TEI: The Text Encoding Initiative

P. S. Langeslag

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Fundamentals

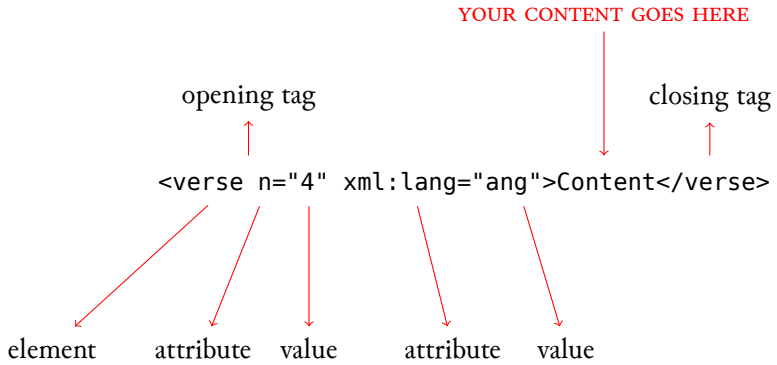
HTML: Hypertext Markup Language

```
<!doctype html>
<html lang="en">
  <head>
    <meta http-equiv="content-type" content="text/html; charset=utf-8"/>
  </head>
  <body>
    <h1>Title Goes Here</h1>
    <p>First paragraph. <em>Abstract</em> <strong>formatting</strong>
     permitted.</p>
  </body>
</html>
```

XML: Extensible Markup Language

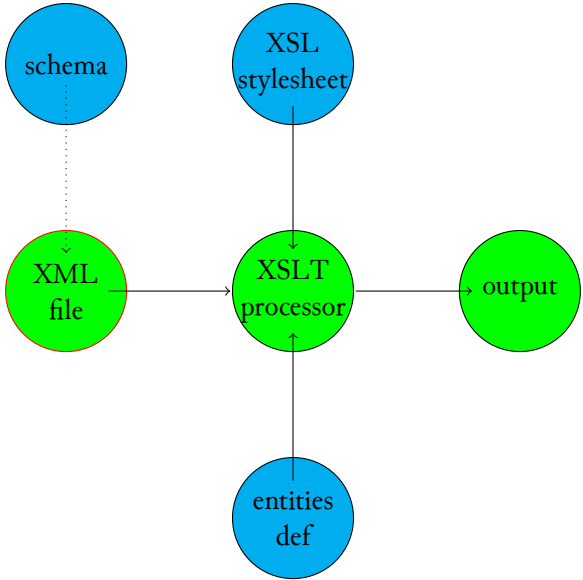
```
<?xml version="1.0" encoding="UTF-8"?>
<library>
  <name>Robarts Library</name>
  <floor n="8">
    <bookcase id="PR">
      <book id="PR1583.K53">
        <title>Klaeber's <em>Beowulf</em></title>
        <editor>Fulk, R. D., R. E. Bjork, and J. D. Niles</editor>
        <year>2012</year>
      </book>
    </bookcase>
  </floor>
</library>
```

Elements and Attributes




```
<title></title>
<title/>
<locus from="" to=""></locus>
```

XML Infrastructure



Who Defines the Building Blocks of the Tree Structure?

Standard	Structure	Authority
HTML	Predefined	W3C
XML	User-defined	(W3C)
TEI	Predefined; applies to XML	TEI-C

In version 4.2.0, revision 736c0acf0, TEI defines:

- 589 elements;
- 268 distinctly named attributes.

Main TEI Document Structure

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI>
    <teiHeader/> <!-- Describes file, text, source(s); defines identifiers -->
    <text>
        <front/> <!-- Preface, prologue, abstract, or similar -->
        <body/>  <!-- Main text; cf. HTML -->
    </text>
</TEI>
```

Text Structure Elements 1 (Selection)

Element	Function
`<front>`	Preface, prologue, abstract, or similar
`<body>`	Main text (cf. HTML)
`<div>`	Upper-level division: book part, chapter, section, act, scene; recursive (cf. HTML)
`<div1>` ... `<div7>`	Explicit hierarchical division
`<p>`	Paragraph (vgl. HTML)
`<ab>`	Block-level element: paragraph-like without the associations of a paragraph
`<s>`	Sentence-like element
`<l>`	Verse line

(Nesting by library logic: library > bookcase > book)

Text Structure Elements 2 and Other Elements (Selection)

Element	Function
`<head>`	Heading
`<seg>`	Lower-level segment; recursive
`<w>`	Word
`<quote>`	Quotation
`<foreign>`	Language identifier
`<name>`	Name, e.g. personal name, place-name, organization
`<placeName>`	Shorthand for `<name type="place">`
`<hi>`	Graphical description, e.g. cursive, colour, font size

Text Structure Elements 2 and Other Elements (Selection)

Element	Function
`<head>`	Heading
`<seg>`	Lower-level segment; recursive
`<w>`	Word
`<quote>`	Quotation
`<foreign>`	Language identifier
`<name>`	Name, e.g. personal name, place-name, organization
`<placeName>`	Shorthand for `<name type="place">`
`<hi>`	Graphical description, e.g. cursive, colour, font size

Availability in Schemata

- Schemata negotiate between frugality and flexibility;
- Modules offer a time-saving compromise.

Modules (Selection)

Module	Covers
`core`	Universal TEI-elements: **<title>**, **<abbr>**
`header`	TEI header: **<fileDesc>**, **<profileDesc>**
`textstructure`	Text structure elements: **<body>**, **<div1>**
`linking`	Linking, division, joining: **<ab>**, **<seg>**, **<join>**
`corpus`	Text corpus elements: **<setting>**, **<textDesc>**
`dictionaries`	Lexicographiscal elements: **<def>**, **<etym>**, **<colloc>**
`analysis`	Interpretative elements: **<s>**, **<w>**, **<interp>**
`msdescription`	MS description: **<msDesc>**, **<rubric>**
`transcr`	Transcription elements: **<facs>**, **<damage>**
`textcrit`	Text-critical apparatus: **<witness>**, **<lacunaStart>**
`verse`	Verse structure: **<rhyme>**, **<caesura>**

Examples With (and Without) Attribute(s)

