

Coding the Sources

Digitales Edieren in den Geisteswissenschaften

Online Summerschool und Workshop

Lübeck/Online, 17.08. – 27.08.2020



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



ODD Customization

Die Erstellung eines projektspezifischen TEI-Schemas

Martina Scholger

Zentrum für Informationsmodellierung – Austrian Centre for Digital Humanities,
Universität Graz
martina.scholger@uni-graz.at

Lübeck/Online, 26.08.2020



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



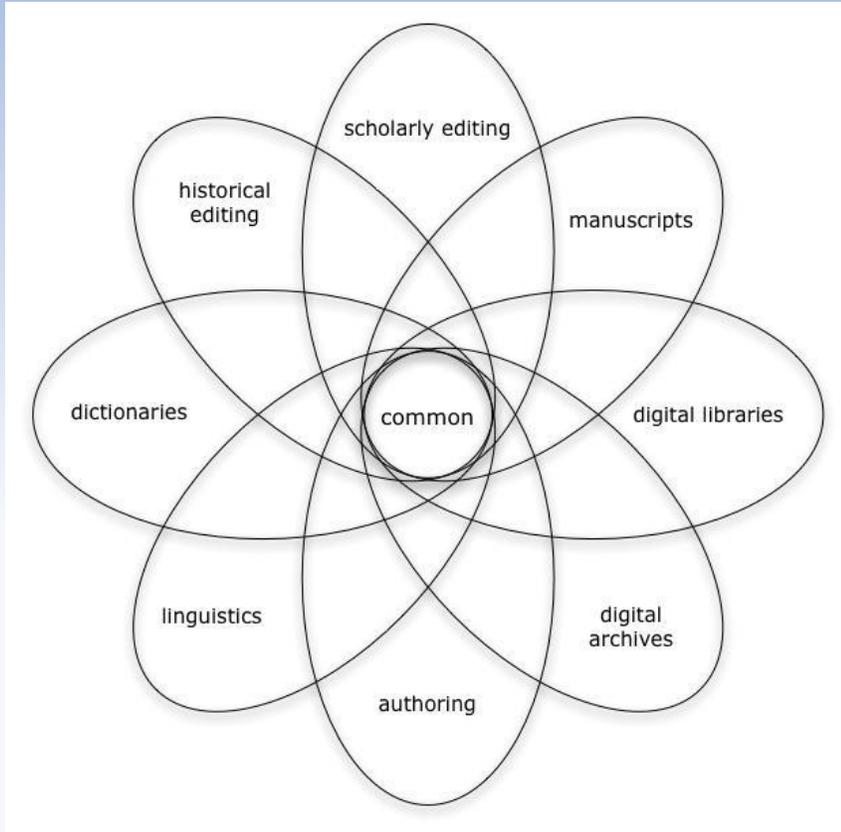
ZM ZENTRUM FÜR
INFORMATIONSMODELLIERUNG
AUSTRIAN CENTRE FOR
DIGITAL HUMANITIES



Projektspezifisches Schema erstellen

- TEI Customization – wozu?
- Wiederholung: Schema
- Wiederholung: TEI Infrastruktur
- TEI Customization mit Roma (gemeinsame Übung)
- ODD Spezifikation im Detail
- ODD in Oxygen modifizieren

