



CENTER FOR TEXTUAL STUDIES AND DIGITAL HUMANITIES

DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2014

Week 9

Today's Class

- TEI Exercise
- TEI Example Project Files
- TEI continued

DIGH 400 - Introduction to Digital Humanities Research

TEI Exercise

```
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>
          <!--Title-->
        </title>
      </titleStmt>
      <publicationStmt>
        <p>
          <!--Publication Information-->
        </p>
      </publicationStmt>
      <sourceDesc>
        <p>
          <!--Information about the source-->
        </p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <!--Some structural division, paragraph, line group, speech, ...-->
    </body>
  </text>
</TEI>
```

TEI Exercise - solution options

- within the <teiHeader> we could add
 - the title of the text
 - <respStmt>
 - <publicationStmt>
 - <sourceDesc>
- within <text> we could add
 - <front> with a dedication and table of contents...
 - <body> with transcription etc of the text
 - <back> with a div of type 'colophon'...

Text Encoding Initiative - Example 1

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>The Epic of Gilgamesh</title>
      <respStmt>
        <resp>editor</resp>
        <name xml:id="AL">Alasdair Livingstone</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <p>Not for distribution.</p>
    </publicationStmt>
    <sourceDesc>
      <p>Transcribed from the diaries of Professor Wilfred
        Lambert</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Text Encoding Initiative - Example 2

```
<text>
<front>
  <div type="dedication">
    <p>personal dedication...</p>
  </div>
  <div type="contents">
    <head>Table of Contents</head>
    <list>
      <item>1. No.1...</item>
      <item>2. No.2...</item>
      <item>3. No.3...</item>
    </list>
  </div>
</front>
<body>
  <p>some body text goes here....</p>
</body>
<back>
  <div type="colophon">
    <p>Physical book conditions...</p>
  </div>
</back>
</text>
```

Text Encoding Initiative - Example Project Documents

TEI Example Project Files...

Ongoing assessment

<http://www.digital-humanities.com>

DIGH 400 - Introduction to Digital Humanities Research

Ongoing assessment

[SciFi Authors](#)

Text Encoding Initiative

More about Textual Phenomena - Structural Features

- Prose elements

- <p> <div> <head> <list> <item> <q> <pb> <seg> <figure>
<table>

- Verse elements

- <lg> <l>

- Drama elements

- <div> <sp> <p> <lg> <l> <seg>

Text Encoding Initiative

Some examples

```
<list>  
<item></item>  
</list>
```

```
<lg type="free">  
<|></|>  
</lg>
```

```
<seg type="preamble">  
<seg></seg>  
<seg></seg>  
</seg>
```

an example with 'quoted' [q](#)

Text Encoding Initiative

Title Pages

- <titlePage> used for transcription and encoding of physical title page
- within <front> or <back> using <titlePage>
- <titlePage> commonly includes <docTitle>
- <docTitle> may include subsections, <titlePart>
- <docAuthor> for name of the author
- <byline> for primary statement of responsibility for a work
- Other elements can include
 - <docEdition>
 - <docDate>
 - <docImprint> may contain <pubPlace> <docDate>
- <publisher>...
 - <epigraph>
 - <imprimatur>
 - <graphic/>

Text Encoding Initiative

More about Textual Phenomena - Logical and Semantic Features

- encodes underlying semantics with specific elements other than <hi>
- highlighting may be used to show
 - <emph> <foreign> <distinct>
- uses of quotation
 - <said> <quote> <cit> <mentioned> <soCalled>
- other examples include
 - <term> <gloss>...
- generic elements may also carry semantic and logical information
 - <title> <name> <num> <measure> <date> <address> <abbr>
<expan>

Text Encoding Initiative

More about Textual Phenomena - Logical and Semantic Features

Referring Strings

- in TEI places, people, objects can use the `<name>` elements with `@type` attribute
 - eg: `<name type="person">Jack</name>`
- proper nouns can also be referred to using the `<rs>` element plus `@type` attribute
 - eg: `<rs type="person">he</rs>`