```
<foreign xml:lang="la">Arma virumque cano</foreign>
            <!-- Language key @ https://www.iana.org/assignments/
             language-subtag-registry/language-subtag-registry -->

<quote source="#augDcd">Vbi eum audiuimus?</quote>


<w lemma="góðr" pos="A" msd="asm">Goðan</w>


<s>How is <hi rend="italic">italic typeface</hi> encoded?</s>
                                    <!-- Orthographical presentation -->

<s>He said <emph>what</emph>?</s>
                                    <!-- Linguistic presentation -->
```

Text Example

Plaintext

```
Aldhelm incorporates the following verse lines into his De laude virginitatis:

Christus passus patibula
Atque leti latibula
Virginem uirgo uirgini
Commendabat tutamini
```

Paragraphs and Verse Lines

```
<p>Aldhelm incorporates the following verse lines into his
 <title level="m">De laude virginitatis</title>:</p>

<lg xml:id="xpassus">
 <l>Christus passus patibula</l>
 <l>Atque leti latibula</l>
 <l>Virginem uirgo uirgini</l>
 <l>Commendabat tutamini</l>
</lg>
```

Context

```
<TEI version="3.5.0" xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader/>
 <text>
  <body>
   <p>Aldhelm incorporates the following verse lines into his
    <title level="m">De laude virginitatis</title>:</p>

   <lg xml:id="xpassus">
    <l>Christus passus patibula</l>
    <l>Atque leti latibula</l>
    <l>Virginem uirgo uirgini</l>
    <l>Commendabat tutamini</l>
   </lg>

  </body>
 </text>
</TEI>
```

Explicit Line Numbering

```
<TEI version="3.5.0" xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader/>
 <text>
  <body>
   <p>Aldhelm incorporates the following verse lines into his
    <title level="m">De laude virginitatis</title>:</p>

   <lg xml:id="xpassus">
    <l n="1">Christus passus patibula</l>
    <l n="2">Atque leti latibula</l>
    <l n="3">Virginem uirgo uirgini</l>
    <l n="4">Commendabat tutamini</l>
   </lg>

  </body>
 </text>
</TEI>
```

Language Specification

```
<TEI version="3.5.0" xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
 <teiHeader/>
 <text>
  <body>
   <p>Aldhelm incorporates the following verse lines into his
    <title level="m" xml:lang="la">De laude virginitatis</title>:</p>

   <lg xml:id="xpassus" xml:lang="la">
    <l n="1">Christus passus patibula</l>
    <l n="2">Atque leti latibula</l>
    <l n="3">Virginem uirgo uirgini</l>
    <l n="4">Commendabat tutamini</l>
   </lg>

  </body>
 </text>
</TEI>
```

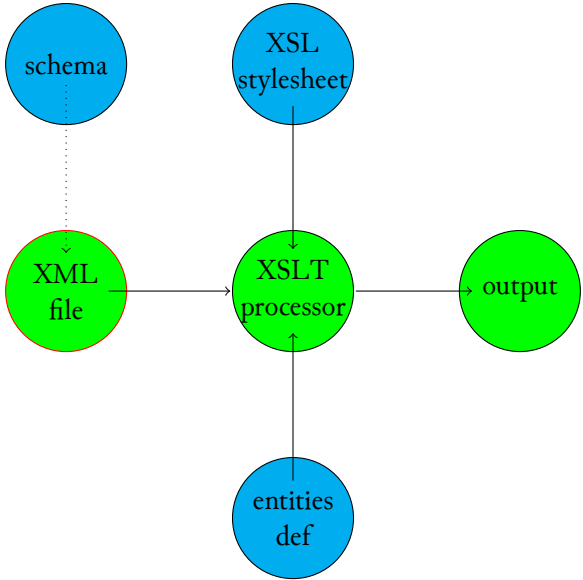
Metre

```
<TEI version="3.5.0" xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
 <teiHeader/>
 <text>
  <body>
   <p>Aldhelm incorporates the following verse lines into his
    <title level="m" xml:lang="la">De laude virginitatis</title>:</p>