TEI Customization

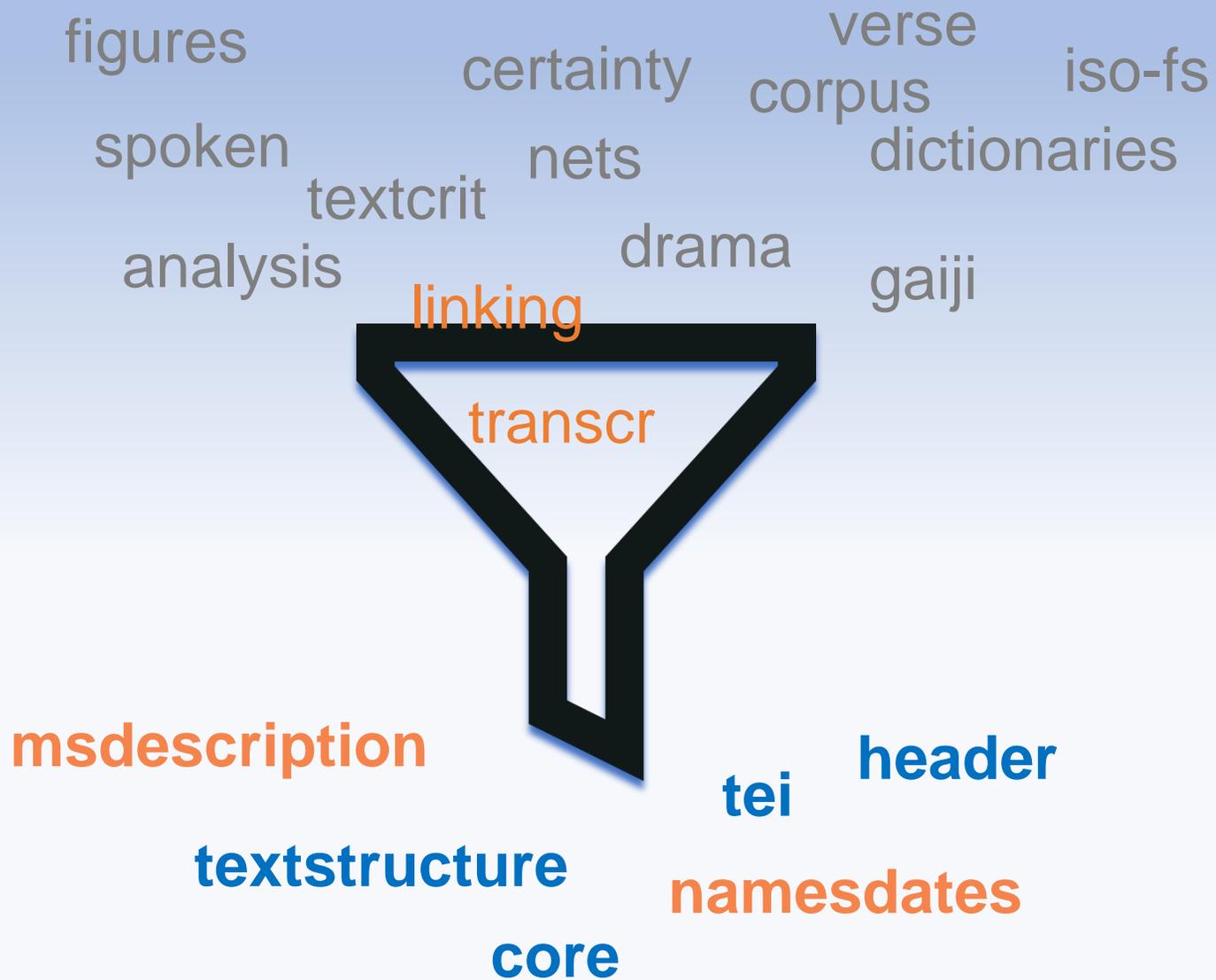


© Woman Writers Project

- Meint die Erstellung eines TEI-Schemas
 - das an projektspezifische Bedürfnisse **angepasst** ist und
 - das die Kodierungspraxis und die Abweichungen vom Standard **dokumentiert**.

[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



Wiederholung: Was ist ein Schema?

- Formale Deklaration von XML-Konstrukten, die in einem Dokument erforderlich, zulässig oder unzulässig sind.
- Inventar an Elementen.
- Relationen von Elementen zueinander.
- Regeln, in welchen Kontexten Elemente und Attribute erlaubt sind.
- Umsetzung: DTD, XSD, RNG, Schematron

- Jede `<Person>` kann eine beliebige Anzahl an `<Beruf>`en, aber nur ein `<Geburtsdatum>` haben.

Wiederholung: Wohlgeformtheit und Gültigkeit

Wohlgeformtheit (well-formed):

- Elemente mit Start- und Endtag
- Attribute unter Anführungszeichen
- Ein einziges Wurzelement
- Keine Überlappung der Elemente

Gültigkeit (valid):

- Korrektes **Vokabular**: alle verwendeten Elemente und Attribute existieren in der verwendeten Kodierungssprache
- Korrekte **Grammatik**: Elemente werden an der richtigen Stelle (Hierarchie), in der richtigen Reihenfolge verwendet.

Selbsttest (welcher Ausschnitt ist richtig?)

- Ein Brief muss mit einem Datum beginnen, gefolgt von einer Begrüßung, muss zumindest einen Absatz haben und eine Signatur
- Absatz, Begrüßung und Signatur können sowohl Text als auch <name>-Elemente enthalten

```
<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>
  <date>2012-02-12</date>
  <salutation>Dear Harry,</salutation>
  <signature>Yours, Larry</signature>
</letter>
```

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

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

Wozu brauche ich ein Schema?