Text Encoding Initiative

More about Textual Phenomena - Logical and Semantic Features

Dates and Time

- <date> and <time>
- system or calendar used may be documented using @calendar attribute
- value of date or time supplied using @when attribute
- normalised representation of <date> should conform to valid W3C datatype

<date when="2003-12-22" calendar="Gregorian">22 Nov 2003</date>

- <date> can also be used to mark a period of time using the attributes
 - @from @to @notBefore @notAfter

Text Encoding Initiative

More about Textual Phenomena - Logical and Semantic Features

Numbers and Measures

- <num> and <measure>

<num type="percentage" value="23">23%</num>

<measure type="volume" quantity="1.5" unit="litre" commodity="wine"
>1.5L bottle of wine</measure>

Text Encoding Initiative

More about Textual Phenomena - Logical and Semantic Features

Addresses

- physical and digital addresses can be encoded
 - <address> <email>
- <email> can also use a @type attribute
- <addrLine> within <address>
- <name> <street> <postCode> <postBox>

Text Encoding Initiative

More about Textual Phenomena - Logical and Semantic Features

Abbreviations and Expansions

- explicitly encode <abbr> and <exp>
- @type attribute may be used with <abbr>
 - <abbr type="title">Dr</abbr>
- <exp> used with <abbr> within <choice>

<choice>

<abbr type="title">Dr</abbr>

<exp>Doctor</exp>

</choice>

Text Encoding Initiative

More about Textual Phenomena - Analytical Features

Notes and Annotations

- <note> can be used to record a textual annotation
- @type attribute used to differentiate notes
- @resp attribute used to assign responsibility
- position of the note can be referenced using the @place attribute

```
<note n="1" place="foot" type="editorial" resp="NJH">a new note...</note>
```

Text Encoding Initiative

More about Textual Phenomena - Analytical Features

Index Entries

- pre-existing in <front> or <back> using <list> inside <div>
- new index using <term> inside <index> at the location of the index item

<index>

<term>new index term...</term>

</index>

Text Encoding Initiative

More about Textual Phenomena - Analytical Features

Errors

- indicated using `<sic>`
- corrected using `<corr>`
- combine `<sic>` and `<corr>` within `<choice>`
- use `@cert` and `@resp` attributes to encode degree of certainty and editor responsible

`<corr cert="high" resp="#NJH">correction...</corr>`

Text Encoding Initiative

More about Textual Phenomena - Analytical Features

Regularisation and Normalisation

- <reg> element for regularisation
- <orig> for original, non-normalised form
- use <reg> in isolation or combined with <orig> within <choice>

<choice>

<orig>thou</orig>

<reg resp="#NJH">you</reg>

</choice>

Text Encoding Initiative

More about Textual Phenomena - Analytical Features

Additions, Deletions & Omissions

- [<gap>](#) element used for omission, both material and editorial
- @reason attribute used to indicate reason for omission
- @extent and @unit attributes can be used to record extent of omission
- editorial omissions should be recorded using <editorialDecl> inside <editorialDesc> in <teiHeader>
- <gap> may be empty or include a <desc> of the material omitted
- <add> and may also be used for words and phrases
- @rend attribute may also be used with <add> and
- <addSpan/> and <delSpan/> for longer passages
- <subst> to contain <add> and with causal relationship
- <unclear> with @reason attribute for difficult to read deletions in the text

Text Encoding Initiative

Non-Textual Phenomena

- graphics such as illustrations, diagrams, drawings, artwork...
- anchor in a text using `<graphic/>` and optional `@url` attribute

`<graphic url="http://www.somewhere.com/image.jpg"/>`

- use `<figure>` element as parent to create a full listing for a graphic

Text Encoding Initiative

Global Attributes

There are currently 6 global attributes:

- @n
- @rend
- @rendition
- @xml:base
- @xml:id
- @xml:lang

Text Encoding Initiative

[TEI P5 Guidelines](#)