TEI Customization

Martina Scholger

CLARIAH-AT
TEI Introductory School
10.-13. September 2019

Venue:
Centre for Information Modelling (ZIM)
Elisabethstraße 59/III, SR 81.31
This session is about

• Why customize?
• The TEI structure
• ROMA
• ODD „One Document Does it All“
• ODD by hand
TEI Customization means the process of creating a TEI schema by

- **customizing** the TEI to your needs and
- **documenting** your encoding practice, especially the deviations from the standard

[TEI] is standardisation by not saying “Do what I do” but instead by saying “Do what you need to do but tell me about it in a language I understand” (Cummings, 2013)
TEI customization for the letter
What is a schema?

- Inventory of elements
- Relations of elements to each other
- Rules in which context elements and attributes are allowed
- Realization: DTD, XSD, RNG
Recap

Well-formed:
- Everything is properly delimited
- A single (one and only one) root element
- No overlap

Validity:
- Correct **vocabulary**: all elements and attributes actually exist in that language
- Correct **grammar**: elements are used in the right place, in the right order
Test yourself!

- A letter must begin with a date, followed by a salutation, at least one paragraph, and a signature.
- A paragraph, signature, or salutation may contain names and transcribed text.

```xml
<letter>
  <date>2012-02-12</date>
  <salutation>Dear Harry,</salutation>
  <paragraph>Symmetrical dates are so elegant!</paragraph>
  <signature>Yours, Larry</signature>
</letter>

<letter>
  <salutation>Dear <name>Larry</name>, </salutation>
  <paragraph>Dates just reveal your enslavement to the space-time continuum.</paragraph>
  <signature>Yours, <name>Harry</name></signature>
</letter>

<letter>
  <salutation>Dear <name>Harry</name>, </salutation>
  <paragraph>My triskaidekaphobia is acting up.</paragraph>
  <signature>Yours, Larry</signature>
  <date>2012-02-13</date>
</letter>
```
Why do I need a schema?

- The tei_all schema is very flexible and often allows multiple ways for one phenomenon. Example: personal names

```xml
<persName>
  Augustinus
</persName>

<name type="person">
  Augustinus
</name>

<rs type="person">
  he [Augustinus]
</rs>
```
Schemas provided by the TEI

- tei_all
- tei_lite (light and basic encoding)
- tei_simplePrint (early modern printed material)
- tei_speech (speech and spoken texts)
- Also available in oXygen – see the TEI Framework
TEI Infrastructure: Modules

- **23 Modules**: elements and attributes are grouped according to their purpose

- Example: `<persName>` is a member of the module?

TEI Infrastructure: Elements

- Currently, 574 elements and a large number of attributes
- An element belongs to one! module
- Elements are also grouped in classes
- A class groups elements or attributes that behave in the same way and that tend to go in the same place.
TEI Infrastructure: Model Classes

- Groups of elements with similar structural roles, access to similar structural locations
- Elements belong to one or more classes in order to be able to appear in those locations
- Classes of elements are named `model.the class` (e.g. `model.global`)

- Example:
  `model.biblPart` is a set of elements that may appear within `<bibl>`. Which elements are a member of `model.biblPart`?
Groups of attributes with related functions
An elements joins the class to get access to its attributes
Classes of attributes are named `att.the class` (e.g. `att.global`)

Example:
- `att.datable.w3c` is a set of attributes to describe dates
- The element `<faith>` is a member of the class `att.datable.w3c`
- As a result the following attributes can be used on `<faith>`: `@when`, `@notBefore`, `@notAfter`, `@from`
Datatypes

• The values which attributes may take in a TEI schema are defined by reference to a TEI datatype specification.
• Mostly from W3C schema datatypes
## Exercise: class assignment

<table>
<thead>
<tr>
<th></th>
<th>Classes</th>
<th>Elements and Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>att.typed</td>
<td><strong>A</strong> &lt;closer&gt;, &lt;postscript&gt;, &lt;signed&gt;, &lt;trailer&gt;</td>
</tr>
<tr>
<td>2</td>
<td>model.divBottomPart</td>
<td><strong>B</strong> &lt;abbr&gt;, &lt;am&gt;, &lt;corr&gt;, &lt;ex&gt;, &lt;expan&gt;, &lt;orig&gt;, &lt;reg&gt;, &lt;seg&gt;, &lt;sic&gt;, &lt;unclear&gt;</td>
</tr>
<tr>
<td>3</td>
<td>att.measurement</td>
<td><strong>C</strong> &lt;bibl&gt;, &lt;biblFull&gt;, &lt;biblStruct&gt;, &lt;listBibl&gt;, &lt;msDesc&gt;</td>
</tr>
<tr>
<td>4</td>
<td>model.choicePart</td>
<td><strong>D</strong> @type, @subtype</td>
</tr>
<tr>
<td>5</td>
<td>att.global</td>
<td><strong>E</strong> &lt;addSpan&gt;, &lt;damageSpan&gt;, &lt;delSpan&gt;, &lt;gap&gt;, &lt;space&gt;, &lt;alt&gt;, &lt;altGrp&gt;, &lt;certainty&gt;, &lt;fLib&gt;, &lt;fs&gt;, &lt;fVLib&gt;, &lt;index&gt;, &lt;interp&gt;, &lt;interpGrp&gt;, &lt;join&gt;, &lt;joinGrp&gt;, &lt;link&gt;, &lt;linkGrp&gt;, &lt;listTranspose&gt;, &lt;precision&gt;, &lt;respons&gt;, &lt;span&gt;, &lt;spanGrp&gt;, &lt;substJoin&gt;, &lt;timeline&gt;, &lt;incident&gt;, &lt;kinesic&gt;, &lt;pause&gt;, &lt;shift&gt;, &lt;vocal&gt;, &lt;writing&gt;, &lt;anchor&gt;, &lt;cb&gt;, &lt;fw&gt;, &lt;gb&gt;, &lt;lb&gt;, &lt;milestone&gt;, &lt;pb&gt;, &lt;note&gt;, &lt;witDetail&gt;, &lt;figure&gt;, &lt;metamark&gt;, &lt;notatedMusic&gt;</td>
</tr>
<tr>
<td>6</td>
<td>model.biblLike</td>
<td><strong>F</strong> @unit, @quantity, @commodity</td>
</tr>
<tr>
<td>7</td>
<td>model.milestoneLike</td>
<td><strong>G</strong> @xml:id, @n, @xml:lang, @rend, @style, @rendition, @xml:base, @xml:space, @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select, @ana, @facs, @change</td>
</tr>
<tr>
<td>8</td>
<td>model.global</td>
<td><strong>H</strong> &lt;anchor&gt;, &lt;cb&gt;, &lt;fw&gt;, &lt;gb&gt;, &lt;lb&gt;, &lt;milestone&gt;, &lt;pb&gt;</td>
</tr>
</tbody>
</table>
Let’s check the Guidelines