- Die TEI ist umfangreich, flexibel und modular. Es existieren oft mehrere Möglichkeiten, um ein Phänomen zu kodieren, z.B. Personennamen

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

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

```
<rs type="person">  
  he [Augustinus]  
</rs>
```

- Viele Funktionalitäten brauche ich gar nicht für mein Projekt
- Ich benötige zusätzliche Funktionalitäten
- Konsistenz der Kodierung wird mit einem Schema erhöht
- Dokument-Modell formal ausdrücken

Vordefinierte TEI-Schemata

- tei_all
- tei_lite (einfache Basiskodierung)
- tei_simplePrint (frühneuzeitliche Drucke)
- tei_speech (gesprochener Text)
- Weitere: <https://tei-c.org/guidelines/customization/>
- Auch in oXygen verfügbar – siehe TEI Framework

TEI Infrastruktur: Module

- 23 Module: Elemente und Attribute werden nach ihrem Verwendungszweck gruppiert.
- <https://tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#STMA>
- Zu welchem Modul gehört `<persName>` ?
- Zu welchem Modul gehört `<particDesc>` ?

- 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 Infrastruktur

- Derzeit 582 Elemente und 266 Attribute
- Ein Element gehört zu EINEM Modul.
- Elemente sind in Klassen (= model classes) gruppiert.
- Eine Klasse gruppiert Elemente und Attribute die ähnliche Eigenschaften aufweisen und üblicherweise im selben Kontext verwendet werden können.

TEI Infrastruktur: Model-Klassen

- Elemente gehören zu einer oder mehreren Klassen.
- Elemente können damit in gleichen Kontexten verwendet werden.
- Elementklassen werden als `model.theClassName` bezeichnet (z.B. `model.global`)
- Beispiel:
 - `model.biblPart` beinhaltet ein Set an Elementen, das innerhalb von `<bibl>` verwendet werden kann.
 - Welche Elemente sind Mitglieder der Klasse `model.biblPart`?

TEI Infrastruktur: Attributklassen

- Gruppiert Attribute mit ähnlichen Funktionen.
- Ein Element erhält über die Mitgliedschaft zu einer Klasse Zugriff auf die in der Klasse definierten Attribute.
- Attributklassen werden nach dem Muster `att.theClass` benannt (z.B. `att.global`).
- Beispiel:
 - `att.dateable.w3c` ist ein Attributset zur Beschreibung von Datumsangaben.
 - Das Element `<faith>` ist Mitglied der Klasse `att.dateable.w3c`.
 - Die folgenden Attribute können daher verwendet werden: `@when`, `@notBefore`, `@notAfter`, `@from`

TEI Infrastruktur: Datentypen

- Die Art des zulässigen Attributwerts wird über eine Referenz auf Datentypen spezifiziert.
- Die meisten TEI Datentypen werden vom [W3C Schema Datatypes](#) abgeleitet.
- Beispiele für Datentypen in TEI
 - teidata.count nicht-negative Ganzzahl
 - teidata.temporal.w3c Werte für Datums- und Zeitangaben
 - teidata.word einzelnes Wort oder Token
 - teidata.enumerated Aufzählung aus einer vordefinierten Liste
- Mehr über Datentypen: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html#DTYPES>

Übung: Klassenzugehörigkeit

	Klassen
1	att.typed
2	model.divBottomPart
3	att.measurement
4	model.choicePart
5	att.global
6	model.biblLike
7	model.milestoneLike
8	model.global

