

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



TEI-Einführung 1

Textauszeichnung mit den TEI-Kernelementen

Roman Bleier
roman.bleier@uni-graz.at
Lübeck/Online, 19.08.2020



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



Übersicht

- Text Encoding Initiative – TEI
- TEI Infrastructure/Tools
- TEI Guidelines
- TEI Grundstrukturen
- Gemeinsame Übungen

Danke an meine KollegInnen des IDE, deren Folien ich als Basis für meinen Vortrag verwendet habe.

Text Encoding Initiative – TEI

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"
- 1990: Veröffentlichung des ersten Entwurfs der TEI Guidelines



Vassar College, Poughkeepsie

Text Encoding Initiative – TEI

- TEI ist eine Community, die gemeinsam an einem Standard zur Auszeichnung von geisteswissenschaftlichen Daten arbeitet, Lehrmaterial, Tools, etc. zur Verfügung stellt.

Information:

- TEI P5 Guidelines
 - <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
 - Version 4.0.0. (letztes Update am 13. Feb 2020)
- Lehrmaterial
 - Tutorials
 - TEI by example (<https://teibyexample.org>)

Text Encoding Initiative – TEI

- Journal: jTEI
 - <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 Wiki: https://wiki.tei-c.org/index.php?title=Main_Page
- SIGs und TEI Conference and Members' Meeting

Text Encoding Initiative – TEI

- TEI Conference and Members' Meeting 2019 in Graz



What is text, really? TEI and beyond (TEI 2019)
September 16 – 20, University of Graz, Austria

[Registration](#) [Program](#) [The Conference](#) [The Location](#) [Privacy statement](#)



The image shows a wide-angle photograph of the main building of the University of Graz, a grand, light-colored stone structure with a central dome and classical architectural details. The building is surrounded by greenery and a clear blue sky. In the foreground, there are streetlights and a paved area.

Text Encoding Initiative – TEI

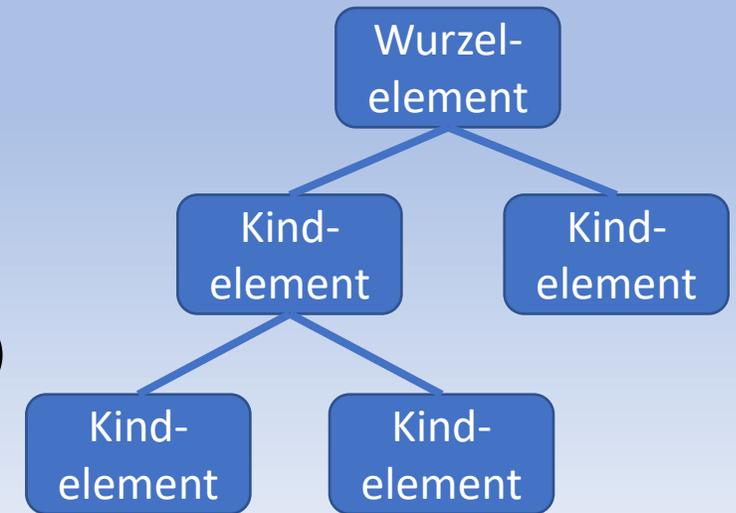
Tools:

- OxGarage: Konversion von und nach TEI
- ROMA: Erstellen von Schemata und Dokumentationen
- Stylesheets: Konversion von TEI-Dokumenten in diverse Formatvorlagen (auch in oXygen – Framework)
- Publishing Tools: z.B. <https://teipublisher.com/index.html>
 - [Weitere Tools: https://wiki.tei-c.org/index.php/Category:Tools](https://wiki.tei-c.org/index.php/Category:Tools)

TEI Guidelines

XML stellt nur die grundlegenden Regeln bereit:

