



Einführung in die TEI

Summer School "Digitale Edition" 2016
Erschließung geisteswissenschaftlicher Quellen mit digitalen
Methoden

5. September 2016, Christiane Fritze

Zentrum für Informationsmodellierung
Austrian Centre for Digital Humanities
Elisabethstraße 59/III, SR 81.31



Überblick

- ▶ Was ist TEI?
 - ▶ Ein Konsortium.
 - ▶ Ein Kodierungsstandard.
- ▶ Was bietet TEI?
 - ▶ Guidelines.
 - ▶ Werkzeuge.
 - ▶ Und Mehr.
- ▶ Wie gehe ich mit TEI um?
 - ▶ Grundaufbau des TEI Datenmodells.
 - ▶ Guidelines benutzen.



Text Encoding Initiative - TEI

▶ Gründung

- ▶ Frühjahr 1987: European workshops on standardisation of historical data (J.P. Genet, M Thaller)
- ▶ Herbst 1987: NEH funds an exploratory international workshop on the feasibility of defining "text encoding guidelines"

Vassar College, Poughkeepsie

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





< Text Encoding Initiative >

Home Guidelines Activities Tools Membership Support About News Online Store

- “The TEI was established in 1987 to develop, maintain, and promulgate hardware- and software-independent methods for encoding humanities data in electronic form.”
- “Over nearly three decades the TEI has been extraordinarily successful at achieving its objective and it is now widely used by scholarly projects and libraries around the world.”
- “TEI is the *de facto* standard for text encoding.”

Angebote



- TEI P5 Guidelines (derzeit 1.646 Druckseiten)
 - <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
 - Release 3.0.0. (letztes Update am 29. März 2016)
 - Lehrmaterial
- Tools:
 - OxGarage: Konversion von und nach TEI
 - ROMA: Erstellen von Schemata und Dokumentationen
 - Stylesheets: Konversion von TEI-Dokumenten in diverse Formatvorlagen
- SIGs, Annual members' meetings
- Journal: [jTEI](http://journal.tei-c.org/journal/index)
 - <http://journal.tei-c.org/journal/index>
- TEI-L Mailingliste: tei-l@listserv.brown.edu
 - Fragen von allen für alle auf allen Niveaus
 - Mail-Archiv:
 - <http://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L>
 - <http://markmail.org/list/edu.brown.listserv.tei-l>





TEI Conference 2016

26th to 30th September, Vienna, Austria



Welcome to TEI conference and members' meeting 2016

The TEI conference 2016 will be hosted by the Austrian Centre for Digital Humanities ACDH from 26th to 30th September 2016 at the Austrian Academy of Sciences in Vienna.

Program

Tweet about the conference: [#TEIconf2016](#)

#TEIconf2016



TEI2German
Translatathon at
[#TEIconf2016](#) in Vienna
(27 Sep 2016):
tei2016.acdh.oew.ac.at/tranlatathon

02 Sep



- ▶ Konvertiere mittels **OxGarage** das Dokument Postkarte.txt nach TEI P5.