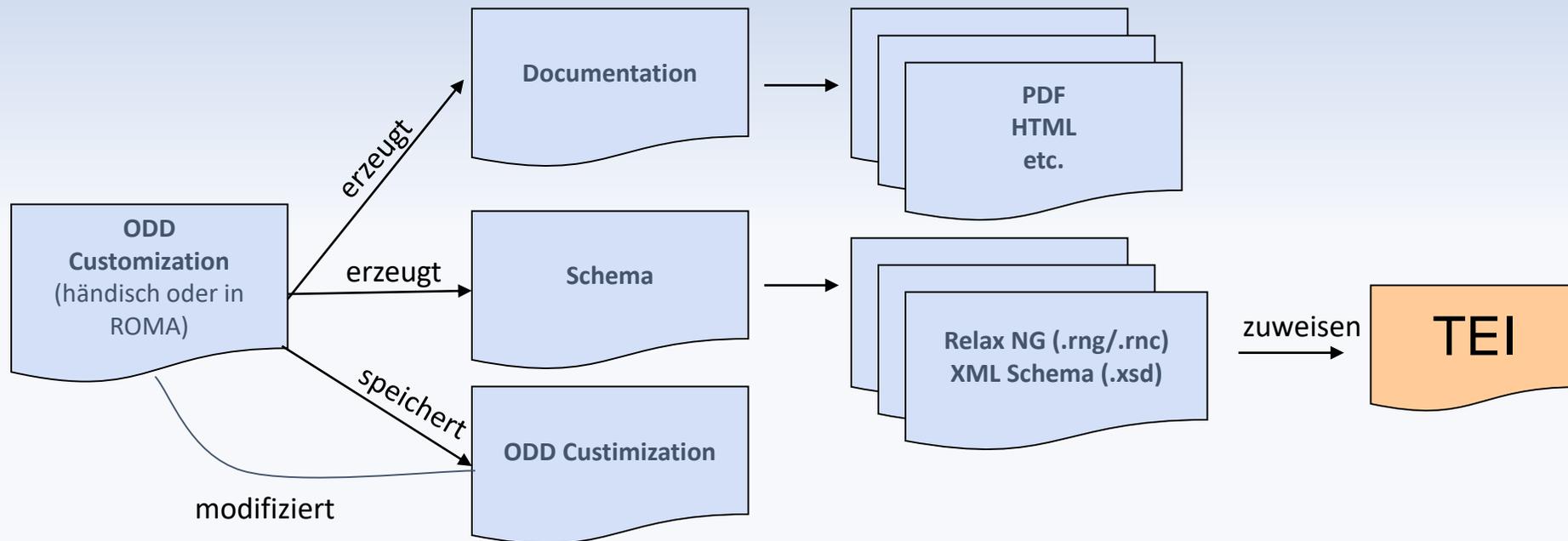
	Element- bzw. Attributzuordnungen
A	<closer>, <postscript>, <signed>, <trailer>
B	<abbr>, <am>, <corr>, <ex>, <expan>, <orig>, <reg>, <seg>, <sic>, <unclear>
C	<bibl>, <biblFull>, <biblStruct>, <listBibl>, <msDesc>
D	@type, @subtype
E	<addSpan>, <damageSpan>, <delSpan>, <gap>, <space>, <alt>, <altGrp>, <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>
F	@unit, @quantity, @commodity
G	@xml:id, @n, @xml:lang, @rend, @style, @rendition, @xml:base, @xml:space, @corresp, @synch, @sameAs, @copyOf, @next, @prev, @exclude, @select, @ana, @facs, @change
H	<anchor>, <cb>, <fw>, <gb>, <lb>, <milestone>, <pb>

ODD – One document does it all

- Ein Metaformat zur Generierung multipler Outputs
- Ein TEI Dokument, das Schemaspezifikationen enthält
- Eine Schemaspezifikation besteht aus:
 - einer formalen Dokumentation von verwendeten Modulen, Elementen und Attributen sowie deren Werteinschränkungen
 - einer deskriptiven Dokumentation – dazu werden Elemente aus dem TEI Modul 22 „Documentation Elements“ verwendet
 - <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/TD.html>

Anpassung des TEI Schemas (mit Roma)

- ODD ist ein TEI Dokument, mit spezifischen Dokumentationselementen



Anpassung des TEI Schemas (mit Roma)

- Module auswählen
- Elemente auswählen (include) oder ausschließen (exclude)
- Zulässige Attributwerte beschränken
- Inhaltsmodell anpassen
- Klassenzuordnung ändern
- Internationalisierte Version erstellen
- Neue Elemente und Attribute hinzufügen
- Element- und Attributnamen ändern (nicht empfohlen)
- Schema (RelaxNG, W3C Schema usw.) generieren
- Dokumentation generieren (HTML, PDF)
- **ODD Customization zur Weiterbearbeitung speichern!**

