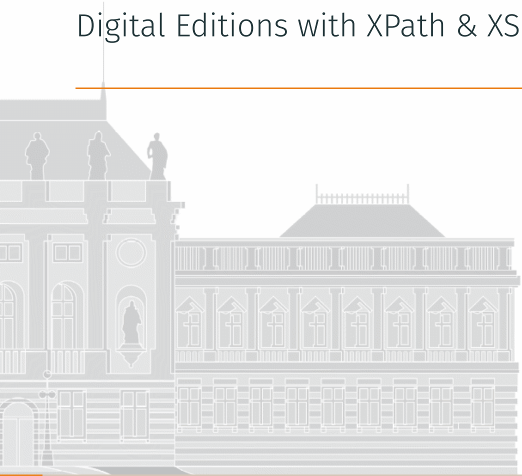
ZENTRUM FÜR
INFORMATIONSMODELLIERUNG
AUSTRIAN CENTRE FOR
DIGITAL HUMANITIES

KARL-FRANZENS-UNIVERSITÄT GRAZ
UNIVERSITY OF GRAZ

Beyond TEI

Digital Editions with XPath & XSLT for the Web & in LaTeX



Sarah Lang

Harvard, April/May 2022

Overview

1. The workshop
2. Annotating with XML markup
3. Text Encoding Initiative

The workshop

Goals

1. get to know XPath & XSLT (and learn how to use it)
2. understand the role of XML/TEI, XPath and XSLT in Digital Editing
3. be able to use XSLT to generate HTML and LaTeX output from TEI
4. Two days isn't enough for you to master XSLT!

Schedule

Day 1, morning XML, TEI and Digital Editing → repetition of the basics, making sure we're all on the same page, understanding why we're even learning XSLT.

Day 1, afternoon Navigating XML documents using XPath, introduction to HTML (& Bootstrap) and LaTeX (& *reledmac*)

Day 2 Transforming XML documents into HTML & LaTeX output formats using XSLT

Single point of entry for all workshop-related materials: LaTeX Ninja blogpost & Github Repository (‘additional resources’ directory)

Introductions

Please introduce yourselves!

1. Name, pronouns, field/topic of study
2. Why did you come to this workshop?
3. Previous experience with Digital Humanities (DH) or editing?

Contact

- @SarahALang_ @latex_ninja
- sarahalang.com latex-ninja.com
- @ sarah.lang@uni-graz.at

Sarah Lang (she/they)

- originally from Germany, now in Graz (Austria)
- Studied Latin, French & History (teacher's education) in Graz & Montpellier (France), then Archaeology Bachelor, Master in Religious Studies & Philosophy
- got a DH certificate & started working at Zentrum für Informationsmodellierung (ZIM) / Centre for Information Modelling in Graz
    - Moral Weeklies/Spectators → gams.uni-graz.at/mws
    - Graz Repository of Ancient Fables (GRaF) → gams.uni-graz.at/graf
    - *PhD thesis:* Decoding alchemical *Decknamen* digitally. A Polysemantic Annotation and Machine Reasoning Algorithm for the Corpus of Iatrochymist Michael Maier (1568–1622)
- *Now:* teaching in Graz, Passau & Vienna; PostDoc in Graz. *Research interests:* history of science (alchemy), Neo-Latin, text mining and computer vision

Annotating with XML markup

TEI, now what?

Why are we doing this workshop? The motivation from our abstract:

- […] the Text Encoding Initiative (TEI) for XML has become the gold standard for scholarly editions of texts.
- …

Goals for the next session

1. wait, what was…
   - ✗ XML?
   - ✗ TEI?
   - ✗ How do I use the TEI for digital editing?

XML: eXtensible Markup Language

- W3Schools Tutorial
- paradigm of the separation of form and content
- XML is a metalanguage

.XML

- RSS, SOAP, XAML
- MathML, GraphML
- XHTML
- RDF
- KML
- Scalable Vector Graphics (SVG)

> “ **Extensible Markup Language (XML)** is a **markup language** and file format for storing, transmitting, and reconstructing arbitrary data. It defines a **set of rules for encoding documents** in a format that is **both human-readable and machine-readable.** (Wikipedia) ”

XML rules

XML can be checked for **validity** (validation if it complies with a standard) and **well-formedness** (following the rules of XML) → will only be parsed if well-formed. Thus: Heed thy error messages!

There are rules on how elements can be named (you can look them up if relevant or will get informed by an error message).

