmt>
```

These notes can be short statements, or many paragraphs long.



## The Source Description

All electronic works need to document their source, **even** 'born digital' ones! There are variety of ways to do so:

- prose description, just a <p>
- <bibl> (厳密でない構造を持つ書誌情報の引用を含む)
- <biblStruct> (構造を持った書誌情報を示す)
- <biblFull> (厳密な構造を持つ書誌情報を示す)
- A <listBibl> (書誌項目引用のリストを示す)
- Specialised elements: <recordingStmt>  
(発話テキストの転記の元になる録音, 録画されたものを示す) & <msDesc>  
(単一の識別可能な手書き資料の解説を示す)
- Authority lists: <listPerson> (特定可能な個人やグループに関する情報のリストを示す),  
<listPlace> (場所のリストを示す), <listOrg>  
(特定可能な団体に関する情報を示す解説のリストを示す)

## &lt;sourceDesc&gt; example

```
<sourceDesc>
  <biblStruct>
    <analytic>
      <title>Letter to Leslie Gunston</title>
      <author>Wilfred Owen</author>
    </analytic>
    <monogr>
      <title>The Wilfred Owen Collection</title>
      <ref
        target="http://www.oucs.ox.ac.uk/ww1lit/
collections/document/5243/4769"> First World War Poetry Digital Archive</ref>
      <imprint>
        <publisher>The First World War Poetry Digital Archive</publisher>
        <pubPlace>Oxford</pubPlace>
        <biblScope type="pp" n="2">Two pages</biblScope>
      </imprint>
    </monogr>
    <relatedItem>
      <bibl>The source of this digital resource is a copy from the
<istributor>Harry Ransom Centre</istributor>.</bibl>
    </relatedItem>
  </biblStruct>
</sourceDesc>
```

## More about <msDesc>

### Why are manuscripts special?

- Manuscripts are *unique objects*, sometimes of great cultural or political value.
- Books, by contrast, exist in multiple copies, and we describe them with basic but well-established and formalised bibliographic conventions.
- For manuscripts, there are several traditions, often descriptive with little consensus internationally.
- Manuscripts can be ancient or modern, and it benefits everyone to use the same forms of description for all types of manuscripts.

Similar concerns apply to other text-bearing objects, and <msDesc> is often also used for incunabula and inscriptions.

## Components of a manuscript description

Inside <msDesc> (単一の識別可能な手書き資料の解説を示す) only <msIdentifier> (解説されている手書き資料を特定するために必要な情報を示す) is required. After this <head> is sometimes used for a brief summary. These are then followed either by one or more paragraphs (<p>), or one or more of the following specialised elements:

- <msContents> (手書き資料の知的内容を、一連の段落または一連の構造化された手書き資料項目として示す)
- <physDesc> (手書き資料の物理的な解説を示す)
- <history> (手書き資料の歴史を表す要素をまとめる)

## Components of a manuscript description (cont.)

- <additional> (当該手書き資料に関する書誌情報などの付随情報まとめる)
- <msPart>  
(元々は別のものであったものを、新しくひとつにまとめ直された手書き資料の情報を示す)

## Inside the top level

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 <msDesc> itself, however, the contents of these first-level and second-level elements need not be this structured, since there is also the option of just using paragraphs (<p>).

## Identification (1)

The <msIdentifier> (解説されている手書き資料を特定するために必要な情報を示す)

Traditional three part specification:

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

```
<msIdentifier>
  <country>United States of America</country>
  <region>Texas</region>
  <settlement>Austin</settlement>
  <institution> The University of Texas at Austin </institution>
  <repository>Harry Ransom Centre</repository>
  <collection>Wilfred Owen Collected Letters</collection>
  <idno type="folio">ff504</idno>
  <altIdentifier>
    <idno>Letter no. 535 Ed. 'Wilfred Owen Collected Letters'</idno>
  </altIdentifier>
  <msName>Letter to Leslie Gunston</msName>
</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 set of `<msItem>` (手書き資料の知的内容の中に出現する、独立した作品または項目を示す) elements, optionally preceded by a prose summary. Unstructured:

```
<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 structured:

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

The <objectDesc> can contain just paragraphs, or <supportDesc> (手書き資料のテキスト部分を作る物理的な素材を示す要素をまとめる) and <layoutDesc> (手書き資料のレイアウト情報をまとめる)

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

## Another <physDesc> example