Erste ODD 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" schematype
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
```

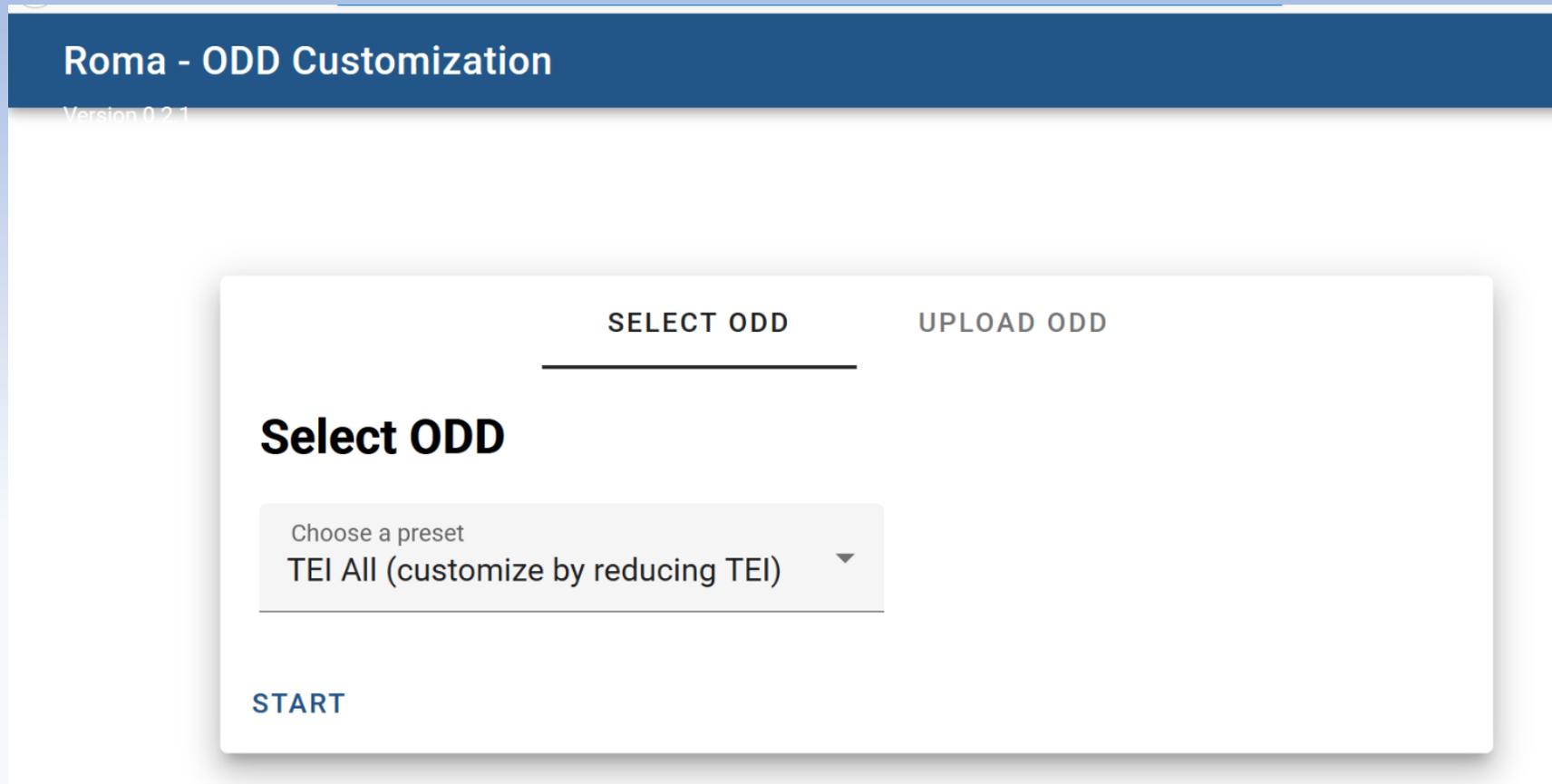
```
<schemaSpec ident="myCustomization">
  <moduleRef key="tei"/>
  <moduleRef key="header"/>
  <moduleRef key="core"/>
  <moduleRef key="textstructure"/>
</schemaSpec>
```

Wie wird ein projektspezifisches Schema erstellt?

- Code schreiben
 - ODD Customization im Oxygen Framework
- Roma als Web-Frontend
 - <https://romabeta.tei-c.org/> (neue Version)
 - <https://roma.tei-c.org/> (alte Version)

Roma interface

<https://romabeta.tei-c.org/>



The screenshot displays the 'Roma - ODD Customization' web interface. At the top, a dark blue header contains the title 'Roma - ODD Customization' and the version 'Version 0.2.1'. Below the header, there are two tabs: 'SELECT ODD' (which is active and underlined) and 'UPLOAD ODD'. The 'SELECT ODD' tab contains a section titled 'Select ODD' with a dropdown menu labeled 'Choose a preset' and the selected option 'TEI All (customize by reducing TEI)'. A 'START' button is located at the bottom left of the main content area.

Roma - ODD Customization
Version 0.2.1

SELECT ODD UPLOAD ODD

Select ODD

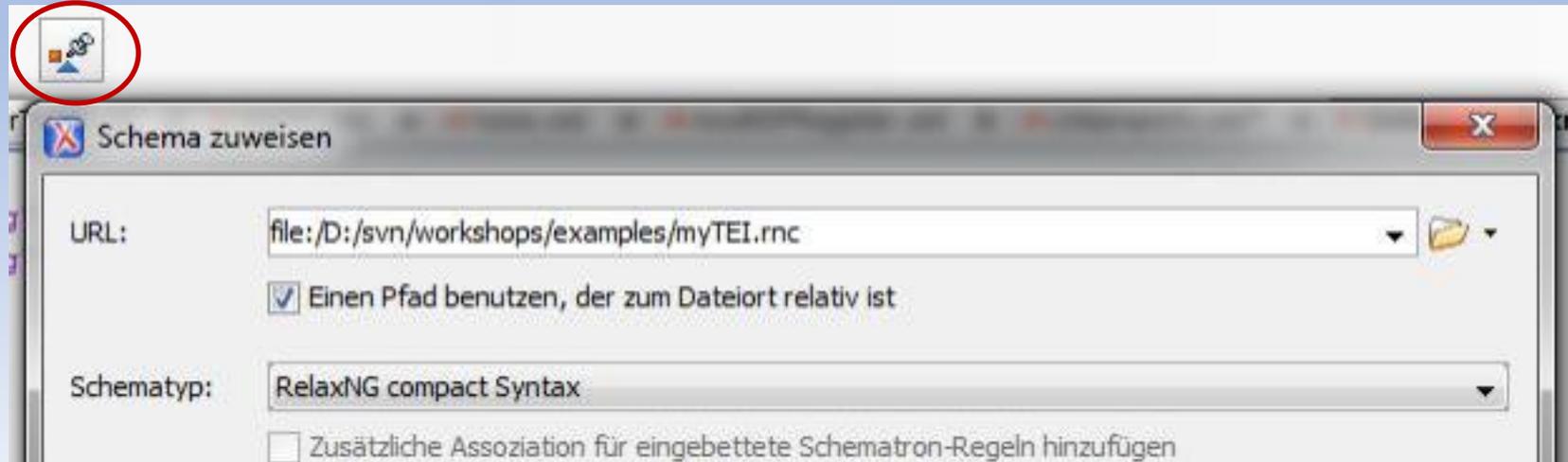
Choose a preset
TEI All (customize by reducing TEI)

START

