MODULE 2: THE TEI HEADER

Ron Van den Branden

Edward Vanhoutte

Melissa Terras
# Table of Contents

1. Desiderius Erasmus: *Colloquia familiaria* .......................................................... 1

2. Thomas Wentworth Higginson: “Letter of 7 November 1885” ................................. 4

3. Christopher Marlowe: *The Tragedie of Doctor Faustus (B text)* ............................ 7


5. Walt Whitman: “After the Argument” ................................................................. 13

6. Oscar Wilde: *The Importance of Being Earnest* ............................................... 16
1. Desiderius Erasmus: *Colloquia familiaria*

This example features the TEI header for the transcription of *Colloquia familiaria*, a series of colloquia written by Desiderius Erasmus. They are encoded and made available by the Stoa Consortium, University of Kentucky.

This is an excellent example of a TEI header. The file description provides the minimal information sections about the title and responsibilities of the electronic text, its publication, and its source. Editorial principles are documented in `<encodingDesc>`, which also has a statement about sampling decisions in `<samplingDesc>` (see section 2.3.2 The Sampling Declaration of the TEI Guidelines). It also contains a formal declaration of a reference system, for which it makes use of `<refState>` elements (see section 2.3.5.3 Milestone Method of the TEI Guidelines). Two classification systems are declared in `<classDecl>`: Library of Congress Subject Headings and Library of Congress Classification. The next header section, `<profileDesc>`, contains the actual classification of the text according to both systems, in `<textClass>`. This is a nice illustration of two classification strategies: using natural language keywords (<`keywords`>) or abstract classification codes (<`classCode`>). Also, the languages of the text are formally declared in `<langUsage>`. Finally, a complete revision history is available in `<revisionDesc>`.

```xml
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <titleStmt>
      <title>Colloquia familiaria</title>
      <author>Desiderius Erasmus</author>
      <respStmt>
        <resp>Encoded by</resp>
        <name xml:id="JN">Jennifer K. Nelson</name>
        <name xml:id="GE">Gretche Ermer</name>
      </respStmt>
      <principal>A. Ross Scaife</principal>
      <sponsor>Cultural Heritage Language Technologies</sponsor>
      <funder>NSF-EU</funder>
    </titleStmt>
    <publicationStmt>
      <publisher>Stoa Consortium</publisher>
      <pubPlace>Lexington, KY</pubPlace>
    </publicationStmt>
  </fileDesc>
</teiHeader>
```
Desiderius Erasmus: Colloquia familiaria

1. Desiderius Erasmus: Colloquia familiaria
Hyphenated words that appear at the end of the line in the Holtze edition have been reformed.

Italics are recorded without interpretation.

[ed] added text of "De utilitate colloquiorum ad lectorem"

1. Desiderius Erasmus: Colloquia familiaria
Module 2: The TEI Header

Example 1. Adapted from a TEI P4 XML encoding of the header for Desiderius Erasmus’ *Colloquia Familiaria* (Erasmus, 1872).


This example shows the TEI header of the digital edition of a letter of 7 November 1885 by the American minister and writer Thomas Wentworth Higginson, encoded and made available by the Lincoln Electronic Text Center of the University of Nebraska.

This TEI header provides detailed documentation about the electronic text in `<fileDesc>`. The title statement not only identifies the people responsible for transcription and markup, but also for the technical processing of the letters by means of stylesheets. The `<extent>` section needs to be completed still; of course, this can only be done after completion of the encoding. Notice the detailed statement of availability in `<availability>`. The source text in which this letter has been published is described using the `<biblFull>` element; notice how its sections reflect the actual file description in the TEI header of the electronic text (apart from the `<sourceDesc>` section). The `<notesStmt>` seems to be used to record some loose annotations about the source text.

The encoding description section only contains a description of the editorial practice in `<editorialDecl>`. This is done in a prose paragraph. The header is concluded by a minimal revision description, recording only one change.

```
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <titleStmt>
      <title>Correspondence of Thomas Wentworth Higginson, 1865–1910</title>
      <author>Thomas Wentworth Higginson</author>
    </titleStmt>
    <respStmt>
      <resp>Transcribed by</resp>
      <name xml:id="MS">Melissa Sinner</name>
    </respStmt>
  </fileDesc>
</teiHeader>
```
Line breaks, paragraph breaks, and indentations have been preserved within the transcription. The layout of the page has been preserved whenever possible. Abbreviations and spellings have been maintained within the transcriptions, and the full word and corrected or modern spellings have been provided. Images of the original letters have been provided in order to show the original page layout and other markings that are not the author's. Such markings include letter head, postcards, postal stamps, etc. and writing by other people. Words or
phrases that are deemed indecipherable have been noted as "unclear." To provide further information as to the context of a particular letter, notations have been provided. 

