

Einführung in TEI

Summer School "Digitale Edition" 2017
Christiane Fritze

#ideschool17



Überblick

- ▶ Was ist TEI?
 - ▶ Ein Konsortium.
 - ▶ Ein Kodierungsstandard.
 - ▶ Eine Community.
- ▶ 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.”

➤ TEI P5 Guidelines (derzeit ~~1.646~~ 1.839 Druckseiten)

- <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- Release 3.2.0. (letztes Update am 10. Juli 2017)
- 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



Frank Fischer @umblaetterer
TEI2German
Translatathon at
[#TEIconf2016](#) in Vienna
(27 Sep 2016):
tei2016.acdh.oew.ac.at/translatathon

02 Sep



Journal of the Text Encoding Initiative

The *Journal of the Text Encoding Initiative* is the official journal of the Text Encoding Initiative Consortium. It publishes selected papers from the annual TEI Conference and Members' Meeting and special issues based on topics or themes of interest to the community or in conjunction with special events or meetings associated with TEI. Its contents are rigorously peer-reviewed.

The objectives of the journal are to disseminate as widely as possible information about the TEI and its applications to scholarship. The *Journal of the Text Encoding Initiative* publishes state-of-the-art reports on electronic textual editing, current trends in TEI encoding, and new use cases for TEI. It provides a forum for articles on the discussion of the interface between the TEI and other communities, and more generally of the role of technological standards in the digital humanities. As a peer-reviewed, regularly published journal, it is the first point of entry to the latest scholarship relating to not only the TEI but also various fields typically associated with TEI practice—for example, digital scholarly editing, linguistic analysis, corpora creation, and much newer areas such as mass digitization, semantic web research, and editing within virtual worlds.

The journal is published at [itei.revues.org](http://tei.revues.org) as part of [Revues.org](http://revues.org), the web platform for journals and book collections of Cléo, the [French Centre for Open Electronic Publishing](#).

Announcements

CFP: Open Topic Issue of the Journal of the Text Encoding Initiative

CFP: Open Topic issue of the Journal of the Text Encoding Initiative

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

Posted: 2017-01-24

[More...](#)

jTEI Issue 9: first 3 articles published

Issue 9 of the Journal of the Text Encoding Initiative (Selected Papers from the 2014 TEI Conference), like issue 8 before it, is being published on a "rolling" basis as articles are completed. The first three articles have now been published on the Journal site: <http://jtei.revues.org/1432>.

Posted: 2016-09-24

[More...](#)

CFP: Issue 10 of the Journal of the Journal of the Text Encoding Initiative: 2015 Conference Issue

CFP: Issue 10 of the Journal of the Text Encoding Initiative

2015 Conference Issue

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

Posted: 2015-12-21

[More...](#)

[More Announcements...](#)

OPEN JOURNAL SYSTEMS

[Journal Help](#)

USER

Username

Password

Remember me

NOTIFICATIONS

- [View](#)
- [Subscribe](#)

FONT SIZE



INFORMATION

- [For Authors](#)



TEI-L Mailingliste

LISTSERV 16.0

[Subscriber's Corner](#) [Email Lists](#) [Log In](#)



TEI-L@LISTSERV.BROWN.EDU



Latest Messages		
Re: heading inside a speech	Janelle Jenstad < log in to unmask >	Sun, 1 Oct 2017 16:16:40 +0000
heading inside a speech	Lou Burnard < log in to unmask >	Sun, 1 Oct 2017 17:07:58 +0100
Re: msDesc - references to digital facsimiles held elsewhere	MLH < log in to unmask >	Sun, 1 Oct 2017 11:40:33 +0000

Search Archives

[Advanced Options](#)

TEI-L [Hide Latest Messages](#)

- [October 2017](#)
- [September 2017](#)
- [August 2017](#)
- [July 2017](#)
- [June 2017](#)
- [May 2017](#)
- [April 2017](#)
- [March 2017](#)
- [February 2017](#)
- [January 2017](#)
- [December 2016](#)
- [November 2016](#)
- [October 2016](#)
- [September 2016](#)
- [August 2016](#)
- [July 2016](#)

Options

[Log In](#)

[Get Password](#)

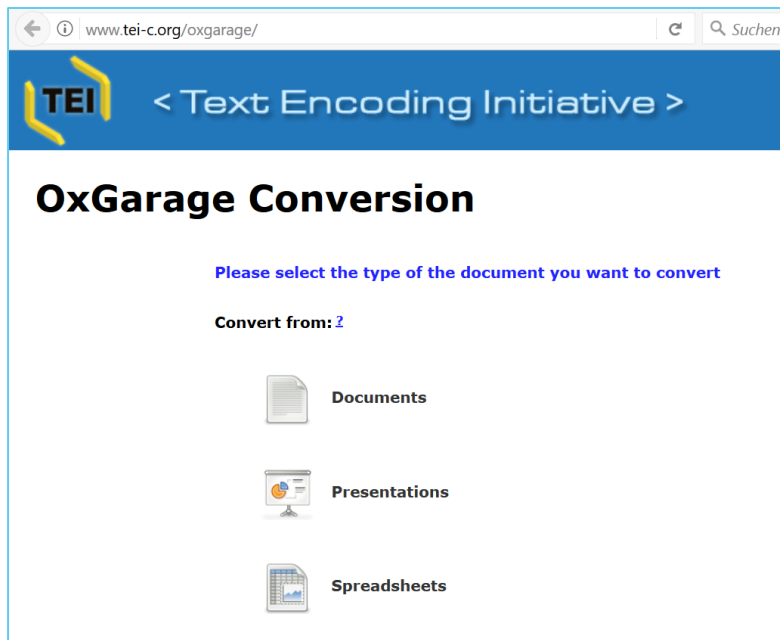
[Search Archives](#)

