



TEI Customization

Martina Scholger

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

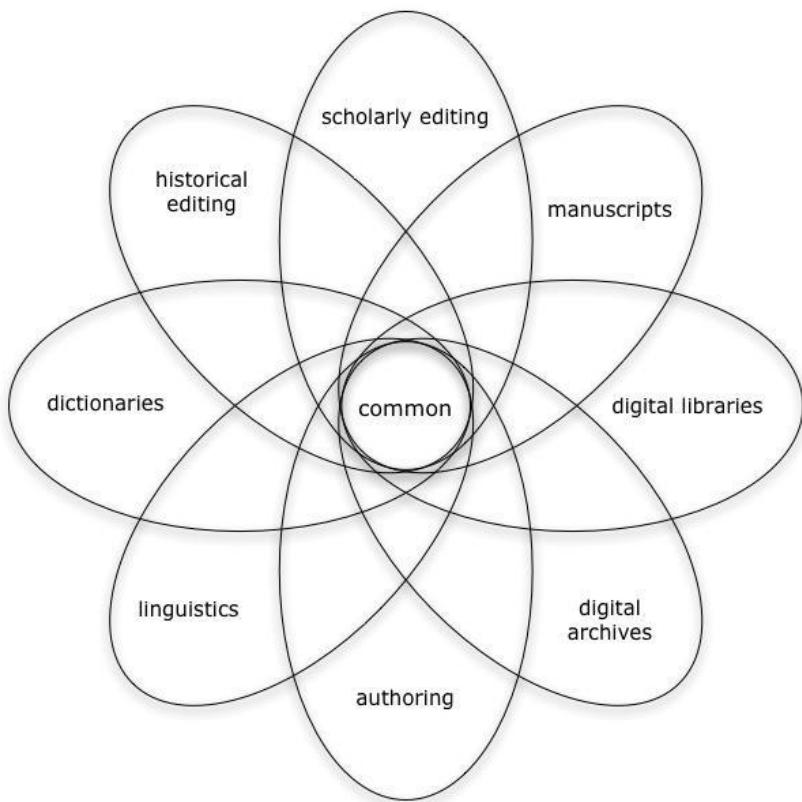
Venue:

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 modules

transcr

tagdocs

textstructure

iso-fs

linking

header

dictionaries

nets

certainty

drama

spoken

gaiji

core

msdescription

tei

verse

namesdates

figures

corpus

textcrit

analysis

TEI customization for the letter

transcr

tagdocs

textstructure

linking

header

dictionaries

certainty

core

drama

spoken

gaiji

namesdates

msdescription

tei

verse

figures

corpus

textcrit

analysis

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

```
<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

```
<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)
- More customizations:
<http://www.tei-c.org/Guidelines/Customization/>
- 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?
- <https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STMA>

- + 1 [The TEI Infrastructure](#)
- + 2 [The TEI Header](#)
- + 3 [Elements Available in All TEI Documents](#)
- + 4 [Default Text Structure](#)
- + 5 [Characters, Glyphs, and Writing Modes](#)
- + 6 [Verse](#)
- + 7 [Performance Texts](#)
- + 8 [Transcriptions of Speech](#)
- + 9 [Dictionaries](#)
- + 10 [Manuscript Description](#)
- + 11 [Representation of Primary Sources](#)
- + 12 [Critical Apparatus](#)
- + 13 [Names, Dates, People, and Places](#)
- + 14 [Tables, Formulæ, Graphics and Notated Music](#)
- + 15 [Language Corpora](#)
- + 16 [Linking, Segmentation, and Alignment](#)
- + 17 [Simple Analytic Mechanisms](#)
- + 18 [Feature Structures](#)
- + 19 [Graphs, Networks, and Trees](#)
- + 20 [Non-hierarchical Structures](#)
- + 21 [Certainty, Precision, and Responsibility](#)
- + 22 [Documentation Elements](#)
- + 23 [Using the TEI](#)

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`?