What is an ODD?

• “One document does it all”
• An ODD file is a TEI file, with specific elements for documentation
What can I customize?

• Select modules
• Exclude elements that you don’t need
• Constrain attribute values
• Change attributes
• Constrain structure
• Manipulate grouping of elements (classes)
• Replace default definition with my customized definition
• Produce an internationalized version
First Customization

```xml
<schemaSpec ident="firstCustomization">
  <moduleRef key="tei"/>
  <moduleRef key="header"/>
  <moduleRef key="core"/>
  <moduleRef key="textstructure"/>
</schemaSpec>
```
How to create a customization?

• Writing by hand
  • ODD Customization in Oxygen Framework

• Using Roma – customization tool
  • Roma - beta (https://romabeta.tei-c.org/)
  • Roma (https://roma.tei-c.org/)
Roma interface

https://romabeta.tei-c.org/
Exercise in Roma: getting started

- Open Roma: [https://romabeta.tei-c.org/](https://romabeta.tei-c.org/)
- Choose a new preset (e.g. TEI Minimal)
- Change „Title“ and „Author“
- Search for the module namesdates and add it
- Delete the elements <analytic> and <citedRange> from the core module
- Delete the elements <revisionDesc> and <encodingDesc> from the header module
- Use the following download options
  - Save the ODD customization for further customizations (.odd)
  - Save the schema as RelaxNG compact (.rnc)
Exercise: Continue in Oxygen

- Open ODD in oXygen and investigate the file
- Open letter.xml in oXygen
- Assign schema to the letter
Assign schema to your XML (Oxygen)

Icon “Assign schema...” in the toolbar or Document > Schema > Assign Schema
Exercise in Roma: Add another module and change attributes

- Add the missing module (linking)
- Search for the element &lt;name&gt; and select it
- Select attributes
- Modify the type attribute of the &lt;name&gt; element
- Provide a closed list with the values “org”, “person”, and “place”.
- Provide a short description for the values
Writing ODD by hand

• Instead of using Roma, we can write our ODD by hand
• Open ODD Customization in Oxygen (Framework)
• Use transformation scenario in Oxygen to generate the schema
Adding more modules

<body>
  <schemaSpec ident="demo01">
    <!-- required modules -->
    <moduleRef key="tei"/>
    <!-- has underpinnings of TEI class system -->
    <moduleRef key="core"/>
    <!-- has <title>, <p>, <emph>, etc. -->
    <moduleRef key="header"/>
    <!-- has <teiHeader>, <fileDesc>, etc. -->
    <moduleRef key="textstructure"/>
    <!-- has <TEI>, <text>, <div>, etc. -->
    <!-- optional modules -->
    <moduleRef key="namesdates"/>
    <moduleRef key="transcr"/>
    <moduleRef key="analysis"/>
  </schemaSpec>
</body>
ODD: delete elements

• Delete elements from selected modules:

  <moduleRef key="namesdates" except="addName affiliation bloc climate"/>
ODD: Documentation

You can add documentation like this:

```html
<p>The <att>type</att> attribute on the <gi>div</gi> element may be used to supply a conventional name for this category of text division in order to distinguish them. Typical values might be <val>book</val>, <val>chapter</val>, <val>section</val>, <val>part</val>, <val>poem</val>, <val>song</val>, etc. TEI simplePrint does not constrain the range of values that may be used here. </p>
```

- <gi> element
- <att> attribute
- <val> attribute value

See Modul 22 for more „Documentation Elements“

Which elements do I need?

XPath expression:
distinct-values(/TEI//name())
Credits

• Woman Writers Project (Syd Bauman, Julia Flanders): https://www.wwp.northeastern.edu/outreach/resources/customization.html
• TEI Technical Council
• Institut für Dokumentologie und Editorik
• ... for sharing workshop materials
Links

- Roma/ODD
  - Customizing the TEI with Roma
  - Getting Started with P5 ODDs
    https://tei-c.org/guidelines/customization/getting-started-with-p5-odds/
  - Lou Burnard, ODD chaining for Beginners:
    http://teic.github.io/TCW/howtoChain.html
  - ODD chaining Wiki page:
    https://wiki.tei-c.org/index.php/ODD_chaining