

Summer School »Digitale Edition«

2. – 5. Oktober 2017



TEI-Customization

Stefan Dumont

*Unter Verwendung von Folien
von Susanne Haaf, IDE-School 2012, Chemnitz*

Was ist ein Schema?

- Inventar von Elementen
- Verhältnis der Elemente zueinander
- Regeln, in welchen Kontexten welche Auszeichnungen (d.h. Elemente und Attribute) möglich sind

- Umsetzung: DTD, XSD, RNG

Beispiele:

- TEI-Schema: `tei_all`
- http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng
- TEI-Element: `<correspDesc>`
- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-correspDesc.html>

Schema-Spezifikation für TEI/P5

- Problem: Das `tei_all`-Schema ist (intentional) sehr flexibel. Dadurch sind oft für ein Textphänomen mehrere Auszeichnungen möglich.
- Beispiel: Eigennametagging
- → verschiedene Möglichkeiten, in TEI Personennamen auszuzeichnen
 - Element: `<rs>` (referencing string)
→ `<rs type="propNounPersName">`
 - Element: `<name>` (name, proper noun)
→ `<name type="person">`
 - Element: `<persName>` (personal name)

Schema-Spezifikation für TEI/P5

- “[...] it is almost impossible to use the TEI scheme without customizing or personalizing it in some way.” (cf. P5 Guidelines of the TEI, ch. 23.2)
 - “Customization is a central aspect of TEI usage and the Guidelines are designed with customization in mind.”
 - (<http://www.tei-c.org/Guidelines/Customization/>)
 - “From the start, the TEI was intended to be used as a set of building blocks for creating a schema suitable for a particular project. This is in keeping with the TEI philosophy of providing a vocabulary for describing texts, not dictating precisely what those texts must contain or might have contained.”
 - (<http://www.tei-c.org/Guidelines/Customization/odds.xml>)
- Die TEI Guidelines unterstützen die Anpassung des Schemas.

TEI-Empfehlungen für Spezifikationen

- <http://www.tei-c.org/Guidelines/Customization/>

Customizations provided by the TEI Consortium

Lite	TEI Lite, the most widely used TEI customization; includes basic elements for simple documents	ODD DTD RNG XSD HTML PDF
TEI Tite	A constrained customization designed for use by keyboarding vendors.	ODD DTD RNG XSD HTML PDF
Bare	TEI Absolutely Bare, a very barebones schema with the absolute minimum of elements	ODD DTD RNG XSD
All	TEI with all modules included	ODD DTD RNG XSD
Corpus	TEI for Linguistic Corpora, includes the modules for encoding linguistic corpora	ODD DTD RNG XSD
MS	TEI for Manuscript Description, includes the elements for describing manuscripts and complex physical aspects of documents	ODD DTD RNG XSD
Drama	TEI with Drama, includes the TEI drama module	ODD DTD RNG XSD
Speech	TEI for Speech Representation, includes the TEI module for spoken language	ODD DTD RNG XSD
Dictionaries	TEI for Dictionaries	ODD DTD RNG XSD

- Beispiel einer externen Customization:
Basisformat des Deutschen Textarchivs
<http://deustextarchiv.de/doku/basisformat>

TEI: Customization

- ODD: „One document does it all“
- Grundlage für die TEI Guidelines, bestehend aus Dokumentation, Beispielen und den formalen Deklarationen (für Elemente, Attribute, Module, ...)
- ODD nutzt das **tagdocs**-Modul → Inventar zur Anpassung der TEI Guidelines an die eigenen Bedürfnisse
- kein direkter Eingriff in das Schema der TEI, sondern übergeordnete Anpassungsregeln

TEI: Customization

Einblick TEI Infrastruktur:

- **Modules:** Elemente und Attribute werden in der TEI nach ihrem Zweck gruppiert → Beispiel [persName](#): Modul "namesdates"
- **Model Classes:** Weitere Gruppierung von Elementen, um sie leichter als Kindelemente einbinden zu können
→ Beispiel [correspDesc](#): Mitglied von [model.profileDescPart](#)
- **Attribute Classes:** Attribute sind häufig ebenfalls in Klassen definiert, um sie leichter in vielen verschiedenen Elementen einbinden zu können
→ Beispiel: [att.canonical](#)
- **Macros:** Gruppiert Model Classes und ggf. Elemente, um sie leicht als bestimmte Inhaltsgruppen einbinden zu können
→ Beispiel: [macro.paraContent](#)

TEI: Customization

1. Elemente entfernen
2. Elemente umbenennen
3. "Content models" von Elementen oder Klassen modifizieren
4. Attribut- und Wertauswahl für Elemente und Klassen modifizieren
5. Klassenzugehörigkeiten modifizieren
6. neue Elemente (zu bestehenden Klassen) hinzufügen

(Vgl. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html#MDMD>)

Mögliche Grundaktionen: add, delete, change, replace

Achtung! Spezifikation vs. Veränderung des TEI-Schemas: Für die Austauschbarkeit von TEI-Dokumenten ist es wünschenswert, dass TEI-Dokumente in Bezug auf tei_all gültig sind.

