

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 2

Metadaten, Text und Transkription

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



Forschungsstelle
für die Geschichte
der Hanse und des Ostseeraums



Übersicht

Metadaten, Text und Transkription

- Elemente im TEI Header
- Elemente zur Transkription

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

Grundstruktur eines TEI Dokuments

Zur Wiederholung:

- Wurzelement <TEI>
- <teiHeader> Bereich
- <text> Bereich

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
  <teiHeader>
    <!-- metadata describing a text -->
  </teiHeader>
  <text>
    <!-- a representation of the text itself -->
  </text>
</TEI>
```

Elemente im TEI Header

Einzige verpflichtende Element im **<teiHeader>** ist **<fileDesc>**:

- **<fileDesc>** (Dateibeschreibung)
- bibliographische Beschreibung des TEI Dokuments
- wichtige Kindelemente:
 - **<titleStmt>**
 - **<publicationStmt>**
 - **<sourceDesc>**
 - Etc.

```
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
```

Elemente im TEI Header

<titleStmt>

- Angaben zum Titel
- umfasst Angaben zum Titel des elektronischen Werks und zu den für seinen Inhalt Verantwortlichen

```
3 ▾ <titleStmt>
4   <title>Shakespeare: the first folio (1623) in electronic form</title>
5   <author>Shakespeare, William (1564-1616)</author>
6 ▾   <respStmt>
7     <resp>Originally prepared by</resp>
8     <name>Trevor Howard-Hill</name>
9   </respStmt>
10 ▾  <respStmt>
11    <resp>Revised and edited by</resp>
12    <name>Christine Avern-Carr</name>
13  </respStmt>
14 </titleStmt>
```

```
3 ▾ <titleStmt>
4   <title>La Parisienne</title>
5   <author>Henry BECQUE</author>
6 </titleStmt>
```

Elemente im TEI Header

< publicationStmt >

- Angaben zur Veröffentlichung
- umfasst Angaben zu Veröffentlichung oder Vertrieb eines elektronischen oder sonstigen Textes

```
3 <publicationStmt>
4   <publisher>Chadwyck Healey</publisher>
5   <pubPlace>Cambridge</pubPlace>
6   <availability>
7     <p>Available under licence only</p>
8   </availability>
9   <date when="1992">1992</date>
10 </publicationStmt>
```

```
3 <publicationStmt>
4   <publisher>Oxford University Press</publisher>
5   <pubPlace>Oxford</pubPlace>
6   <date>1989</date>
7 </publicationStmt>
```

Elemente im TEI Header

<sourceDesc>

- Beschreibung der Quellen
- beschreibt die Quelle, von der sich der elektronische Text ableitet
- üblicherweise eine bibliografische Beschreibung

```
<sourceDesc>
  <bibl>The complete writings of Thomas Paine, collected and edited
    by Phillip S. Foner (New York, Citadel Press, 1945)</bibl>
</sourceDesc>
```

```
3 <sourceDesc>
4   <bibl>
5     <title level="a">The Interesting story of the Children in the Wood</title>. In
6     <author>Victor E Neuberg</author>, <title>The Penny Histories</title>.
7     <publisher>OUP</publisher>
8     <date>1968</date>. </bibl>
9 </sourceDesc>
```

Elemente im TEI Header

Weitere Kindelemente des **<teiHeader>**:

- **<fileDesc>** (Dateibeschreibung)
- **<encodingDesc>** (Beschreibung der Kodierung)
 - dokumentiert das Verhältnis zwischen dem elektronischen Text und seiner Quelle oder den Quellen, von denen er sich ableitet
 - die editorische Praxis wird beschrieben
- **<profileDesc>** (Beschreibung des Textprofils)
 - Kodierung inhaltlicher Informationen über den Text, z.B. verwendete Sprachen und Subsprachen, Entstehungsbedingungen eines Textes
- **<revisionDesc>** (Beschreibung der Dateihistorie)
 - Informationen über Änderungen am TEI-Dokument