- Datenstruktur:
 - Baumstruktur (ein Wurzelement, korrekte Schachtelung)
- Konventionen für die Kodierung von
 - Elementen (`<elementName />`)
 - Attributen (`<elementName attributeName="Wert" />`)
 - Entity-Referenzen (`&entity;` z.B. `<`)
 - Kommentaren (`<!-- ... -->`) usw.
- Benennungsregeln für Elemente und Attribute



TEI Guidelines

TEI Guidelines:

- Zusätzliche Regeln und eine Einschränkung der 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?

```
5 <Text>
6   <Seitenbruch/>
7   <Absatz>
8     <Zeilenbruch/>Im Nahmen Gottes,
9     <Zeilenbruch/>Zu wißen seÿ hiemitt;
10    <Zeilenbruch/>entwichenem Jahr alhie
11    <Zeilenbruch/>geschriebenen Hänßisch
12    <Zeilenbruch/>bereits hievor in <D
13    <Zeilenbruch/>meinem Rathschlage auf
14  </Absatz>
15 </Text>
```

unendliche
Möglichkeiten in XML

```
5 <doc>
6   <newpage/>
7   <para>
8     <newline/>Im Nahmen Gottes,
9     <newline/>Zu wißen seÿ hiemitt; Alß beÿ der in negst
10    <newline/>entwichenem Jahr alhier in <name>Lubeck</name> auß=
11    <newline/>geschriebenen Hänßischen versamblung die
12    <newline/>bereits hievor in <date>anno 1662</date>. zum ge=
13    <newline/>meinem Rathschlage aufgesetzte puncta
14  </para>
15 </doc>
```

TEI Guidelines bieten Regeln
und Auszeichnungsvorschläge

```
5 <div>
6   <pb/>
7   <p>
8     <lb/>Im Nahmen Gottes,
9     <lb/>Zu wißen seÿ hiemitt; Alß beÿ der in negst
10    <lb/>entwichenem Jahr alhier in <placeName>Lubeck</placeName> auß=
11    <lb/>geschriebenen Hänßischen versamblung die
12    <lb/>bereits hievor in <date>anno 1662</date>. zum ge=
13    <lb/>meinem Rathschlage aufgesetzte puncta
14  </p>
15 </div>
```

TEI Guidelines



TEI ist in 21 Module gegliedert:

- Core: Elemente, die in allen TEI Dokumenten verfügbar sind
- Common Metadata: Elemente für den TEI header
- Transcription of Primary Sources
- Etc.
- TEI Infrastructure: <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ST.html>

TEI Guidelines:

- In 23 Kapiteln und Anhängen wird die Anwendung der TEI Elemente und Attribute beschreiben:
 - <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

TEI Grundstruktur(en)

- Wurzelement <TEI>
- Namespace @xmlns
- Ein TEI Dokument enthält das Element <teiHeader> für Metadaten
- Ein weiteres Element zur Representation des Inhalts der digitalen Ressource:
 - <text> für Text/Transkriptionen
 - <sourceDoc> quellennahe Transkription
 - <facsimile> Bilder

```
7 <TEI xmlns="http://www.tei-c.org/ns/1.0">
8   <teiHeader>
9     <fileDesc>
10      <titleStmt>
11        <title>Title</title>
12      </titleStmt>
13      <publicationStmt>
14        <p>Publication Information</p>
15      </publicationStmt>
16      <sourceDesc>
17        <p>Information about the source</p>
18      </sourceDesc>
19    </fileDesc>
20  </teiHeader>
21  <text>
22    <body>
23      <p>Some text here.</p>
24    </body>
25  </text>
26 </TEI>
27
```