`<key>value</key>` . XML as a key value notation

Rules

- Hierarchical nesting below the root
- exactly one root element, i.e. one out-most russion doll
- start and end tag
- tag names are case-sensitive (!)
- empty elements allowed (& can be shortened)

Minimal example

```xml
<?xml version="1.0" ?>
<root>
  <element attribute="value">
    content
  </element>
  <!-- comment -->
</root>
```

XML rules i

Prolog

`<xml version="1.0" encoding="utf-8">`	XML deklaration
`<?xsl-stylesheet type="text/xsl" href="mein.xsl"?>`	processing instructions (optional)

you can include document models (optional)
DTD, XML Schema, RelaxNG, Schematron

entities ‘protected’ characters because they have a meta meaning in XML like:

`&lt;`	<
`&gt;`	>
`&amp;`	&

XML family and vocabularies

XML structured description of data
XPath navigating xml documents
XML Schema strict data model
XSL Extensible Stylesheet Language
XSLT XSL-Transformations, i.e. transforming XML documents
XSL-FO formatted output (e.g. print)
XQuery query langauge for XML databases
and more

- **(X)HTML** Hypertext Markup Language
- **EAD** Encoded Archival Description
- **TEI** Text Encoding Initiative
- **CEI** Charters Encoding Initiative
- **MEI** Music Encoding Initiative
- **LIDO** Lightweight Information Describing Objects (describing museum or collection objects)
- **SVG** Scalable Vector Graphics
- **KML** Keyhole Markup Language (Geographie)
- **MathML**
- **CML** Chemical Markup Language, …

Text Encoding Initiative

TEI Primer

Text Encoding Initiative

.XML

XML-Standard, i.e. convention on how to use XML so that resulting data will be interoperable between different projects. (founded in 1987, consortium since 2000)

> “The Text Encoding Initiative (TEI) is a text-centric community of practice in the academic field of digital humanities, operating continuously since the 1980s. The community currently runs a mailing list, meetings and conference series, and maintains the TEI technical standard, a journal, a wiki, a GitHub repository and a toolchain. (Wikipedia)”

TEI minimal example

```
<TEI> <!-- root element -->
    <teiHeader>
      <!-- author, title, dating,
           sources, edition rules, etc.
    </teiHeader>
    <text> ... </text>
</TEI>
```

Resources

- Learn TEI
- Teach Yourself
- P5 = 5. Proposal
- MEI for music
- CEI for charters
- http://www.tei-c.org/

TEI Header

fileDesc = bibliographical description of the contents of the document

encodingDesc = connection of electronic document to source (i.e. transcription rules, etc.)

```
<TEI> <!-- root element -->
    <teiHeader>
        <fileDesc> ... </fileDesc> <!-- obligatory -->
        <encodingDesc> <!-- optional -->
        <profileDesc> <!-- optional -->
        <revisionDesc> <!-- optional -->
    </teiHeader>
    <text> ... </text>
</TEI>
```

profileDesc = decribes all non-bibliogaphical aspects of the text (i.e. creation, languages)

revisionDesc = tracks changes in the digital document

Using TEI i

Gentle Intro to XML

TEI Core

- **div** (division)
- **p** (paragraph)
- **head** (heading)
- **lb** (linebreak)
- **pb** (page break / beginning)
- **hi** (highlight)
- **l** (line)
- **lg** (line group)
- **list**
- **item**
- **listBibl**
- **bibl** (bibliographical information)

Attributes

- **@n** (label)
- **@type** (typing)
- **xml:id** (unique identifier)
- **xml:lang** (language)
- **@rend** (rendering)
- @ana (interpretation)

Using TEI ii

```
<foreign xml:lang="en">word</foreign>
<term type="homonym"/>
<date when="2009-04-27"/>
<time when="12:00:00"/>
<name type="person"/>
<persName n="Caesar" xml:id="#44BC">Caesaris</persName>
<!-- or -->
<persName key="ID.01.208"/>
<person/>
<emph/> <hi rend="italic">italic text</hi>
<seg/> <abbr type="acronym"/>
<placeName xml:id="#Whitby">Abbey</placeName>
```

Using TEI iii

Name spaces identified via URI

<präfix:name> e.g. *<tei:p>* (‘I mean the *<p>* according to the TEI standard.’)

declaration <element xmlns=“URI“> ...
<prefix:element xmlns:prefix=“URI“> ...
e.g.

```
<tei:p xmlns:tei=“http://www.tei-c.org/ns/1.0“>...
```

TEI is organized in modules

Acts of speech (reference) if speaker name is mentioned, otherwise TEIs ‘said’:

```
<sp who="#person">
    <speaker>1.</speaker> <p>Bla, bla, bla.</p>
</sp>

<said who="#Adolphe">- Alors, Albert, quoi de neuf?</said>
```

Letters in TEI (reference)

```
<div type="letter" n="14">
    <head>Letter XIV: Miss Clarissa Harlowe to Miss Howe</head>
        <opener>
            <dateline>Thursday evening, March 2.</dateline>
            <salute>Hallo,</salute>
        </opener>
    <p>On Hannah's depositing my long letter ...</p>
    <closer>
        <salute>Yours more than my own,</salute>
        <signed>Clarissa Harlowe</signed>
    </closer>
</div>
```