Erste Übung mit Roma

- Roma öffnen: <https://romabeta.tei-c.org/>
- Vordefiniertes Set wählen (z.B. TEI Minimal)
- „Title“ und „Author“ ändern
- Das Modul “namesdates” hinzufügen
- Nicht benötigte Elemente, z.B. <nym>, <object> und <objectName> löschen
- ODD und Schema herunterladen
 - ODD Customization für die Weiterbearbeitung speichern (.odd)
 - Schema (z.B. RelaxNG .rng) speichern
- Schema mit 1669_Rezess_Hanse.xml assoziieren
- Schema durch fehlende Module/Elemente ergänzen (z.B. das Modul “linking”)
- Werteliste von @type im <place>-Element auf ‘town’ und ‘building’ beschränken. Es soll ‘required’ sein und die Werte sollen eine zusätzliche Beschreibung enthalten.
- Den Datentyp des @n-Attributs im <pb>-Element auf teidata.count ändern

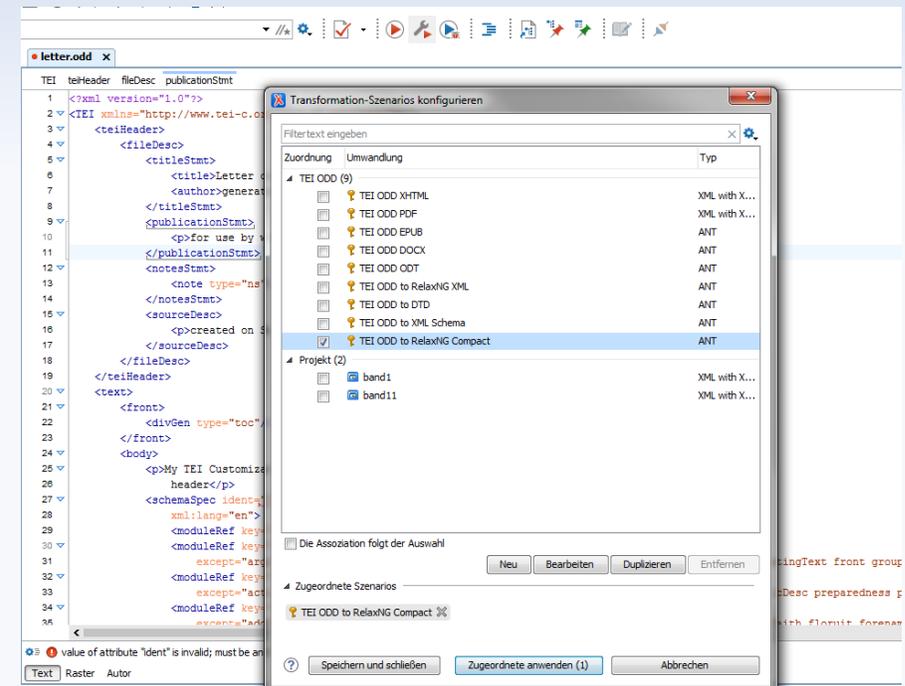
Schema mit XML assoziieren (Oxygen)



- Icon “Schema zuweisen...” in der Werkzeugleiste oder Dokument > Schema > Schema zuweisen

ODD in oXygen bearbeiten

- In oXygen stehen vordefinierte Transformationsszenarien zur Verfügung
 - Transformationsszenario konfigurieren
 - Gewünschtes Szenario (z.B. TEI ODD to RelaxNG XML Schema) wählen
 - „Zugeordnete anwenden“ klicken
- Das Resultat wird im Ordner „out“ gespeichert
- Zuordnung zum Schema erneut erstellen



ODD Spezifikation: Module inkludieren

Ein Metaformat zur Generierung multipler Outputs