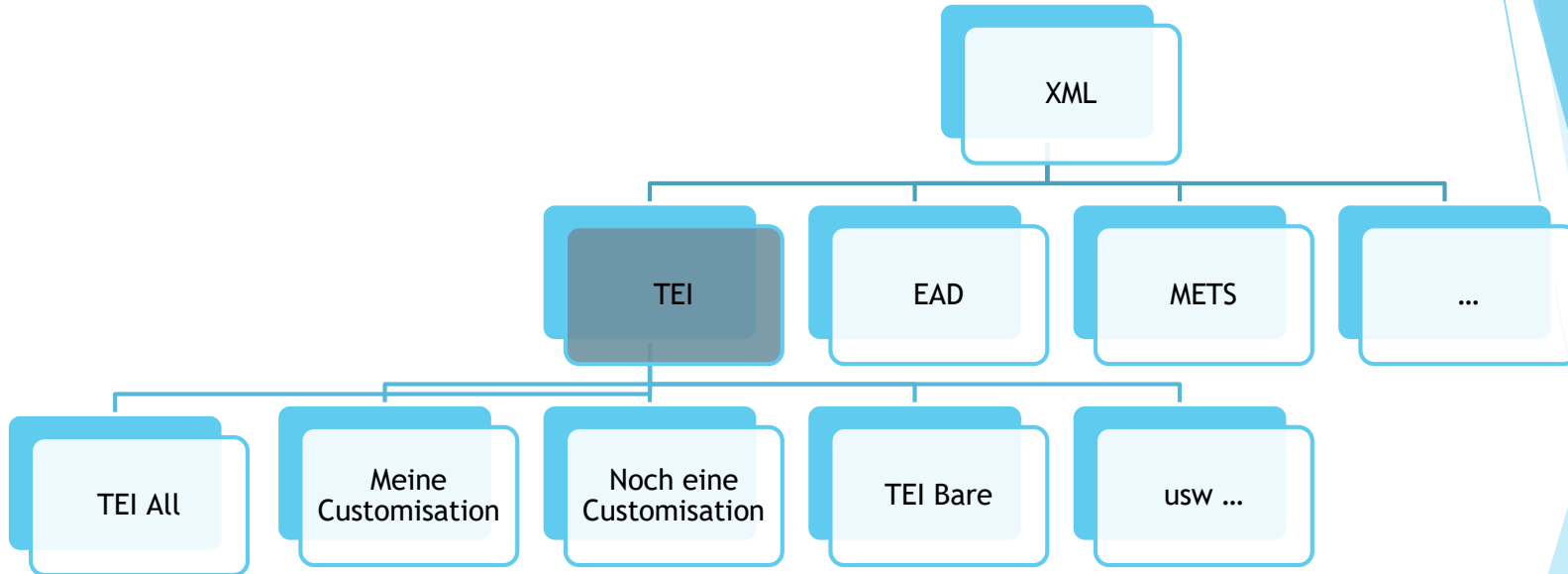
<http://www.tei-c.org/oxgarage/>

- ▶ Speichere das entstandene TEI-Dokument und öffne es im Oxygen-XML-Editor.



- ▶ XML: stellt nur die grundlegenden Regeln bereit:
 - ▶ Baumstruktur (ein Wurzelelement, korrekte Schachtelung)
 - ▶ Konventionen für die Kodierung von
 - ▶ Elementen (`<element/>`)
 - ▶ Attributen (`<element attribut= " wert" />`)
 - ▶ Entity-Referenzen (`&entity;` z.B. `<`)
 - ▶ Kommentaren (`<!-- ... -->`) usw.
 - ▶ Benennungsregeln für Elemente und Attribute

TEI ist XML



- ▶ sind eine Einschränkung der im Prinzip unendlichen Möglichkeiten von XML
- ▶ Bieten formalisierte Schemata zur Validierung von XML-Dateien

Fragen, die geklärt werden müssen:

- ▶ Welche Tags für Elemente und Attribute werden bereitgestellt?
- ▶ Wie dürfen die Tags verschachtelt werden?
- ▶ Wie kann ich die gegebenen Möglichkeiten um eigene Regeln erweitern?