Example 2. Adapted from a TEI P4 XML encoding of a letter by Thomas Higginson (1885). TEI XML source file is not publicly available.

3. Christopher Marlowe: The Tragedie of Doctor Faustus (B text)

This example contains the TEI header of the digital edition of Christopher Marlowe's The Tragedie of Doctor Faustus (B text), encoded and made available by the Perseus Digital Library.

This TEI header provides decent descriptions of the publication details of the electronic text (<publicationStmt>), and the languages occurring in the text (<lang usage>). A reference system is declared in the <encodingDesc> section of the header, using <refState> elements (see section 2.3.5.3 Milestone Method of the TEI Guidelines).

The revision description is interesting both in a positive and a negative way. It clearly contains a detailed list of the changes. The list seems to be generated by an automated versioning system, which allows one to keep complete track of a file's historical states, and document changes with log messages. Integrating automated revision control in the <revisionDesc> section of the TEI header is an interesting idea, as it combines processability and expressiveness. However, on the encoding level, this integration could be improved. In this case, a single <change> element is (ab)used to record the complete revision history. If the output of the automated version control system would be formatted to distinct <change> elements per revision (either directly, or via a post-processing step), this would make the information much more compliant with the semantics of the TEI header.
One essential point of critique concerns the lacking description of the source document in `<sourceDesc>`. In this case, the title and author of the source work (that can be recollected from the information in the `<titleStmt>` subsection) still provide cues to its origin, but this could be much harder for less known texts. It is reasonable to suppose that the source texts of the files in the Perseus Digital Library are documented externally, but then the TEI header sections of these files should at least contain a pointer to these resources.

```xml
<teiHeader xmlns="http://www.tei-c.org/ns/1.0" type="text">
  <fileDesc>
    <titleStmt>
      <title>The Tragedie of Doctor Faustus (B text)</title>
      <author>Christopher Marlowe</author>
      <editor role="editor">Hilary Binda</editor>
      <sponsor>Perseus Project, Tufts University</sponsor>
      <principal xml:id="GC">Gregory Crane</principal>
      <respStmt>
        <resp>Prepared under the supervision of</resp>
        <name xml:id="LC">Lisa Cerrato</name>
        <name xml:id="WM">William Merrill</name>
        <name xml:id="EM">Elli Mylonas</name>
        <name xml:id="DS">David Smith</name>
      </respStmt>
      <funder n="org:Tufts">Tufts University</funder>
    </titleStmt>
    <publicationStmt>
      <publisher>Trustees of Tufts University</publisher>
      <pubPlace>Medford, MA</pubPlace>
      <authority>Perseus Project</authority>
      <availability status="free">
        <p>This text may be freely distributed, subject to the following restrictions: <list>
          <item>You credit Perseus, as follows, whenever you use the document: <quote>Text provided by Perseus Digital Library, with funding from Tufts University. Original version available for viewing and download at http://www.perseus.tufts.edu/hopper/.</quote></item>
          <item>You leave this availability statement intact.</item>
          <item>You use it for non-commercial purposes only.</item>
          <item>You offer Perseus any modifications you make.</item>
        </list>
      </availability>
    </publicationStmt>
  </fileDesc>
</teiHeader>
```

The following example illustrates the TEI header for a sonnet by William Shakespeare, containing a detailed metrical analysis of the poem. Both the electronic text and its source are bibliographically described in the `<fileDesc>` section. The text encoding process is described in `<encodingDecl>`, providing details about the encoding project ( `<projectDesc>`), the editorial policy ( `<editorialDecl>`), and the system used to analyse the metre of the poem ( `<metDecl>`). Notice how the `<editorialDecl>` subsection had to be repeated, as it both documents features that can be encoded in a TEI category ( `<segmentation>` and `<interpretation>`), and features for which no such TEI labels are available. ( `<p>`). The standard TEI scheme does not allow both systems (formal and informal) to be mixed, hence the repetition of the `<encodingDesc>` section. The same goes for the `<metDecl>` sections: as both a formal ( `<metSym>` ) and informal ( `<p>` ) description is provided for the metrical system, repeating the `<metDecl>` element was the easiest solution. Of course, this could have been addressed as well by adapting the TEI schema.

```xml
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <titleStmt>
      <title>A Selection of Sonnets: electronic edition encoded in XML with a TEI DTD</title>
      <author>William Shakespeare</author>
      <respStmt>
        <resp>Transcribed and encoded by</resp>
        <name>Mubina Islam</name>
      </respStmt>
    </titleStmt>
    <extent>64 KB</extent>
    <publicationStmt>
      <publisher>University College London</publisher>
      <pubPlace>London</pubPlace>
    </publicationStmt>
    <sourceDesc>
      <bibFull>
```