```
<physDesc>
  <objectDesc form="folio">
    <supportDesc material="paper">
      <support>A single folio of <material>paper</material> in
the collection as ff504 recto and verso</support>
    </supportDesc>
    <layoutDesc>
      <layout columns="1" writtenLines="20">Written full width
as a single column, with approximately 20 lines per
page</layout>
    </layoutDesc>
  </objectDesc>
  <handDesc hands="1">
    <handNote>Written in <persName ref="#WO">Wilfred
Owen's</persName> hand.</handNote>
  </handDesc>
</physDesc>
```

## &lt;history&gt;

- <origin>: where it all began (手書き資料の出自に関する解説や情報を示す)
- <provenance>: everything in between  
(手書き資料を入手するまでの歴史に関する、特定可能なエピソードに関する情報を示す)
- <acquisition>: how you acquired it  
(当該手書き資料が入手された経緯についての情報を示す)

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

## &lt;history&gt; 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>

```

## <additional> information

- <adminInfo> (当該手書き資料, またはその記録文書そのものの管理・利用形態についての情報を示す)
- <surrogates> (手書き資料が電子化されたものまたは写真の表現に関する情報を示す)
- <accMat> (当該手書き資料と関連する情報の詳細を記述するための要素)
- <listBibl> (書誌項目引用のリストを示す)

## Administrative information

- record history
- availability
- custodial history
- miscellaneous remarks

```
<adminInfo>
  <custodialHist>
    <custEvent type="conservation" notBefore="1961-03" notAfter="1963-02">
      <p>Conserved between March 1961 and February 1963 at Birgitte Dalls
      Konserveringsværksted.</p>
    </custEvent>
    <custEvent type="photography" notBefore="1988-05-01" notAfter="1988-05-
    30">
      <p>Photographed in May 1988 by AMI/FA.</p>
    </custEvent>
  </custodialHist>
</adminInfo>
```



## 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. (元々は別のものであったものを、新しくひとつにまとめ直された手書き資料の情報を示す)

```
<msDesc>
  <msIdentifier>
    <settlement>Amiens</settlement>
    <repository>Bibliothèque Municipale</repository>
    <idno>MS 3</idno>
    <msName>Maurdramnus 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>
```

## Association between header and text

By default everything asserted by a header is true of the text to which it is prefixed. This can be over-ridden:

- as when a text header over-rides or amplifies a corpus-header setting
- when model.declarable elements are selected by means of the *@decls* attribute (available on all model.declaring elements)
- using special purpose selection/definition elements e.g. <catRef> and <taxonomy>

Most components of the encoding description are declarable.

# Encoding Description

<encodingDesc> (電子テキストとその元資料との関係を示す) including:

- <projectDesc> (制作過程に関する情報も含めて、電子ファイルが作られた目的の詳細を示す)
- <samplingDecl>  
(コーパス等を作成する際、テキストを標本化する原理や手法に関する、散文による解説を含む)
- <editorialDecl> (テキストを符号化する際に適用される編集方針や編集方法の詳細を示す),  
with: <correction>, <normalization>, <quotation>,  
<hyphenation>, <segmentation>, <interpretation>
- <classDecl> (当該テキスト中で使用されている分類コードを定義する、ひとつ以上の分類法を示す)
- <tagsDecl> (タグ付けに関する詳細な情報を示す)

## <encodingDesc> Example

```
<encodingDesc>
  <projectDesc>
    <p>The TEI@Oxford project created these teaching materials for teaching TEI
workshops.</p>
  </projectDesc>
  <editorialDecl>
    <correction>
      <p>Apparent errors have been marked as <gi>sic</gi> but correct
readings not provided.</p>
    </correction>
    <hyphenation>
      <p>Hyphens have been transcribed as they appear.</p>
    </hyphenation>
  </editorialDecl>
</encodingDesc>
```

## <rendition> element

- <rendition> (元資料テキスト中にある、ひとつ以上の要素の描出や現れ方に関する情報を示す)

```
<tagsDecl>
  <rendition xml:id="r-center" scheme="css">text-align: center;</rendition>
  <rendition xml:id="r-small" scheme="css">font-size: small;</rendition>
  <rendition xml:id="r-large" scheme="css">font-size: large;</rendition>
</tagsDecl>
```

which you can easily point to from the text:

```
<hi rendition="#r-center #r-large">this bit of text was large and centred</hi>
```

but compare:

```
<hi rend="large center">this bit of text was large and centred</hi>
```

## <applInfo> element

<applInfo> (TEIファイルを編集したソフトウェアに関する情報を示す)

```
<applInfo>
  <application version="1.8.2.2" ident="IMT" notAfter="2012-
06-01">
    <label>Image Markup Tool</label>
    <ptr target="#P1"/>
    <ptr target="#P2"/>
  </application>
</applInfo>
```

# Profile Description

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

- <creation> (テキストの作成に関する情報を示す)
- <langUsage> (テキスト中にある言語, 特殊言語, 社会的方言, 方言などを示す)
- <textDesc> (状況パラメータにより, テキストの情報を示す) and <textClass> (標準的な分類スキーム, 分類語彙などにより, テキストの性格や話題を示す 情報をまとめる)
- <particDesc> (言語交流における, 特定可能な発話者, 声, その他の参加者を示す) and <settingDesc> (言語交流が行われた状況設定を, 散文または一連の要素settingにより示す)
- <handNotes> (元資料にある特定可能な筆致を記録する, ひとつ以上の要素handNote (筆致) を示す)

