



CENTER FOR TEXTUAL STUDIES AND DIGITAL HUMANITIES

DIGH 400 - Introduction to Digital Humanities Research

Fall Semester 2013

Week 9

Today's Class

- Text Encoding Initiative (TEI) continued

Discussion Analysis

The Newton Project

Text Encoding Initiative

Getting Started: Common Data Structure & Elements

- expressed using XML
- elements for information
- attributes for additional information
- comments use standard delimiters
 <!--this is a comment in TEI-->

Text Encoding Initiative

TEI Document

- <teiHeader> = documents all metadata for the TEI document
- <text> = contains the document proper
- these elements are mandatory for all TEI documents
- this structure is contained within the <TEI> element

<TEI>

<teiHeader>...</teiHeader>

<text>...</text>

</TEI>

Text Encoding Initiative

TEI Header <teiHeader>

- mandatory for all TEI documents
- <fileDesc> = contains a description of the electronic file
- <fileDesc> consists of at least three mandatory elements
 - <titleStmt>
 - <publicationStmt>
 - <sourceDesc>

Text Encoding Initiative

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Around the World in Eighty Days</title>
      <respStmt>
        <resp>editor</resp>
        <name xml:id="PS">Peter Shillingsburg</name>
      </respStmt>
    </titleStmt>
    <publicationStmt>
      <p>Not for distribution.</p>
    </publicationStmt>
    <sourceDesc>
      <p>Transcribed from 1873 English Edition</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Text Encoding Initiative

<!-- transcriber note -->

<note resp="#NH">Possible mis-interpretation of author correction.
</note>

<!-- editorial note -->

<note resp="#AL">correction meant as addition and not
correction</note>

Text Encoding Initiative

<text>

<body>

- minimally contains <body>
- <body> contains lower level text structures such as
 - <p>
 - or different structures for genres other than prose

<text>

<body>

<p>some text...</p>

</body>

</text>

Text Encoding Initiative

<text>

<front>

- optional element containing front matter such as
 - title pages, headers, prefaces, dedications...
- also consider prologues for drama, forewords and introductions for prose
- except for title, front matter should be encoded using generic standard elements

<front>

<div type="dedication">

<p>a small dedication...</p>

</div>

</front>

Text Encoding Initiative

<text>

<back>

- optional element and grouping
- may contain all back matter for a text
- either numbered or un-numbered divisions with @type attribute
 - appendix
 - glossary (<list> of terms and their explanations)
 - notes
 - bibliogr (list of bibliographical citations <listBibl>)
 - index
 - colophon

<back>

<div type="colophon">

<p>Printed and bound by.....</p>

</div>

</back>

Text Encoding Initiative

```
<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

Unitary or Composite Texts

<group>

- group structurally related text

<teiCorpus>

- corpus of diverse texts

DIGH 400 - Introduction to Digital Humanities Research

```
<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>
```

Text Encoding Initiative

- 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

```
<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

```
<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

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>