[Subscribe or Unsubscribe](#)

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

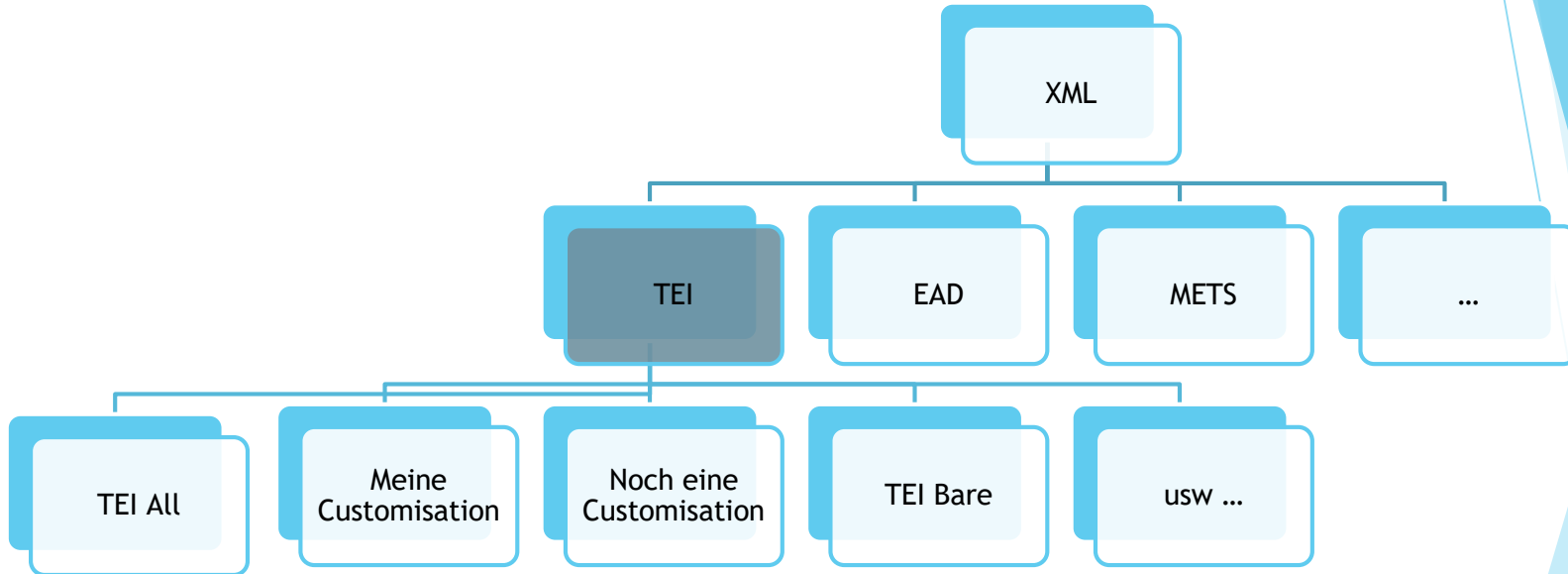
<http://www.tei-c.org/oxgarage/>

- ▶ Speichern Sie das entstandene TEI-Dokument und öffnen Sie 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 nicht gleich TEI.

← → ↻ www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html     

 **< Text Encoding Initiative >**

P5 Guidelines — English

P5: Guidelines for Electronic Text Encoding and Interchange

Version 3.2.0. Last updated on 10th July 2017, revision 0fc651

[English] [Deutsch] [Español] [Italiano] [Français] [日本語] [한국어] [中文]

<h3>Front Matter</h3> <p>Title</p> <ul style="list-style-type: none">i. Releases of the TEI Guidelinesii. Dedicationiii. Preface and Acknowledgmentsiv. About These Guidelinesv. A Gentle Introduction to XMLvi. Languages and Character Sets <h3>Back Matter</h3> <ul style="list-style-type: none">Appendix A Model ClassesAppendix B Attribute ClassesAppendix C ElementsAppendix D Attributes	<h3>Text Body</h3> <ul style="list-style-type: none">1 The TEI Infrastructure2 The TEI Header3 Elements Available in All TEI Documents4 Default Text Structure5 Characters, Glyphs, and Writing Modes6 Verse7 Performance Texts8 Transcriptions of Speech9 Dictionaries10 Manuscript Description11 Representation of Primary Sources12 Critical Apparatus13 Names, Dates, People, and Places14 Tables, Formulae, Graphics, and Notated Music	<h3>TEI sourcecode</h3> <ul style="list-style-type: none">• Getting and Using the TEI Sources.• TEI GitHub Repository• Bug Reports, Feature Requests, etc.
---	---	--