<titleStmt>
<title>The Complete Works of William Shakespeare</title>
<author>
  <name>William Shakespeare</name>
</author>
<editor>
  <name>Peter Alexander</name>
</editor>
</titleStmt>

<publicationStmt>
<publisher>Collins</publisher>
<pubPlace>London</pubPlace>
<date>1978</date>
{idno type="ISBN">0-00-435634-9</idno>
</publicationStmt>

<projectDesc>
<p>A total of ten sonnets collected and encoded according to the metrical interpretation of the verse by Mubina Islam, August 2004. This document was created as part of a Master's dissertation on the markup of poetic metre, for the course MA Electronic Communication and Publishing at UCL.</p>
</projectDesc>

<editorialDecl>
<segmentation>
<p>Each sonnet has been divided into the stanzaic line groupings.</p>
<p>Line groups have been further divided to mark individual lines of verse.</p>
<p>Segmentation tags have been used to represent the division of each line into metrical feet.</p>
</segmentation>

<interpretation>
<p>The metrical interpretation of the text, defined with the segmentation of the text into units of feet, was added by hand by the encoder. This has not been checked and may be subject to alternative readings.</p>
</interpretation>
</editorialDecl>
All punctuation marks, excluding dashes or hyphenation, have been encoded as entities.

Caesuras and line enjambement have been recorded in this document as accurately as possible by the encoder.

Metrically prominent syllables are marked '+' and other syllables '-'. Foot divisions are marked by a vertical bar, and line divisions with a solidus.

This notation may be applied to any metrical unit, of any size (including, for example, individual feet as well as groups of lines).

The 'real' attribute has been used to indicate possible variations in the iambic base metre. Where this attribute is not included, it is assumed each foot inherits the iambic metre defined for the overall division of text.

The 'met' attribute has been used in feet which have a missing or additional syllable rather than the two syllables expected, although the line may still confirm to the metre of the poem.
5. Walt Whitman: “After the Argument”

This example contains the TEI header of the digital edition of a manuscript draft of “After the Argument,” a poem by Walt Whitman. It was encoded and made available by the Walt Whitman Archive.

This TEI header contains a detailed description of the electronic text in <fileDesc>. Apart from the required subsections, the edition of the electronic text is identified briefly in <editionStmt>. The <notesStmt> element contains a general remark about the dating of the manuscript.

Besides the file description, the header contains a detailed account of the file’s history in <revisionDesc>. Functioning as the header of a manuscript transcription, however, one would have expected at least an <encodingDesc>, documenting how the electronic version relates to the source text. When this text is seen in isolation, this header falls short in explaining the editorial choices (that are referred to, however, in the <revisionDesc>). Of course, this text probably features in the wider context of the Walt Whitman Archive, where uniform encoding practices were used for all texts. Still, without repeating boilerplate information in each text of the archive, it would have made sense to provide an <editorialDecl> section with at least pointers to the external documentation of these practices available at http://www.whitmanarchive.org/about/editorial.html and http://www.whitmanarchive.org/mediawiki/index.php/Whitman_Encoding_Guidelines. Furthermore, as the transcription is fairly detailed in the recording of editorial phenomena (additions, deletions, substitutions), identification of the different document hands in <profileDesc> could have made sense.

(Of course, these are only minor remarks, relative to the quality of the surrounding documentation of the archive in which this text is embedded. Yet, even if such external documentation exists, it makes sense to provide pointers in the document.)

```xml
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <titleStmt>
      <title level="m" type="main">After the Argument</title>
      <title level="m" type="sub">a machine readable transcription</title>
      <author>Walt Whitman</author>
      <editor>Ken Price</editor>
      <editor>Ed Folsom</editor>
    </titleStmt>
  </fileDesc>
</teiHeader>
```
Example 5. Encoding of a manuscript of Walt Whitman's poem "After the Argument " (1890).