```
<body>
  <schemaSpec ident="hanse">

    <!-- erforderliche Module -->
    <moduleRef key="tei"/>      <!-- enthält das TEI Klassensystem -->
    <moduleRef key="core"/>    <!-- enthält <title>, <p>, <emph> etc. -->
    <moduleRef key="header"/>  <!-- enthält <teiHeader>, <fileDesc> etc. -->
    <moduleRef key="textstructure"/> <!-- enthält <TEI>, <text>, <div> etc. -->

    <!-- optionale Module -->
    <moduleRef key="namesdates"/>
    <moduleRef key="transcr"/>
    <moduleRef key="analysis"/>

  </schemaSpec>
</body>
```

ODD Spezifikation

```
<schemaSpec ident="hanse" docLang="de">  
  <moduleRef key="tei"/>  
  <moduleRef key="core"/>  
  <moduleRef key="header"/>  
  <moduleRef key="textstructure"/>  
</schemaSpec>
```

<code><schemaSpec></code>	Elternelement für die Schemaspezifikation
<code>@docLang</code>	Dokumentationssprache
<code>@ident</code>	Dateiname
<code><moduleRef></code>	Einbindung von Modulen
<code>@key</code>	Name des TEI Moduls
<code>@except/@include</code>	Ausschluss/Einbindung von Elementen

ODD Spezifikation: Elemente exkludieren/inkludieren

- Elemente aus einem selektierten Modul entfernen:

```
<moduleRef key="namesdates" except="addName affiliation bloc  
climate"/>
```

- Elemente in ein Modul inkludieren:

```
<moduleRef key="textstructure" include="TEI text div back body"/>
```

ODD Spezifikation: Elemente einzeln inkludieren

```
<elementRef key="persName"/>
```

```
<elementRef key="zone"/>
```

Achtung: wenn das Modul transcr nicht referenziert ist, dann steht für <zone> die Attributklasse att.coordinated (@start, @ulx, @uly, @lrx, @lry, @points) nicht zur Verfügung.

ODD: Element spezifizieren / Attribut löschen

```
<elementSpec ident="div" mode="change" module="textstructure">  
  <attList>  
    <attDef ident="decls" mode="delete"/>  
    <attDef ident="met" mode="delete"/>  
    <attDef ident="synch" mode="delete"/>  
  </attList>  
</elementSpec>
```

<code><elementSpec></code>	Element näher spezifizieren
<code>@ident</code>	Name des Elements, das modifiziert wird
<code>@mode</code>	Aktion: <code>delete</code> <code>change</code> <code>add</code> <code>replace</code>
<code>@module</code>	Modulname, aus dem das Element stammt

<code><attList></code>	Liste von Attributen
<code><attDef></code>	Attributdefinition
<code>@ident</code>	Name des Attributs, das modifiziert wird
<code>@mode</code>	Aktion: <code>delete</code> <code>change</code> <code>add</code> <code>replace</code>

ODD Spezifikation – Attributklassen entfernen

- Attributklasse aus Schema entfernen:

```
<classSpec type="atts" ident="att.personal" mode="delete"/>
```

- Attributklasse aus einem spezifischen Element entfernen:

```
<elementSpec module="core" ident="pb" mode="change">  
  <classes mode="change">  
    <memberOf key="att.typed" mode="delete"/>  
  </classes>  
</elementSpec>
```

ODD Spezifikation – Attribute entfernen

- Einzelne Attribute aus einer Attributklasse entfernen:

```
<classRef key="att.global" except="xml:base xml:space"/>
```

- Spezifisches Attribut aus einem Element entfernen:

```
<elementSpec module="core" ident="note" mode="change">  
  <attList>  
    <attDef ident="anchored" mode="delete"/>  
  </attList>  
</elementSpec>
```

ODD: Liste von Attributwerten

```
<elementSpec ident="div" mode="change" module="textstructure">
  <attList>
    <attDef ident="type" mode="change" usage="req">
      <valList type="closed" mode="replace">
        <valltem ident="Abstract" />
        <valltem ident="Inhaltsverzeichnis" />
      </valList>
    </attDef>
  </attList>
</elementSpec>
```

@usage Die Verwendung des Attributs ist obligatorisch bzw. optional: **req** | **opt**

<valList> Elternelement der Werteliste

@type Offene oder geschlossene Listen: **open** | **closed**

<valltem> Liste der zulässigen Attributwerte

ODD: eigenes Beispiel hinzufügen