Names, Dates, Places

Named Entities & Indirekte reference

TEI 13: Names, Dates, People, Places

- **persName** for personal names, **<rs>** for *referring string* when mentioned indirectly (‘he’, ‘the woman’, etc.) →**@key** or **@ref** to specify who it is (reference).
- **forename**
- **surname**
- **roleName** (z.B. ‘king’)
- **genName** (‘the Younger’)
- **addName**
- **nameLink** (‘von’).

```xml
<name role="writer" type="person"
ref="http://d-nb.info/gnd/118540238">
Goethe</name>
<person>
  <addName type="Former">Murray</addName>
  <forename>Wilhelmina</forename>
  <addName type="nickname">Mina</addName>
</person>
```

Metadata in the TEI Header i

```
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>
<!-- title of the resource -->
   </title>
  </titleStmt>
  <publicationStmt>
   <p>
<!-- Information about distribution of the resource -->
   </p>
  </publicationStmt>
  <sourceDesc>
   <p>
<!-- Information about source from which the resource derives -->
   </p>
  </sourceDesc>
 </fileDesc>
</teiHeader>
```

Metadata in the TEI Header ii

The title and author in the *<titleStmt>* isn't the bibliographic data from the source! It describes the digital document and its authors or editors.

If you want to desribe your source documents, you need elements like *<sourceDesc>* or *<msDesc>*:

```
<sourceDesc>
 <bibl>
  <title level="a">The Interesting story of the Children in the Wood</title>. I
 <author>Victor E Neuberg</author>, <title>The Penny Histories</title>.
 <publisher>OUP</publisher>
  <date>1968</date>. </bibl>
</sourceDesc>
```

```
<sourceDesc>
 <p>Born digital: no previous source exists.</p>
</sourceDesc>
```

Metadata in the TEI Header iii

```xml
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>Thomas Paine: Common sense, a
       machine-readable transcript</title>
   <respStmt>
    <resp>compiled by</resp>
    <name>Jon K Adams</name>
   </respStmt>
  </titleStmt>
  <publicationStmt>
   <distributor>Oxford Text Archive</distributor>
  </publicationStmt>
  <sourceDesc>
   <bibl>The complete writings of Thomas Paine, collected and edited
       by Phillip S. Foner (New York, Citadel Press, 1945)</bibl>
  </sourceDesc>
 </fileDesc>
</teiHeader>
```

<msDesc>

```xml
<msDesc>
 <msIdentifier>
  <settlement>Oxford</settlement>
  <repository>Bodleian Library</repository>
  <idno type="Bod">MS Poet. Rawl. D. 169.</idno>
 </msIdentifier>
 <msContents>
  <msItem>
   <author>Geoffrey Chaucer</author>
   <title>The Canterbury Tales</title>
  </msItem>
 </msContents>
 <physDesc>
  <objectDesc>
   <p>A parchment codex of 136 folios, measuring approx
       28 by 19 inches, and containing 24 quires.</p>
   <p>The pages are margined and ruled throughout.</p>
   <p>Four hands have been identified in the manuscript: the first 44
       folios being written in two cursive anglicana scripts, while the
       remainder is for the most part in a mixed secretary hand.</p>
  </objectDesc>
 </physDesc>
</msDesc>
```

<titlePage>

To describe a title page (e.g. early modern print copperplates, etc.), use *<titlePage>*:

```
<titlePage>
 <docTitle>
  <titlePart type="main">THOMAS OF Reading.</titlePart>
  <titlePart type="alt">OR, The sixe worthy yeomen of the West.</titlePart>
 </docTitle>
 <docEdition>Now the fourth time corrected and enlarged</docEdition>
 <byline>By T.D.</byline>
 <figure>
  <head>TP</head>
  <p>Thou shalt labor till thou returne to duste</p>
  <figDesc>Printers Ornament used by TP</figDesc>
 </figure>
 <docImprint>Printed at <name type="place">London</name> for <name>T.P.</name>
  <date>1612.</date>
 </docImprint>
</titlePage>
```

<front>

You might also need *<front>* (front matter): contains any prefatory matter (headers, abstracts, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.

```
<front>
 <epigraph>
  <quote>Nam Sibyllam quidem Cumis ego ipse oculis meis vidi in ampulla
     pendere, et cum illi pueri dicerent: <q xml:lang="grc">Σίβυλλα τί
       θέλεις</q>; respondebat illa: <q xml:lang="grc">ἀποθανεῖν θέλω.</q>
  </quote>
 </epigraph>
 <div type="dedication">
  <p>For Ezra Pound <q xml:lang="it">il miglior fabbro.</q>
  </p>
 </div>
</front>
```

TEI practice!

Fill out the `teiHeader` or `msDesc`.

Use websearch (‘tei msDesc’) to learn how to use new elements (overview plus examples view).