- ▶ 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>
```



Drama

<actor> <move>
 <role>
 <castList>
<performance>

textstructure

<div> <front>
 <opener> <floatingText>
 <text>
<back> <epigraph>

certainty

<certainty>
 <precision>
 <respons>

Core

<p> <item> <note>
 <quote>
<date> <bibl> <foreign>
<list> <abbr> <name>

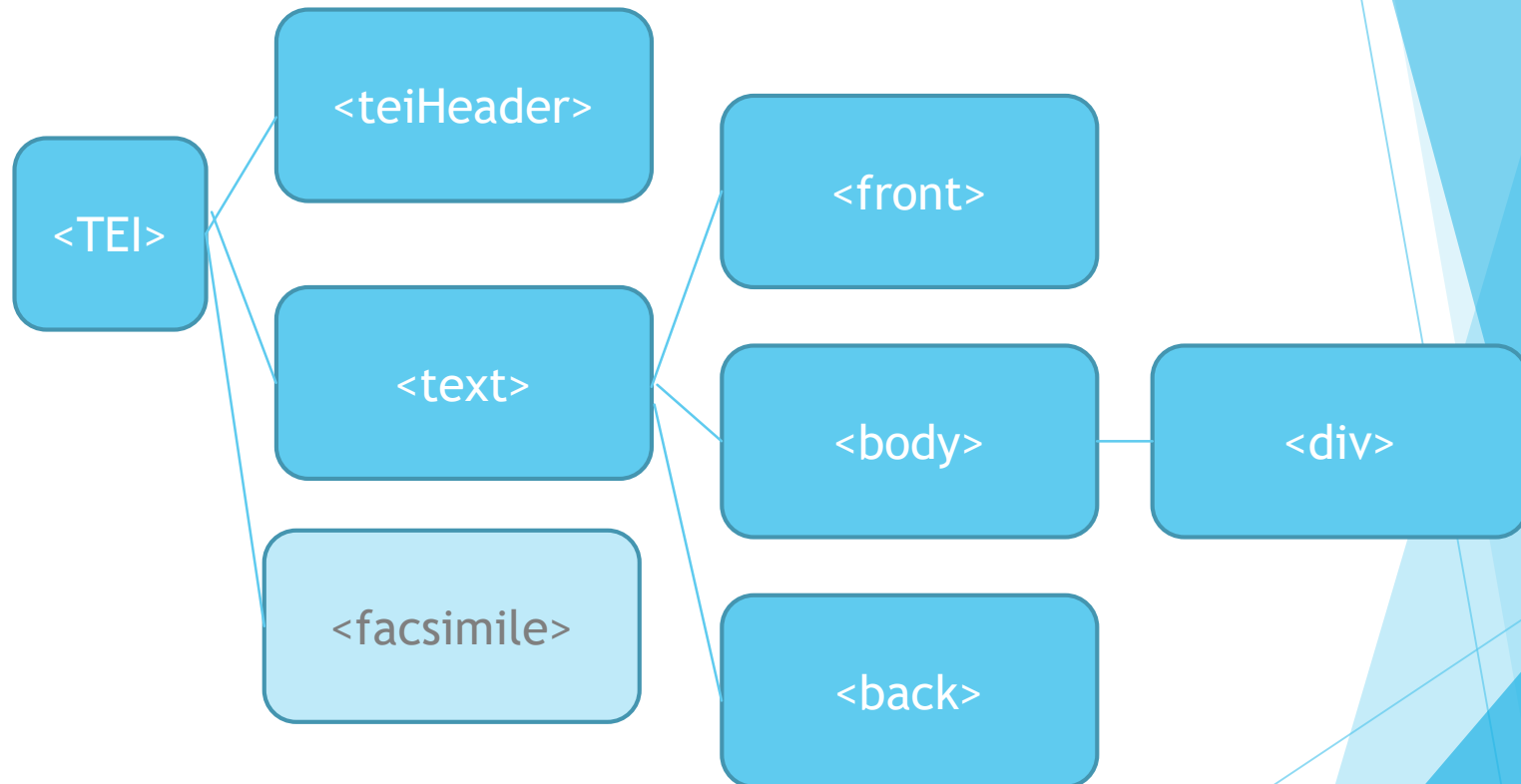
namesdate

<geo> <sex>
 <location> <persName>
<placeName>
 <nym> <relation>

textcrit

<app> <lem>
 <rdg> <lacunaStart>
 <wit> <lacunaEnd>

TEI Grundgerüst



- 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.

<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 Mittwoch 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>

Alles wird gut