TEI Grundstruktur(en)

für genetische Edition /
topographische Transkriptionen

- quellennahe Transkription
- <sourceDoc>
- <surfaceGrp>, <surface>, <zone>, <line>

```
7 <TEI xmlns="http://www.tei-c.org/ns/1.0">
8   <teiHeader>
9     <fileDesc>
10      <titleStmt>
11        <title>Title</title>
12      </titleStmt>
13      <publicationStmt>
14        <p>Publication Information</p>
15      </publicationStmt>
16      <sourceDesc>
17        <p>Information about the source</p>
18      </sourceDesc>
19    </fileDesc>
20  </teiHeader>
21  <sourceDoc>
22    <surfaceGrp n="leaf1">
23      <surface facs="page1.png">
24        <zone>All the writing on page 1</zone>
25      </surface>
26      <surface>
27        <graphic url="page2-highRes.png"/>
28        <graphic url="page2-lowRes.png"/>
29        <zone>
30          <line>A line of writing on page 2</line>
31          <line>Another line of writing on page 2</line>
32        </zone>
33      </surface>
34    </surfaceGrp>
35  </sourceDoc>
36 </TEI>
```

TEI Grundstruktur(en)

- <facsimile> für Verknüpfung mit Bildern, Bildbeschreibungen, etc.

```
6 <TEI xmlns="http://www.tei-c.org/ns/1.0">
7   <teiHeader>
8     <fileDesc>
9       <titleStmt>
10        <title>A TEI Document containing four page images </title>
11      </titleStmt>
12     <publicationStmt>
13       <p>Unpublished demonstration file.</p>
14     </publicationStmt>
15     <sourceDesc>
16       <p>No source: this is an original work.</p>
17     </sourceDesc>
18   </fileDesc>
19 </teiHeader>
20 <facsimile>
21   <graphic url="page1.png"/>
22   <graphic url="page2.png"/>
23   <graphic url="page3.png"/>
24   <graphic url="page4.png"/>
25 </facsimile>
26 </TEI>
```

Gemeinsame Übung

- TEI P5 Guidelines
 - <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- Suche die Elemente `<persName>` und `<lb>`
 - Wie wird das Element definiert?
 - Welche Attribute sind erlaubt?
 - Welche Kindelemente?
 - Findet ihr die Anwendungsbeispiele?

Gemeinsame Übung

OxGarage:

- Im Übungsordner befindet sich das Word Dokument „Transkript Hanserezess 1669 Lübeck.docx“
- Konvertiere mittels OxGarage das Dokument nach TEI P5:
<http://www.tei-c.org/oxgarage/>
- Speichere das entstandene TEI-Dokument und öffne es im Oxygen-XML-Editor

Gemeinsame Übung

Konvertiere das Word Dokument mit oXygen nach TEI P5

The screenshot displays the oXygen XML Editor interface. On the left, the Project Explorer shows a project named 'CodingTheSources.xpr' with a sub-project 'Coding The Sources Lübeck'. The Archive Browser on the right shows the file structure of the 'Transkript Hanserezess 1669 Lübeck.docx' file, including folders like '_rels', 'word', and 'theme', and files like 'document.xml', 'fontTable.xml', and 'header1.xml'. The main editor window shows the XML code for 'document.xml', starting with the XML declaration and the root element <w:document>. A dialog box titled 'Configure Transformation Scenario(s)' is open in the foreground, showing a table of associations:

Association	Scenario	Type
<input checked="" type="checkbox"/>	DOCX TEI P5	ANT
<input type="checkbox"/>	DOCX DITA	ANT

Below the table, there are buttons for 'New', 'Edit', 'Duplicate', and 'Remove'. The 'Associated scenarios' section shows 'DOCX TEI P5' selected. At the bottom of the dialog, there are buttons for 'Save and close', 'Apply associated (1)', and 'Cancel'. The background XML code shows the beginning of the document structure, including namespace declarations and the start of the <w:body> element.

Weiterführende Literatur:

Lou Burnard: **What is the Text Encoding Initiative?**

<https://books.openedition.org/oep/679?lang=en>

DariahTeach. DH Lehr- und Lernmaterialien. <https://teach.dariah.eu/>

Fragen?

roman.bleier@uni-graz.at



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



Christian-Albrechts-Universität zu Kiel

Historisches Seminar