6. Oscar Wilde: The Importance of Being Earnest

This example contains the TEI header for an electronic edition of Oscar Wilde's The Importance of Being Earnest, encoded and made available by Corpus of Electronic Texts (CELT), a project of University College, Cork.

This is an excellent TEI header example, featuring quality descriptions of the electronic text (<fileDesc>), its relation to the source text (<encodingDesc>), the context in which it came about (<profileDesc>), and a revision history (<revisionDesc>).

An outstanding feature of this example is the level of detail for the bibliographic description of the source text, in <sourceDesc>. It contains a complete bibliography, in three sections: “select editions,” “select bibliography,” and “the edition used in the digital edition.” The former two categories consist of bibliographic lists, with a <listBibl> element grouping the separate <bibl> elements. The actual edition used for the electronic text is described in detail with a <biblStruct> element.
<name xml:id="ML">Margaret Lantry</name>
</respStmt>
<funder>University College, Cork</funder>
</titleStmt>
<editionStmt>
<edition n="2">Second draft.</edition>
<respStmt>
<resp>Proof corrections by</resp>
<name>Margaret Lantry</name>
</respStmt>
</editionStmt>
<extent>
<measure type="words">23410</measure>
</extent>
<publisher>CELT: Corpus of Electronic Texts: a project of University College, Cork</publisher>
<address>
<addrLine>College Road, Cork, Ireland—http://www.ucc.ie/celt</addrLine>
</address>
<date>1997</date>
<date>2008</date>
<distributor>CELT online at University College, Cork, Ireland. </distributor>
<idno type="celt">E850003-002</idno>
<availability status="restricted">
<p>Available with prior consent of the CELT programme for purposes of academic research and teaching only.</p>
</availability>
</publicationStmt>
<notesStmt>
<note>There is not as yet an authoritative edition of Wilde's works.</note>
</notesStmt>
<sourceDesc>
<listBibl>
<head>Select editions</head>
<bibl n="1">The writings of Oscar Wilde (London; New York: A. R. Keller & Co. 1907) 15 vols.</bibl>
</listBibl>
</sourceDesc>

Complete works of Oscar Wilde (Glasgow: HarperCollins, 1994).

Select bibliography

'Notes for a bibliography of Oscar Wilde', Books and book-plates (A quarterly for collectors) 5, no. 3 (April 1905), 170-183.


Richard Ellmann (ed), The Artist as Critic: Critical Writings of Oscar Wilde (Chicago 1982).


Frank Harris, Oscar Wilde, including My memories of Oscar Wilde, by George Bernard Shaw and an introductory note by Lyle Blair (London: Robinson, 1992).


6. Oscar Wilde: The Importance of Being Earnest
Module 2: The TEI Header

`<projectDesc>`
```xml
<p>CELT: Corpus of Electronic Texts</p>
</projectDesc>`

`<samplingDecl>`
```xml
<p>All the editorial text with the corrections of the editor has been retained.</p>
</samplingDecl>`

`<editorialDecl>`
```xml
<correction status="medium">
<p>Text has been checked, proof-read and parsed using NSGMLS.</p>
</correction>
<normalization>
<p>The electronic text represents the edited text.</p>
</normalization>
<quotation>
<p>Direct speech is marked <emph>q</emph>.</p>
</quotation>
<hyphenation>
<p>The editorial practice of the hard-copy editor has been retained.</p>
</hyphenation>
<segmentation>
<p><emph>div0</emph> = the whole text.</p>
</segmentation>
<interpretation>
<p>Names of persons (given names), and places are not tagged. Terms for cultural and social roles are not tagged.</p>
</interpretation>
</editorialDecl>`

`<refsDecl>`
```xml
<p>The <emph>n</emph> attribute of each text in this corpus carries a unique identifying number for the whole text.</p>
<p>The title of the text is held as the first <emph>head</emph> element within each text.</p>
<p><emph>div0</emph> is reserved for the text (whether in one volume or many).</p>
</refsDecl>`

`<encodingDesc>`
```
```

`<profileDesc>`
```xml
<creation>By Oscar Wilde (1854-1900). <date>1895</date></creation>`

6. Oscar Wilde: The Importance of Being Earnest 20

Higginson, Thomas Wentworth. 1885. “Letter of November 7, 1885.” Encoded and made available by the Lincoln Electronic Text Center of the University of Nebraska at http://higginson.unl.edu/letters/LC1885k07.html.