## <creation> example

```
<creation>
  <date when="1918-05"/>
  <placeName>Ripon</placeName>
  <listChange ordered="true">
    <change xml:id="CHG-1">First stage, written in pencil in Owen's hand
  </change>
    <change xml:id="CHG-2">Second stage, revised in pencil in Owen's
hand</change>
    <change xml:id="CHG-3">Fixation of the revised passages and further minor
revisions by Owen using ink</change>
    <change xml:id="CHG-4">Addition of another stanza with a different ink,
probably at a later stage</change>
  </listChange>
</creation>
```

Here <listChange> records stages in changes to the document. Further down, in <revisionDesc> the same element is used to record changes to the electronic file.



## Language and character set usage

The <langUsage> element is provided to document usage of languages and writing systems 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>
```

## Detailed characterization of a text

<textDesc> (状況パラメータにより、テキストの情報を示す)

```
<textDesc n="novel">
  <channel mode="w">print; part issues</channel>
  <constitution type="single"/>
  <derivation type="original"/>
  <domain type="art"/>
  <factuality type="fiction"/>
  <interaction type="none"/>
  <preparedness type="prepared"/>
  <purpose type="entertain" degree="high"/>
  <purpose type="inform" degree="medium"/>
</textDesc>
```

## Revision Description

- `<revisionDesc>` (ファイルの改訂履歴を示す)
- A list of `<change>` (研究者間で共有されている電子テキストの特定の版に対して施された変更や修正を示す) elements, each with a `@date` and `@who` attributes, indicating significant stages in the evolution of a document. Most recent first.
- Can be grouped into `<listChange>` elements. Used here it is about the electronic file, used in `<creation>` it is about the document.

```

<revisionDesc>
  <listChange>
    <change when="2012-07-03">
      <persName>James Cummings</persName> improved the header.</change>
    <change when="2012-02">
      <persName>Renée van Baalen</persName> transcribed the <title>Letter to
      Leslie Gunston</title> document. </change>
    </listChange>
  </revisionDesc>

```

# Names, People, and Places

Instances of names are distinct from the entities which they reference. One entity (person, place, organisation) might be known by many names. We get these in our TEI schema by including the **namesdates** module.

# Names

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

- `<rs>` (一般的な意味での名前や参照文字列) e.g. 'the girl you mentioned', 'my husband'...
- `<name>` (固有名詞) e.g. 'Siegfried Sassoon', 'Kyoto', 'Sony' ...
- Specific forms: `<persName>`, `<placeName>`, `<orgName>`
- Components: `<surname>`, `<forename>`, `<geogName>`, `<geogFeat>` etc.

# Named Entities

- <person> (特定可能な個人の情報を示す。  
例えば、言語交流の参加者、歴史資料中で参照される人物など) corresponding with <persName>
- <place> (地理上の場所のデータを示す) corresponding with <placeName>
- <org> (特定可能な団体の情報を示す。例えば、会社、集団など、人の集まり)  
corresponding with <orgName>
- and in addition <relation>, <event> etc.

## 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
- ‘Wilfred Owen’ refers to a single real world entity; ‘Tokyo’ and ‘Mount Fuji’ to others: a specific place, a specific river respectively
- When we translate between natural languages, often the proper names don’t change, or are conventionally equivalent

## Name 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 Humperdinck</name> was born in 1854... </p>
```

```
<p>... <name role="singer">Engelbert Humperdinck</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>Leslie Gunston</persName>....
<persName>Leslie</persName> ....
<rs>Wilfred's cousin</rs>
```

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

```
....
<persName ref="#LG">Leslie Gunston</persName>....
<persName ref="#LG">Leslie</persName> ....

<rs ref="#LG">Wilfred's cousin</rs>
<!-- ... elsewhere -->
<person xml:id="LG">
  <persName>Leslie Gunston</persName>
  <!-- everything we want to say about Leslie -->
</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

```
<p>
  <persName>
    <forename>Wilfred</forename>
    <forename>Edward</forename>
    <forename>Salter</forename>
    <surname>Owen</surname>
  </persName> did not know
  <persName ref="#jsbach" xml:lang="fr">
    <forename type="composé">Jean-Sébastien</forename>
    <surname>Bach</surname>
  </persName>
</p>
```

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') ...

## Components of place names

- <placeName> (絶対的, 相対的場所名を示す)
- <geogName> (ウィンドラッシュ峡谷, シナイ山などの地理的特性に関する名前)
- <geogFeat> (地理上の名前, 例えば, 谷, 山などに含まれている地物を特定する名前を示す)