TEI Infrastructure: Attribute classes

- Groups of attributes with related functions
- An element 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
- More on datatypes: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#DTYPES>

Exercise: class assignment

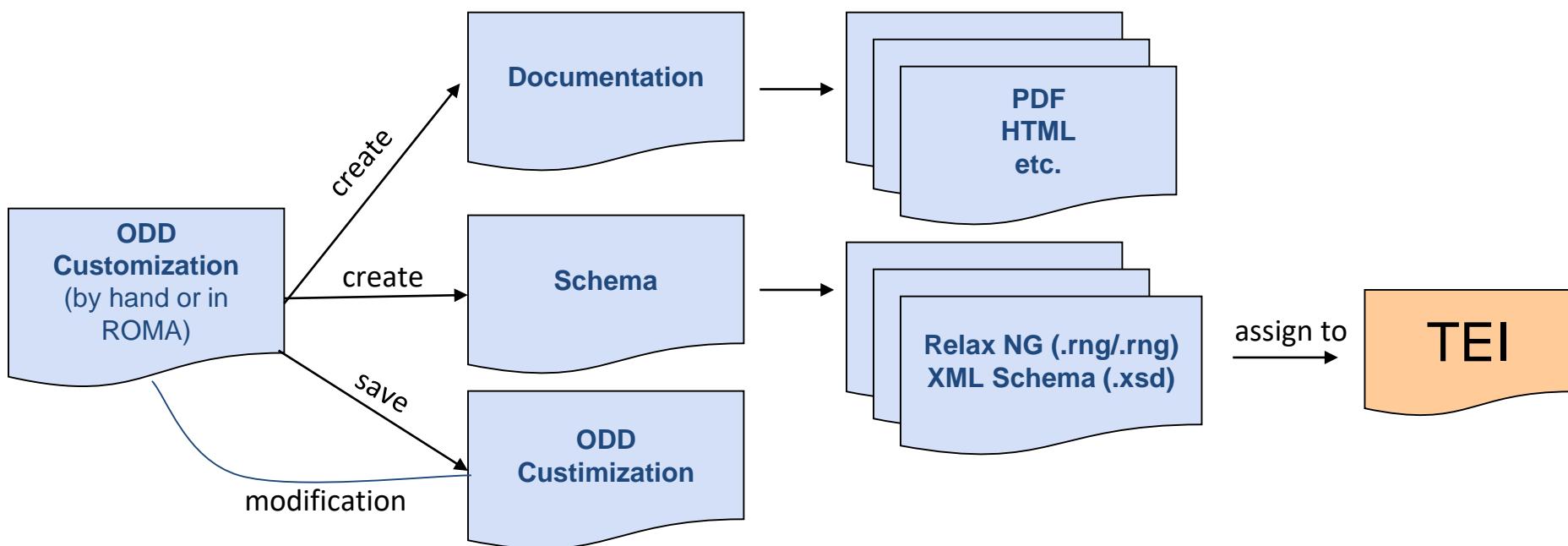
	Classes		Elements and Attributes
1	att.typed	A	<closer>, <postscript>, <signed>, <trailer>
2	model.divBottomPart	B	<abbr>, <am>, <corr>, <ex>, <expan>, <orig>, <reg>, <seg>, <sic>, <unclear>
3	att.measurement	C	<bibl>, <biblFull>, <biblStruct>, <listBibl>, <msDesc>
4	model.choicePart	D	@type, @subtype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