   <lg xml:id="xpassus" xml:lang="la" type="quatrain" met="+-+-+-+-" rhyme="aa">
    <l n="1">Christus passus patibula</l>
    <l n="2">Atque leti latibula</l>
    <l n="3">Virginem uirgo uirgini</l>
    <l n="4">Commendabat tutamini</l>
   </lg>

  </body>
 </text>
</TEI>
```

Names and Cross-References

```
<TEI version="3.5.0" xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">
 <teiHeader/>
 <text>
  <body>
   <p><name type="person" ref="#aldhelm">Aldhelm</name> incorporates the
    following verse lines into his <title level="m" xml:lang="la" ref="#aldDLV">
    De laude virginitatis</title>:</p>
   <lg xml:id="xpassus" xml:lang="la" type="quatrain" met="+-+-+-+-" rhyme="aa">
    <l n="1">Christus passus patibula</l>
    <l n="2">Atque leti latibula</l>
    <l n="3">Virginem uirgo uirgini</l>
    <l n="4">Commendabat tutamini</l>
   </lg>
  </body>
 </text>
</TEI>
```

XML Infrastructure



XML Declarations and the TEI Header

Declarations

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE TEI [
    <!ENTITY punctvers "&#xF1EA;">  <!-- punctus versus -->
    <!ENTITY pbardes "&#xA751;">    <!-- p with bar through descender -->
    ]>
<?xml-model href="
 https://tei-c.org/release/xml/tei/custom/schema/relaxng/tei_lite.rng"
            type="application/xml"
            schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-stylesheet type="text/css" href="render.css"?>
```

TEI Header

```
<teiHeader>
 <fileDesc/>
 <encodingDesc/>
 <profileDesc/>
 <xenoData/>
 <revisionDesc/>
</teiHeader>
```

TEI Header

```
<teiHeader>
 <fileDesc/>        <!-- Mandatory -->
 <encodingDesc/>    <!-- Recommended -->
 <profileDesc/>     <!-- Optional -->
 <xenoData/>        <!-- Optional -->
 <revisionDesc/>    <!-- Recommended -->
</teiHeader>
```

TEI Header

```
<teiHeader>
 <fileDesc/>        <!-- Describes the XML file -->
 <encodingDesc/>    <!-- Describes the encoding of sources -->
 <profileDesc/>     <!-- Describes the text -->
 <xenoData/>        <!-- Place for non-TEI metadata -->
 <revisionDesc/>    <!-- Changelog -->
</teiHeader>
```

Manuscript Signature in **<fileDesc>**

```
<sourceDesc>
 <msDesc>
  <msIdentifier>
   <settlement>Vercelli</settlement>
    <repository>Biblioteca Capitolare</repository>
     <idno>MS CXVII</idno>
  </msIdentifier>
  <msContents>
   <msItem>
    <locus from="9v" to="12r"/>
    <title>Vercelli homily II</title>
   </msItem>
  </msContents>
 </msDesc>
</sourceDesc>
```

Dating in <fileDesc>

```
<sourceDesc>
 <msDesc>
  <msIdentifier/>
  <origin>
   <origDate notBefore="0963" notAfter="0987"/> <!-- If known use format -->
  </origin>                                     <!-- when="0982-05-01" -->
 </msDesc>
</sourceDesc>
```

Bibliography in **<fileDesc>**

```
<sourceDesc>
 <bibl>
  <relatedItem type="otherEdition">
   <biblStruct>
    <monogr>
     <editor>Donald <surname>Scragg</surname></editor>
     <title level="m">The Vercelli Homilies</title>
     <imprint>
      <pubPlace>Oxford</pubPlace>
      <publisher>Oxford University Press</publisher>
      <date>1992</date>
     </imprint>
     <biblScope unit="pp" from="52" to="62"/>
    </monogr>
   </biblStruct>
  </relatedItem>
 </bibl>
```

Changelog in <revisionDesc>

```
<revisionDesc>
 <change n="0.3" when="2019-05-21" who="#el">added facsimile links</change>
 <change n="0.2" when="2019-05-20" who="#psl">added basic markup</change>
 <change n="0.1" when="2019-05-19" who="#psl">transcribed the text</change>
</revisionDesc>
```

Bibliography

TEI Consortium. “TEI P5: Guidelines for Electronic Text Encoding and Interchange,” August 19, 2020. https://www.tei-c.org/guidelines/P5/.