```
<placeName>  
  <geogFeat>Mount</geogFeat>  
  <geogName>Fuji</geogName>  
</placeName>
```

# Named Entity Metadata

```
<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</occupation>
  <note>Among the foremost representatives of early-20th century Russian
Futurism.</note>
</person>
```

## Another basic &lt;person&gt;

```
<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>
```

## Traits, States, and Events

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

- <state>: more general-purpose, but usually a time-related property (人物, 場所, 組織などの, 現在の社会的状態や地位を示す)
- <trait>: if you want to a distinguish between time-bound and static, use this for properties that (usually) don't change over time (人物や場所の文化的な特性を示す)
- <event>: an independent event in the real world which may lead to a change in state or trait (人生における特定の重要事象の解説を示す)

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

# Traits of a Person

## Some typical traits of a person

- <faith> (人物の信仰・宗教を示す)
- <langKnowledge> (個人の言語学的知識を散文または要素 langKnown のリストでまとめる)
- <nationality> (人物の国籍や市民権の形式的でない解説を示す)
- <sex> (人物の性別を示す)
- <socecStatus> (ある個人の、社会的・経済的状态を示す、形式的でない解説を示す)



# Traits of a Place

Some typical traits of a place:

- <climate> (ある場所の天候に関する情報を示す)
- <location> (地理上の座標集合, 地政学上の名前付き実体, アドレスなどにより場所を定義する)
- <population> (ある場所の人口に関する情報を示す)
- <terrain> (場所の地形情報を示す)

# States

## Some typical states for a person

- <occupation> (人物の仕事や職業の、形式的でない説明を示す)
- <residence> (人物の、現在または過去の住居を示す)
- <affiliation> (人物が所属している、またはしていた団体に関する情報を示す。  
例えば、雇い主や出資者など)
- <education> (人物の学歴を示す)
- <floruit> (人物が活躍した時期に関する情報を示す)

## 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 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>
```

## 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 type="soverignState">
  <placeName>United Kingdom</placeName>
  <placeName type="full">United Kingdom of Great Britain and
    Northern Ireland</placeName>
  <place type="country">
    <placeName>Scotland</placeName>
    <place xml:id="edinburgh" type="city">
      <placeName>Edinburgh</placeName>
      <place xml:id="craiglockhart">
        <placeName>Craiglockhart War Hospital</placeName>
        <location>
          <geo>55.91812, -3.24019</geo>
        </location>
      </place>
    </place>
  </place>
</place>
</place>
```

## Organizational names

Organizations have names as well. These are 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>
```

## Events

For persons, only two specific event elements are defined: <birth> (人物の出生に関する情報を示す。例えば、日時や場所など) and <death> (人物の死亡に関する情報を示す。例えば、日付や場所など). Anything else must be defined using the generic <event> (人生における特定の重要事象の解説を示す) element and its @type attribute.

```
<person xml:id="SS">
  <persName>Siegfried Loraine Sassoon</persName>
  <birth when="1886-09-08">
    <placeName>
      <placeName>Weirleigh Mansion</placeName>
      <settlement>Matfield</settlement>
      <region>Kent</region>
    </placeName>
  </birth>
  <death when="1967-09-01"/>
  <event when="1914-08-04" type="military">
    <desc>In service with Sussex Yeomanry on the day the United
      Kingdom declared war</desc>
  </event>
  <event when="1933-12" type="marriage">
    <desc>Married Hester Gatty in December 1933</desc>
  </event>
  <event when="1945" type="separation">
    <desc>Seperated from his wife in 1945</desc>
  </event>
</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* 日付や時間を、標準形式で示す

*@notBefore* 当該事象の一番古い日付を、標準形式で示す

*@notAfter* 当該事象の一番新しい日付を、標準形式で示す

*@from* 標準形で当該時間幅の始点を示す

*@to* 標準形で当該時間幅の終点を示す

The 'standard form' is that defined by W3C. All dates are normalised to the Gregorian calendar, but there are ways to use other calendars.



## Example of <relation>

<relation> (指定された参加者グループ間の関連性の種類を示す)

```
<person xml:id="SS">
  <persName>Siegfried Loraine Sassoon</persName>
</person>
<person xml:id="HG">
  <persName>Hester Gatty</persName>
</person>
<person xml:id="GS">
  <persName>George Sassoon</persName>
</person>
<!--...-->
<relationGrp type="children">
  <relation name="parent" active="#SS" passive="#GS"/>
<!--...-->
</relationGrp>
```

# Nyms

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

- <nym> (各種名前の、標準的な名前または名前の部分の定義を示す)
  - 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> (別名, すなわち, 一般的に使われている名前のリストを示す)
- @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>
```

## Next

We're going to do the third and final exercise.