Let's check the Guidelines

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

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

```
TEI text body schemaSpec moduleRef
1  <?xml version="1.0" encoding="UTF-8"?>
2  <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_odds.rng" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
3  <?xml-model href="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_odds.rng" type="application/xml"
4    schematypens="http://purl.oclc.org/dsdl/schematron"?>
5  <TEI xmlns="http://www.tei-c.org/ns/1.0">
6    <teiHeader>
7      <fileDesc>
8        <titleStmt>
9          <title>My first ODD</title>
10         <author>Martina Scholger</author>
11       </titleStmt>
12       <publicationStmt>
13         <p>Publication Information</p>
14       </publicationStmt>
15       <sourceDesc>
16         <p>Information about the source</p>
17       </sourceDesc>
18     </fileDesc>
19   </teiHeader>
20   <text>
21     <body>
22       <schemaSpec ident="myTEI">
23         <moduleRef key="core"></moduleRef>
24         <moduleRef key="tei"></moduleRef>
25         <moduleRef key="header"></moduleRef>
26         <moduleRef key="textstructure"></moduleRef>
27       </schemaSpec>
28     </body>
29   </text>
30 </TEI>
31
```

```
<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/>

The screenshot shows the 'Roma - ODD Customization' interface. At the top, there is a blue header bar with the title 'Roma - ODD Customization' and 'Version 0.2.1'. Below the header, there are two buttons: 'SELECT ODD' and 'UPLOAD ODD'. A large section titled 'Select ODD' is centered. It contains a dropdown menu with the options 'Choose a preset' and 'TEI All (customize by reducing TEI)'. At the bottom of this section is a large blue button labeled 'START'.

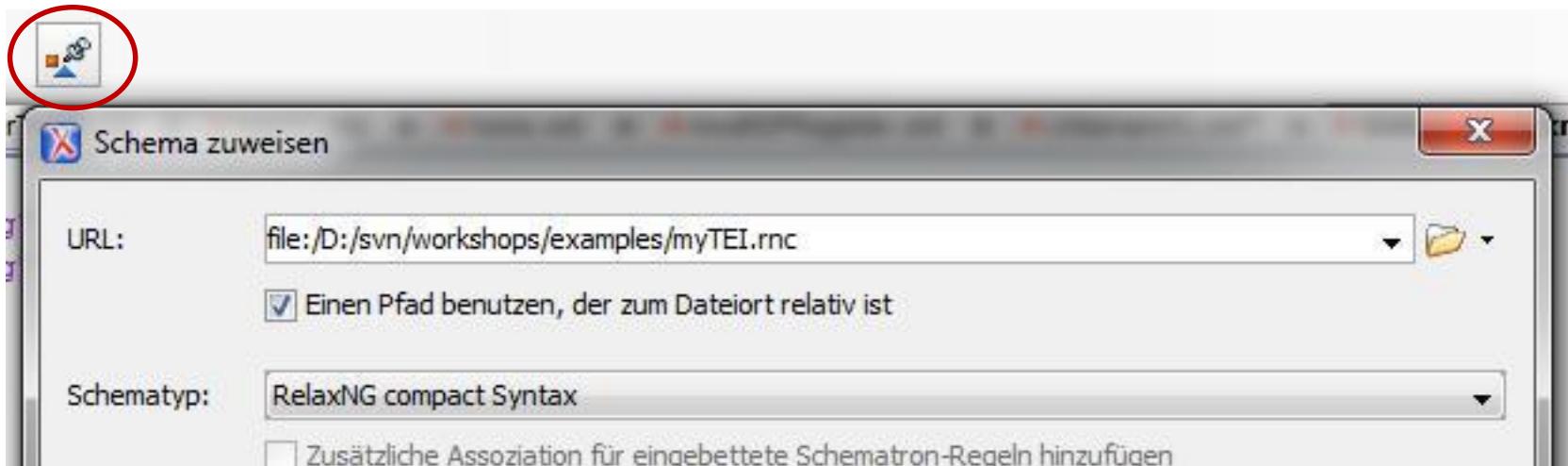
Exercise in Roma: getting started

- Open Roma: <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 <name> and select it
- Select attributes
- Modify the type attribute of the <name> 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:

<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“

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>

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
<https://tei-c.org/guidelines/customization/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