Roma

**Webtool beim Erstellen eines ODD-Dokuments:
Roma <http://www.tei-c.org/Roma/>**

Roma: Startpunkt auswählen

– Roma: generating customizations for the TEI

These pages will help you design your own TEI customization, as a DTD, RELAX NG or W3C Schema.

Create a new or upload existing customization

- Build up: create a new customisation by adding elements and modules to the smallest recommended schema
- Reduce: create a new customization by removing elements and modules from the largest possible schema
- Create customization from template
- Open existing customization

TEI Absolutely Bare	▼
TEI Absolutely Bare	
TEI Lite	
TEI Tite	
TEI for Linguistic Corpora	
TEI for Manuscript Description	
TEI with Drama	
TEI for Speech Representation	
TEI for authoring ODD	
TEI with SVG	
TEI with MathML	
TEI with XInclude (experimental)	
TEI with W3C ITS	
TEI for Dictionaries (experimental)	

Start

Roma was written by Arno Mittelbach and is maintained by

Ioan Bernevig. Please direct queries to the [TEI@Oxford](#) project.

Roma: Metadaten für das ODD spezifizieren



Set your parameters

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#) [Sanity Checker](#)

Set your parameters

Title	<input type="text" value="IDE autumn school TEI customization"/>
Filename	<input type="text" value="tei_all_ideAutSchool"/>
Namespace for new elements	<input type="text" value="http://www.example.org/ns/nonTEI"/>
Prefix for TEI pattern names in schema	<input type="text"/>
Language	<input checked="" type="radio"/> English, <input type="radio"/> Deutsch, <input type="radio"/> Italiano, <input type="radio"/> Español, <input type="radio"/> Français, <input type="radio"/> Portugues, <input type="radio"/> Russian, <input type="radio"/> Svenska, <input type="radio"/> 日本語, <input type="radio"/> 中文
Author name	<input type="text" value="generated by Roma 3.6"/>
Description	<input type="text" value="This TEI customization contains some adjustments to tei_all."/>

[Save](#)

Roma was written by Arno Mittelbach and is maintained by Sebastian Rahtz. Sanity check written by Ioan Bernevig. Documentation language en. Please direct queries to the [TEI@Oxford](#) project. This is Roma version 4.9, last updated 2012-06-16. Using TEI P5 version 2.1.0

Module und Elemente auswählen

- Roma:
 - Module: „Modules“ > „List of selected Modules“
 - Elemente: „List of elements in module:[Modulname]“
- ODD:
 - Module: <moduleRef> (module reference) mit @key
 - Elemente: @except oder @include in <moduleRef>

- Beispiel:

```
<moduleRef key="textstructure" except="div1 div2 div3 div4 div5 div6  
div7 group"/>
```

```
<moduleRef key="drama" except="camera caption epilogue move  
performance prologue set sound spGrp tech view"/>
```

- *Achtung!* Module ohne <moduleRef> im ODD werden nicht eingebunden in das Schema.

Roma: Module auswählen

Modules

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#) [Sanity Checker](#)

List of TEI Modules			
	Module name	A short description	Changes
add	analysis	? Simple analytic mechanisms	
add	certainty	? Certainty and uncertainty	
add	core	? Elements common to all TEI documents	
add	corpus	? Corpus texts	
add	dictionaries	? Dictionaries	
add	drama	? Performance texts	
add	figures	? Tables, formulæ, notated music, and figures	
add	gaiji	? Character and glyph documentation	
add	header	? The TEI Header	
add	iso-fs	? Feature structures	
add	linking	? Linking, segmentation and alignment	
add	msdescription	? Manuscript Description	
add	namesdates	? Names and dates	
add	nets	? Graphs, networks, and trees	
add	spoken	? Transcribed Speech	
add	tagdocs	? Documentation of TEI modules	
add	textcrit	? Critical Apparatus	
add	textstructure	? Default text structure	
add	transcr	? Transcription of primary sources	
add	verse	? Verse structures	