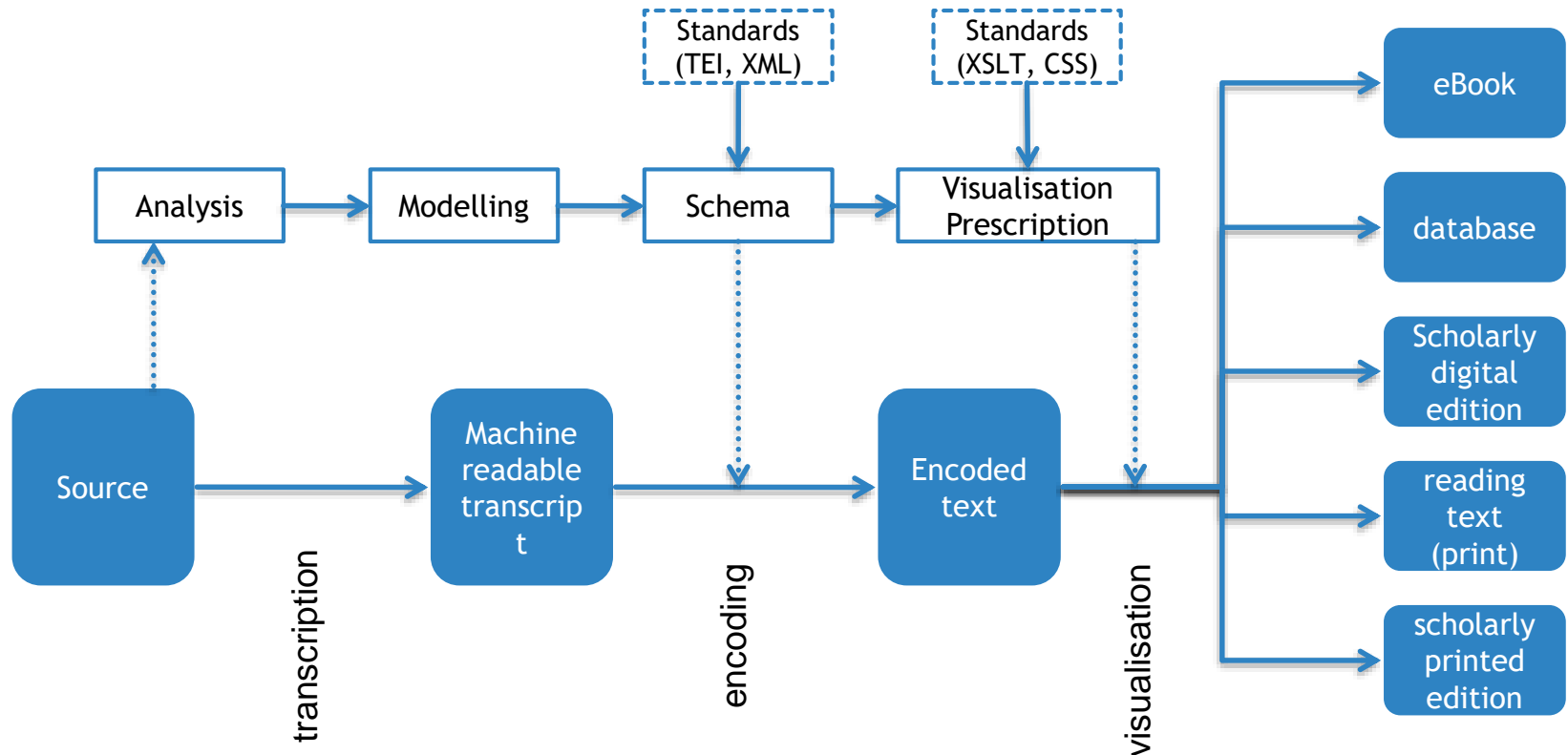
- ▶ Öffnen Sie im Oxygen-XML-Editor ein neues XML-TEI P5-Dokument (TEI Lite):
 - ▶ [STRG + N] → Framework-Vorlagen → TEI P5 → TEI Lite

TEI ist XML



```
<?xml version="1.0" encoding="UTF-8"?>
<TEI>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Title</title>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>Some text here.</p>
    </body>
  </text>
</TEI>
```

Alles wird gut



TEI-Grundstruktur



- Wurzelelement **<TEI>**
- enthält mindestens **zwei** Unterelemente, nämlich
 - **<teiHeader>**
für Metadaten und **muss** immer vorhanden sein **und**
 - **<text>**
für Texte aller Art **und/oder**
 - **<sourceDoc>**
für genetische Edition / topographische Transkriptionen **und/oder**
 - **<facsimile>**
für Verknüpfung mit Bildern **und/oder**
 - **<fsdDecl>** ‚Feature Structure Declaration‘
v.a. für Textanalysen, linguistische Merkmale u.ä.

TEI-Header



<teiHeader> kann aus vier Hauptteilen bestehen:

- **<fileDesc>** (file description)
Bibliographische Beschreibung des TEI-Dokuments
- **<encodingDesc>** (encoding description)
Beschreibung der editorischen Praxis
- **<profileDesc>** (profile description)
Kodierung inhaltlicher Informationen über den Text, v.a. bei Sprachcorpora
- **<revisionDesc>** (revision description)
Informationen über die Änderungsgeschichte' des TEI-Dokuments

N.B.: Only **<fileDesc> is mandatory.**

TEI-Guidelines: 3 Elements Available in All TEI Documents

- **<p>** (paragraph): Absatz
- **<ab>** (anonymous block): Irgendein Textblock
- **<head>** (head line): Überschrift