```
<elementSpec ident="place" mode="change">
  <attList>
    <attDef ident="type" mode="change" usage="req">
      <desc versionDate="2020-08-26" xml:lang="de"/>
      <valList type="closed" mode="change">
        <valItem mode="add" ident="town">
          <desc versionDate="2020-08-26" xml:lang="de">Der Ort ist eine Stadt</desc>
        </valItem>
        <valItem mode="add" ident="building">
          <desc versionDate="2020-08-26" xml:lang="de">Der Ort ist ein Gebäude</desc>
        </valItem>
      </valList>
    </attDef>
  </attList>
  <exemplum>
    <egXML xmlns="http://www.tei-c.org/ns/Examples">
      <place xml:id="antwerpen_stadt" type="town">
        <placeName>Antwerpen</placeName>
        <region ref="#flandern">Flandern</region>
        <country>Spanische Niederlande</country>
        <idno>https://www.geonames.org/2803138/antwerpen.html</idno>
      </place>
      <place xml:id="antwerpen_stadt_hansekantor" type="building">
        <placeName>Hansisches Kontor, Antwerpen</placeName>
        <placeName>Antwerpischen haüses</placeName>
        <settlement ref="#antwerpen_stadt">Antwerpen</settlement>
        <region ref="#flander">Flandern</region>
        <country>Spanische Niederlande</country>
      </place>
    </egXML>
  </exemplum>
</elementSpec>
```

ODD: Dokumentieren

<p>Die Einbindung des Moduls <tag>moduleRef key="namesdates"</tag> ist für die erweiterte Kodierung von Personeninformationen zwingend notwendig.</p>

<p>Das Element <gi>idno</gi> muss das Attribut <att>type</att> enthalten. Dieses kann die Werte <val>URLWeb</val> oder <val>URLXML</val> aus der vordefinierten Liste annehmen.</p>

<gi>	Element
<att>	Attribut
<val>	Attributwert
<tag>	Tag

Modul 22 „Documentation Elements“ für weitere Dokumentationselemente

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

Beispiel ODDs

- Demo

https://wvp.northeastern.edu/outreach/seminars/_current/demos/ODDs/imls_tight.odd

- JTEI Journal

https://tei-c.org/release/xml/tei/custom/odd/tei_jtei.odd

Welche Elemente verwende ich bereits in meinem TEI Dokument?

- XPath:

```
distinct-values(//*/name())
```

Nachmittagsübung

- Erstellen Sie ein projektspezifisches Schema für das Dokument 1969_Rezess_Hanse_full.xml.
- Ergänzen Sie weitere notwendige Module.
- Entfernen Sie alle Elemente, die nicht benötigt werden.
- Erstellen Sie eine geschlossene Werteliste für das @type-Attribut des <roleName>-Elements. Fügen Sie eine entsprechende Dokumentation für die einzelnen Werte hinzu.
- Entfernen Sie Attributklassen aus dem Schema, die Sie nicht benötigen.
- Entfernen Sie aus den Elementen jene Attribute, die Sie nicht benötigen.
- Fügen Sie eine projektspezifische Dokumentation unter Verwendung der Dokumentationselemente (<gi>, <att>, <val>) hinzu.
- Fügen Sie zu einem beliebigen Element ein eigenes Beispiel hinzu.
- Generieren Sie eine HTML-Dokumentation.

Weiterführende Links

- Raffaele Viglianti, « One Document Does-it-all (ODD): a language for documentation, schema generation, and customization from the Text Encoding Initiative »
<https://www.balisage.net/Proceedings/vol24/html/Viglianti01/BalisageVol24-Viglianti01.html>
- Raffaele Viglianti, « Creating A TEI Customization with the new Roma »
<http://www.itisnotsound.com/2019/11/18/video-creating-a-tei-customization-with-the-new-roma/>
- Syd Bauman, « A TEI Customization for Writing TEI Customizations », *Journal of the Text Encoding Initiative* [Online], Issue 12 | July 2019 - May 2020, Online since 15 November 2019, connection on 21 August 2020. URL: <http://journals.openedition.org/jtei/2573>; DOI: <https://doi.org/10.4000/jtei.2573>
- Syd Bauman: Schematron Introduction
https://www.wwp.neu.edu/outreach/seminars/xpath_2019-09_tei/presentations/xpath_intro/taste_of_Schematron_00.xhtml

Danke für Ihre Aufmerksamkeit!

Fragen?

martina.scholger@uni-graz.at

Danke für das Teilen von Workshopmaterial:
Woman Writers Project (Syd Bauman, Julia Flanders)
TEI Technical Council
Institut für Dokumentologie und Editorik



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



 ZENTRUM FÜR
INFORMATIONSMODELLIERUNG
AUSTRIAN CENTRE FOR
DIGITAL HUMANITIES