Liste der verfügbaren TEI-Module

List of selected Modules

[remove](#) [tei](#)

[remove](#) [core](#)

[remove](#) [analysis](#)

[remove](#) [certainty](#)

[remove](#) [corpus](#)

[remove](#) [dictionaries](#)

[remove](#) [drama](#)

[remove](#) [figures](#)

[remove](#) [gaiji](#)

[remove](#) [header](#)

[remove](#) [iso-fs](#)

[remove](#) [linking](#)

[remove](#) [msdescription](#)

[remove](#) [namesdates](#)

[remove](#) [nets](#)

[remove](#) [spoken](#)

[remove](#) [textcrit](#)

[remove](#) [textstructure](#)

[remove](#) [transcr](#)

[remove](#) [verse](#)

[remove](#) [tagdocs](#)

Liste der für die eigene Spezifikation ausgewählten TEI-Module

Roma: Elemente auswählen

Change module

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#) [Sanity Checker](#)

[back](#)

List of elements in module:namesdates

	Include	Exclude	Name	Description	Attributes
addName	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="addName"/>	? (additional name) contains an additional name component, such as a nickname, epithet, or alias, or any other descriptive phrase used within a personal name.	Change attributes
affiliation	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="affiliation"/>	? (affiliation) contains an informal description of a person's present or past affiliation with some organization, for example an employer or sponsor.	Change attributes
age	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="age"/>	? (age) specifies the age of a person.	Change attributes
birth	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="birth"/>	? (birth) contains information about a person's birth, such as its date and place.	Change attributes
bloc	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="bloc"/>	? (bloc) contains the name of a geo-political unit consisting of two or more nation states or countries.	Change attributes
climate	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="climate"/>	? (climate) contains information about the physical climate of a place.	Change attributes
country	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="country"/>	? (country) contains the name of a geo-political unit, such as a nation, country, colony, or commonwealth, larger than or administratively superior to a region and smaller than a bloc.	Change attributes
death	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="death"/>	? (death) contains information about a person's death, such as its date and place.	Change attributes
district	<input type="radio"/>	<input checked="" type="radio"/>	<input type="text" value="district"/>	? contains the name of any kind of subdivision of a settlement, such as a parish, ward, or other administrative or geographic unit.	Change attributes
education	<input checked="" type="radio"/>	<input type="radio"/>	<input type="text" value="education"/>	? contains a description of the educational experience of a	Change attributes

Liste der Elemente eines gewählten Moduls

Module und Elemente auswählen

- Nötiges Vorwissen:
 - In welchen Modulen sind Elemente enthalten, die ich benötige?
 - Welche Module kann ich hingegen ganz eliminieren?
 - In welchen der von mir ausgewählten Module sind die Elemente enthalten, die ich verändern möchte?
- Hilfreich dabei: TEI-Elemente gruppiert nach Modulen
- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html>
- → „Show by Module“

Module und Elemente auswählen – Beispiel

Beispiel: Eigennamen (vgl. Folie 3)

- Modul „namesdates“ (und somit auch das in diesem Modul enthaltene Element <persName>) entfernen
 - **Roma:** „Modules“ → in der „List of selected modules“: „remove namesdates“
 - ODD: auf eine <moduleRef key="namesdates"/> verzichten

- Element <rs> (aus dem Modul „core“) entfernen
 - **Roma:** „Modules“ → in der „List of selected modules“ das Modul „core“ auswählen; die „List of elements in module:core“ erscheint; dort mittels „Exclude“ das Element <rs> aus dem Modul entfernen
 - ODD: <moduleRef key="core" except="rs"/>

Module und Elemente auswählen – Beispiel

Beispiel: Eigennamen (vgl. Folie 3)

- Element `<name>` (aus dem Modul „core“) für die Auszeichnung von Eigennamen erhalten
 - **Roma:** „Modules“ → in der „List of selected modules“ das Modul „core“ auswählen; die „List of elements in module:core“ erscheint; dort mittels „Include“ das Element `<name>` in dem Modul erhalten
 - ODD:
 - `<moduleRef key="core"/>`: erlaubt sämtliche Elemente des Moduls „core“
 - `<moduleRef key="core" insert="name p quote"/>`: erlaubt nur die genannten Elemente des Moduls „core“ (also auch `<name>`)
 - `<moduleRef key="core" except="rs q said"/>`: erlaubt alle Elemente des Moduls „core“, außer den genannten (also auch `<name>`)

