

Summer School »Digitale Edition«

2. – 5. Oktober 2017



TEI-Vertiefung 1

Metadaten

Stefan Dumont

*enthält Material aus den Folien zu
„TEI-Vertiefung 1: Metadaten, Text und Transkription“
von Roman Bleier, IDE-School 2016*

http://www.i-d-e.de/wp-content/uploads/2016/09/Bleier_TEI_Header_Text.pdf

TEI-Dokument

```
<TEI version="5.0" xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>The shortest TEI Document Imaginable</title>
      </titleStmt>
      <publicationStmt>
        <p>First published as part of TEI P2, this is the P5
          version using a name space.</p>
      </publicationStmt>
      <sourceDesc>
        <p>No source: this is an original work.</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <p>This is about the shortest TEI document imaginable.</p>
    </body>
  </text>
</TEI>
```

TEI-Header

- enthält **Metadaten** über den edierten Text und die Datei, z.B.:
 - Herausgeber
 - Quelle (Archivalie, Druck ...)
 - Entstehungskontext des Textes (Autor, Ort, Zeit ...)
 - Kodierungsrichtlinien ...
- Metadaten sind:
 - normalisiert
 - standardisiert
 - im Idealfall maschinenlesbar (→ Verwendung von URIs, maschinenlesbaren Datumsangaben etc.)

Wozu Metadaten?

- Identifikation
- Beschreibung
- Vernetzung
- Suche
- Darstellung

Struktur des <teiHeader>

- **<fileDesc> (File Description)**
- <encodingDesc> (Encoding Description)
- <profileDesc> (Profile Description)
- <revisionDesc> (Revision Description)

teiHeader/fileDesc

- bibliografische Beschreibung des elektronischen Dokuments (TEI)

- Kindelemente:

<titleStmt>

<publicationStmt>

<sourceDesc>

<titleStmt>

umfasst Angaben zum Titel eines Werks
und für seinen Inhalt Verantwortlichen:

```
<titleStmt>
  <title>Alexander von Humboldt an Karl Ludwig Willdenow. Havanna, 21. Februar
    1801</title>
  <editor>
    <persName ref="http://d-nb.info/gnd/139028285">
      <surname>Päßler</surname>
      <forename>Ulrich</forename>
    </persName>
  </editor>
  <respStmt>
    <persName ref="http://d-nb.info/gnd/120113856">
      <surname>Gerlach</surname>
      <forename>Klaus</forename>
    </persName>
    <resp>
      <note type="remarkResponsibility">Mitarbeit</note>
    </resp>
  </respStmt>
</titleStmt>
```

<publicationStmt>

umfasst Angaben zu Veröffentlichung oder Distribution des TEI Dokuments

```
<publicationStmt>
  <publisher>
    <ref target="http://www.bbaw.de">Berlin-Brandenburgische Akademie
      der Wissenschaften (BBAW)</ref>
  </publisher>
  <pubPlace ref="http://www.geonames.org/2950159">Berlin</pubPlace>
<availability>
  <licence target="https://creativecommons.org/licenses/by-sa/4.0/">
    Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)
  </licence>
</availability>
<idno type="URLWeb">http://edition-humboldt.de/H0001181</idno>
<idno type="URLXML">http://edition-humboldt.de/v2/H0001181.xml</idno>
</publicationStmt>
```

<sourceDesc>

beschreibt den Quelltext, von dem der elektronische Text abstammt oder erzeugt wurde

```
<sourceDesc>
  <msDesc rend="manuscript">
    <msIdentifier>
      <institution>Deutsches Literaturarchiv, Marbach</institution>
      <repository>Handschriftenabteilung</repository>
      <idno>68.593 A:Humboldt</idno>
    </msIdentifier>
  </msDesc>
  <listWit>
    <witness>
      <bibl type="print" n="1">Humboldt 1801a</bibl>
    </witness>
  </listWit>
</sourceDesc>
```

Struktur des <teiHeader>

- <fileDesc> (File Description)
- **<encodingDesc> (Encoding Description)**
- <profileDesc> (Profile Description)
- <revisionDesc> (Revision Description)

<encodingDesc>

dokumentiert das Verhältnis zwischen digitalem Text und seiner/n Quelle(n)

```
<encodingDesc>
```

```
  <p>Basic encoding, capturing lexical information only. All  
    hyphenation, punctuation, and variant spellings normalized. No  
    formatting or layout information preserved.</p>
```

```
</encodingDesc>
```

<encodingDesc>

```
<encodingDesc>
  <projectDesc>
    <p>Originally prepared for use in the production of a series of old-spelling
      concordances in 1968, this text was extensively checked and revised for use during the
      editing of the new Oxford Shakespeare (Wells and Taylor, 1989).</p>
  </projectDesc>
  <editorialDecl>
    <correction method="silent">
      <p>Errors are silently corrected.</p>
    </correction>
    <normalization method="markup">
      <p>Where both upper- and lower-case i, j, u, v, and vv have been normalized, to modern
        20th century typographical practice, the <gi>choice</gi> element has been used to
        enclose <gi>orig</gi> and <gi>reg</gi> elements giving the original and new values
        respectively. ... </p>
    </normalization>
    <normalization method="silent">
      <p>Spacing between words and following punctuation has been regularized to zero spaces;
        spacing between words has been regularized to one space.</p>
    </normalization>
  </editorialDecl>
```

Struktur des <teiHeader>

- <fileDesc> (File Description)
- <encodingDesc> (Encoding Description)
- **<profileDesc> (Profile Description)**
- <revisionDesc> (Revision Description)

<profileDesc>

Inhaltliche Informationen über einen Text, z.B.:

- Autor, weitere Beteiligte
- Entstehungszeit
- Sprachen und Subsprachen

```
<profileDesc>
  <creation>
    <persName>Mathieu, Claude Louis</persName>
    <placeName ref="http://www.geonames.org/2988507">Paris</placeName>
    <date notBefore="1812-01-01" notAfter="1812-12-31" cert="high"/>
  </creation>
</profileDesc>
```

Struktur des <teiHeader>

- <fileDesc> (File Description)
- <encodingDesc> (Encoding Description)
- <profileDesc> (Profile Description)
- **<revisionDesc> (Revision Description)**

<revisionDesc>

- Änderungen am Dokument
- Status des Dokuments im Workflow
- insbesondere sinnvoll, wenn schon veröffentlicht

```
<revisionDesc status="cleared">  
  <change when="1991-11-11" who="#LB"> deleted chapter 10 </change>  
</revisionDesc>
```