```
4 <TEI>
5   <teiHeader>
6     <fileDesc>
7       <titleStmt>
8         <title>Title</title>
9       </titleStmt>
10      <publicationStmt>
11        <p>Publication Information</p>
12      </publicationStmt>
13      <sourceDesc>
14        <p>Information about the source</p>
15      </sourceDesc>
16    </fileDesc>
17    <encodingDesc><!--Beschreibung der Kodierung--></encodingDesc>
18    <profileDesc><!-- Beschreibung des Textprofils --></profileDesc>
19    <revisionDesc><!-- Beschreibung der Dateihistorie --></revisionDesc>
20  </teiHeader>
21  <text>
22    <body>
23      <p>Some text here.</p>
24    </body>
25  </text>
26 </TEI>
```

Elemente zur Transkription

- Innerhalb von `<text>`
- `<front>` `<body>` `<back>`
- `<div>` Textabschnitt

```
23 <text>
24   <front>
25     <!--front matter-->
26   </front>
27   <body>
28     <!--text-->
29   </body>
30   <back>
31     <!--back matter-->
32   </back>
33 </text>
```

```
23 <text>
24   <body>
25     <div>
26       <!--Kapitel 1-->
27     </div>
28     <div>
29       <!--Kapitel 2-->
30     </div>
31   </body>
32 </text>
```

Elemente zur Transkription

- Blockelemente
- Absatz: `<p>` paragraph
- Auszeichnung von Versen und Versgruppen:
 - `<l>` line (ein Vers)
 - `<lg>` (eine Versgruppe)
- Alternative Blockstruktur:
 - `<ab>` anonymous block

Elemente zur Transkription

Den Beginn von Seiten, Spalten und Zeilen:

- `<pb>`, `<cb>`, `<lb>`
- `<lb break="no">`

Anführungszeichen:

- `<q>` quoted (Text in Gänsefüßchen)
- `<quote>` quotation (Zitat)

Interpunktion:

- `<pc>` punctuation character

Elemente zur Transkription

Eingriffe durch die AutorIn:

- `<add>` (Hinzufügung)
 - `@place`, z.B. "above", "blow"
- `` (Tilgung)
 - `@rend`, z.B. "strikethrough"

`<subst>` (substitution)

```
Tis moonlight
<subst>
  <del>upon</del>
  <add>over</add>
</subst> Oman's sky
```

```
The story I am going to relate is true as to its main facts,
and as to the consequences <add place="above">of
these facts</add> from which this tale takes its title.
```

Elemente zur Transkription

Eingriffe durch die EditorIn:

- <choice> (Alternative)
- <sic> und <corr> (Korrekturen, seltsame Schreibung)
- <orig> und <reg> (Normalisierung)
- <abbr> und <expan> (Abkürzung)

```
6 indicate that the
7 <choice>
8     <corr>dates</corr>
9     <sic>date's</sic>
10 </choice> mentioned
```

```
4 The address is Southmoor
5 ▾ <choice>
6     <expan>Road</expan>
7     <abbr>Rd</abbr>
8 </choice>
```

Elemente zur Transkription

Eingriffe durch die EditorIn:

- `<gap>` (Auslassung)
 - weist in einer Transkription auf eine Stelle hin, in der Text aus der Vorlage ausgelassen worden ist und
- `<unclear>` (unklare Stelle)
 - enthält ein Wort, einen Satz oder eine Textpassage, welche in der Quelle nicht oder nur schwer lesbar
- `<note>` (Anmerkung)
 - `@type`, `@resp`

Einige weitere Elemente

- Listen: <list> "Container" für die Liste, <item> ein Listeneintrag
- Verweise: <ptr> und <ref> und dem @target
- Zahlen und Maßangaben: <num>, <measure>
- Adressen: <address>
- Elemente f. Dramentexte: <sp> speech (direkte Rede), <speaker> Sprecher, <stage> Regieanweisung
- Namen, Orte: <persName>, <placeName>

Etc.

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