Attribute & Werte auswählen

- Die laut TEI erlaubten Attribute können eliminiert oder hinsichtlich ihrer Werte konkretisiert werden
- → Klassenweise (und somit Element-übergreifend)
 - **Roma:** „Change Classes“ → für jede Klasse unter „Change attributes“ mittels „Include“ Attribute einschließen oder sie mittels „Exclude“ entfernen; nähere Festlegungen zur Attributauswahl können von dort aus für jedes Attribut unter „Change attribute“ vorgenommen werden
 - ODD: <classSpec> (class specification)
 - Beispiel:
 - <classSpec ident="att.datcat" module="tei" type="atts" mode="delete"/>
 - <classSpec ident="att.typed" module="tei" type="atts" mode="change">
 <attList><attDef ident="subtype" mode="delete"/></attList>
</classSpec>

Attribute & Werte auswählen

- Die laut TEI erlaubten Attribute können eliminiert oder hinsichtlich ihrer Werte konkretisiert werden
- → Elementweise
 - **Roma:** „Modules“ → ... „List of elements in module:[X]“ → „Change Attributes“
 - ODD: <elementSpec> (element specification) → <attDef> (attribute definition)
 - Beispiel:

```
<elementSpec id="figure" module="figures" mode="change">
  <attList>
    <attDef id="place" mode="delete"/>
    <attDef id="type" mode="change" usage="opt">
      <valList type="closed" mode="replace">
        <valItem id="notatedMusic"/>
      </valList>
    </attDef>
    <attDef id="n" mode="delete"/>
  </attList>
</elementSpec>
```

Attribute & Werte auswählen

- Hilfreich dabei: die Auflistung sämtlicher Elemente, die von einer bestimmten Attributklasse Gebrauch machen
→ z.B. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-att.global.html>
- *Übrigens*: Übersicht über die Attributklassen in den TEI Guidelines: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-CLASSES-ATTS.html>

Schema erstellen mit Roma

- ~~ODD auf Fehler überprüfen mit dem „Sanity Checker“~~
→ derzeit nicht funktionsfähig
- ODD speichern unter „Save customization“
- Achtung!* Das ODD sollte immer abgespeichert werden, damit die bestehende Spezifikation späteren möglichen Anpassungen zugrunde gelegt werden kann.
- Schema erstellen unter „Schema“
→ Auswahl des bevorzugten Schema-Formats (RNG empfohlen)



Time to give you a schema

[New](#) [Customize](#) [Language](#) [Modules](#) [Add Elements](#) [Change Classes](#) [Schema](#) [Documentation](#) [Save Customization](#) [Sanity Checker](#)

Creating a schema

Which format do you prefer?

- RELAX NG schema (XML syntax)
- RELAX NG schema (compact syntax)
- RELAX NG schema (XML syntax)
- ISO Schematron
- Schematron
- W3C schema (in ZIP archive)
- DTD

[Generate](#)

We use RELAX NG

http://www.w3.org/wiki/XML_Schema_Language_comparison

Arbeiten mit dem erstellten Schema

- Das Schema wird an zweiter Stelle (nach der XML-Deklaration) in die TEI-Datei eingebunden mit dem Ausdruck:
- `<?xml-model href="[URL zum Schema]"
schematypens="[Namensraum]"?>`
- z.B. das RNG von tei.all:
- `<?xml-model href="http://www.tei-
c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng"
schematypens="http://relaxng.org/ns/structure/1.0"?>`
- Im **oXygen** kann das Schema auch durch Verwendung des Menüs „Dokument“ > „Schema“ > „Schema zuweisen“ eingebunden werden:
 - Pfad zum Schema angeben, Schematyp festlegen, OK
 - Resultat: Schema wird an der richtigen Stelle in der Datei eingebunden

Literatur und Links

- TEI Guidelines zu ODD und Customization: Kapitel 22 und 23
- Einführung zu ODD: <http://www.tei-c.org/Guidelines/Customization/odds.xml>
- ODD-Wiki: <http://wiki.tei-c.org/index.php/ODD>
- Roma: <http://www.tei-c.org/Roma/>

- Zum DTA-Basisformat im Vergleich mit anderen TEI-Formaten: *Geyken/Haaf/Wiegand*, The DTA 'base format': A TEI-Subset for the Compilation of Interoperable Corpora (Proceedings der Konvens 2012, <http://www.oegai.at/konvens2012/proceedings.pdf#page=383>)
- Dokumentation des DTA-Basisformats inkl. ODD und RNG:
- www.deutschestextarchiv.de/doku/basisformat