Außerdem:

- **<lb/>** (line break): Zeilenumbruch (z.B. wenn Zeilenumbrüche der Vorlage mit transkribiert werden)
 - **<pb/>** (page break): Seitenumbruch, normalerweise der der Vorlage
- ... mit **@n**, **@type**, **@xml:id** spezifizierbar

Strukturelle Gliederungselemente



TEI-Guidelines: 4 Default Text Structure

- **<div>** (Division): Abschnitte im Dokument
z.B. Buch, Kapitel, einzelne Gedichte, Akte/Szenen u.ä.
- Wichtige Attribute:
 - **@n** (Nummerierung)
z.B. "1.1.2.a" (entweder aus der Quelle übernommen
oder selbst erstellt)
 - **@type**
z.B. "book", "chapter", "poem")
 - **@xml:id** (eindeutiger Identifikator)
 - darf im gesamten Dokument nur einmalig vergeben
werden
 - muss mit einem Buchstaben beginnen (i.d.R. selbst
vergeben oder automatisch erzeugt)

Hervorhebung und wörtliche Rede



- **<hi>** (highlighted)
 - allgemeiner Tag für Hervorhebungen
 - z.B. für *kursiv*, **fett**, `g e s p e r r t` etc.
 - spezifizierbar durch **@rend**
- **<foreign>**, **<emph>**, **<distinct>**
 - verschiedene Hervorhebungsarten bzw. Markierung ‚ungewöhnlicher‘ Textteile (Fremdsprachiges, Betonung, Archaismen etc.)
- **<q>** (quoted)
 - für wörtliche Rede, Fachausdrücke, „sog.“ (anstelle von Anführungsstrichen)
- **<quote>** (quotation) und **<cit>** (citation) für Zitate

- Physikalische Einheiten
 - `<pb/>`, `<cb/>`, `<lb/>`, `<gb/>` → der „break“ beginnt die Einheit
- Strukturelle Texteinheiten
 - `<div>`, `<p>`, `<list>`, `<table>`, `<lg>`, `<index>`
- Semantische Texteinheiten
 - `<head>` (nicht: `<title>!`), `<fw>`, `<note>`, `<quote>`, `<term>`
 - `<ref>`, `<bibl>`, `<rs>`
- Entitäten
 - `<persName>`, `<orgName>`, `<placeName>`, `<name>`
 - `<date>`



- ▶ TEI-Standard stellt mehrere hundert Elemente und Attribute bereit, z.B. für
- ▶ Modularer Aufbau der TEI erlaubt Definition von Untermengen des TEI-Tagsets
- ▶ d.h. mein Schema muss nicht alle Elemente und Attribute der TEI enthalten (customisations)
 - ▶ Textkritische Editionen
 - ▶ Linguistische Corpora
 - ▶ Bibliographische Beschreibung von Handschriften
 - ▶ Verknüpfung von Transkriptionen mit digitalen Faksimiles
- ▶ Eigenes Datenmodell zusammenstellen
 - ▶ Mehr dazu am Freitag mit Stefan Dumont

TEI Guidelines online nutzen



- ▶ Inhaltsverzeichnis
 - ▶ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- ▶ Inspiration im Elementverzeichnis (Appendix C)
 - ▶ <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/REF-ELEMENTS.html>
- ▶ Elementbeschreibung aus Oxygen heraus aufrufen
- ▶ Elementbeschreibung lesen

TEI ... Hilfe?!



- ▶ TEI P5 Guidelines:
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index-toc.html>
- ▶ IDE: Folien, Mailinglist, Mitglieder
- ▶ TEI by Example (tutorial):
<http://www.tbe.kantl.be>
- ▶ Dokumentkonvertierung: <http://www.tei-c.org/oxgarage/>
- ▶ TEI @ Oxford (school material):
<http://tei.oucs.ox.ac.uk/Talks/>;
z.B.: <http://tei.oucs.ox.ac.uk/Talks/2014-01-toronto/>
- ▶ TEI Mailingliste: TEI-L@listserv.brown.edu
- ▶ <http://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L>
- ▶ <http://markmail.org/list/edu.brown.listserv.tei-l>

Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).



Alle darin verwendeten Werke anderer Urheber sind Zitate zu wissenschaftlichem Gebrauch.