

Referat 04: Metadaten mit dem `<teiHeader>` (eine Auswahl)

DARIAH-DE Tutorial *Textkodierung mit TEI*
Christof Schöch (Univ. Würzburg)
Version 1.0, 02/2014

Grundlage der Folien: DH@Oxford 2012, Talk 3.
Lizenz: Creative Commons Attribution 4.0 International (CC-BY)



Gliederung

1. Die Struktur des Headers und der „minimale Header“
2. Erweiterungen des Headers

1. Die Struktur des Headers und der “minimale Header”

Ziele des <teiHeader>s

Grundlegendes Ziel

- Metadaten dokumentieren
- Aber was für Metadaten?

Erste Perspektive: Bibliothekare und Archivare

- bibliographische Metadaten über digitale Objekte
- Strukturierung und Standardisierung

Zweite Perspektive: Textanalysten und Textkodierer

- Kodierungspraxis dokumentieren
- Kombination strukturierte Daten und freie Prosa

TEI Header: 4 Komponenten

<fileDesc> (file description)

- notwendig (als einzige Komponente)
- bibliographische Beschreibung des Dokuments

<encodingDesc> (encoding description)

- Beziehung zwischen kodiertem Text und Quelle

<profileDesc> (text-profile description)

- Viele weitere Aspekte: Sprache, Kontext, etc.

<revisionDesc> (revision description)

- Fasst den Bearbeitungsverlauf des Dokuments zusammen.

Obligatorische Teile der <fileDesc>

<titleStmt> - title statement

- <title> Titel des digitalen Dokuments (nicht der Quelle)

<publicationStmt> - publication statement

- Aussage zum Status des Dokuments bzgl. Publikation
- Mindestens ein unstrukturierter Absatz (<p>)

<sourceDesc> - source description

- Beschreibung der Quelle, die dem digitalen Dok. zugrunde liegt
- Mindestens ein unstrukturierter Absatz (<p>)

Der minimale Header

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Titel des digitalen Dokuments: was?</title>
    </titleStmt>
    <publicationStmt>
      <p>Publikationskontext: wer, wie, wann?</p>
    </publicationStmt>
    <sourceDesc>
      <p>Quelle des Dokuments: woher?</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Beispiel: minimaler Header

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>My 'Strange Meeting' document</title>
    </titleStmt>
    <publicationStmt>
      <p>An exercise for learning TEI.</p>
    </publicationStmt>
    <sourceDesc>
      <p>The primary resource of this file is <ref
        target="http://www.oucs.ox.ac.uk/wllit/
        collections/item/3350"> Strange Meeting</ref> from
        Jon Stallworthy's edition, available on the WWI
        Poetry Digital Archive. </p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```


2. Erweiterungen für den Header

Erweiterungen für das <titleStmt>

Standard-Rollen

- <author> - Autor des kodierten Texts oder des „born digital“ Dokuments
- <editor> - Herausgeber des kodierten Texts
- <principal> - wichtigster beteiligter Forscher, Projektleiter

und, für spezifischere Rollen

- <respStmt> mit <resp> und <name>
- <resp> - frei definierbare Aufgabe, bspw. „Transkription und Kodierung“ oder „Kodierung der Personen- und Ortsnamen“

Beispiel <titleStmt>

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Monster: eine XML-Edition</title>
      <author>Wir sind Helden</author>
      <editor>Franz Sahle</editor>
      <respStmt>
        <resp>Kodierung durch: </resp>
        <name>Patrick Fischer</name>
      </respStmt>
    </titleStmt>
    <!-- (restlicher Header) --!>
  </fileDesc>
</teiHeader>
```

Erweitertes <publicationStmt>

```
<publicationStmt>
  <publisher>TEI @ Oxford</publisher>
  <distributor>DH@Oxford Summer School</distributor>
  <authority>James Cummings</authority>
  <pubPlace>
    <address>
      <orgName>Oxford Univ. Computing Services</orgName>
      <street>13 Banbury Road</street>
      <settlement>Oxford</settlement>
      <postCode>OX2 6NN</postCode>
      <country>United Kingdom</country>
    </address>
  </pubPlace>
  <date when="2012-07-03">3 July 2012</date>
  <idno>dhoxss-tei-ex03</idno>
  <availability>
    <licence>Licensed with a <ref target="[URL]">Creative
      Commons Attribution</ref> licence.</licence>
  </availability>
</publicationStmt>
```

Erweiterte <sourceDesc>

Mehrere Möglichkeiten

- Unstrukturiert: freie Prosa in einem <p>
- Leicht strukturiert: Mischung von Prosa und bibliographischen Elementen wie <author> oder <publisher>, in einem <bibl>
- Stark strukturiert mit <biblStruct>: spezielle Elemente, sodass bibliographische Standards respektiert werden (und Zitierstile oder Datenimport möglich werden)

Außerdem spezielle Elemente:

- für gesprochene Sprache: <recordingStmt>
- für Manuskripte: <msDesc>
- Referenz-Listen: <listPerson>, <listPlace>, <listOrg>

Beispiel: erweiterte <sourceDesc>

```
<sourceDesc>
  <biblStruct>
    <title>Letter to Leslie Gunston</title>
    <author>Wilfred Owen</author>
    <monogr>
      <title>The Wilfred Owen Collection</title>
      <ref target="[URL]">The First World War Poetry Digital
      Archive</ref>
      <imprint>
        <publisher>The First World War Poetry Digital
        Archive</publisher>
        <pubPlace>Oxford</pubPlace>
        <biblScope type="pp" n="2">Two pages</biblScope>
      </imprint>
    </monogr>
    <relatedItem>
      <bibl>The source of this digital resource is a copy from the
      <distributor>Harry Ransom Centre</distributor>.</bibl>
    </relatedItem>
  </biblStruct>
</sourceDesc>
```

Optionale Komponenten der <fileDesc>

<editionStmt>

- Auch digitale Editionen können mehr als eine Auflage/Version haben

<seriesStmt>

- Auch sie können in Reihen publiziert werden

<extent>

- Angabe des Umfangs der Edition, in einem geeigneten Maß
- kb, MB, GB, Zeichen, Wörter, Zeilen, etc.

<notesStmt>

- Anmerkungen zu Aspekten, die anderswo nicht abgedeckt sind
- Eine oder mehrere <note> Elemente

Beispiel: Ausgabe und Umfang

```
<teiHeader>
  <fileDesc>

    <!-- (andere Angaben) --!>

    <editionStmt>
      <edition>Version 2.1</edition>
    </editionStmt>
    <extent>123kb</extent>
  </fileDesc>
</teiHeader>
```


Encoding Description

<encodingDesc>

- enthält Informationen zum Kodierungs-Vorgang
- entweder als Prosa in einem <p>, oder mit speziellen Elementen

Elemente

- <projectDesc>: Ziele des Projekts
- <samplingDecl>: Prinzipien der Auswahl der Texte
- <editorialDecl>: Editionsprinzipien, bspw. <normalization>, <hyphenation>, <segmentation>.
- <classDecl>: classification system/s used
- sowie einige weitere Elemente

Beispiel <encodingDesc>

```
<encodingDesc>
  <projectDesc>
    <p>TEI@Oxford project created materials for
    DHOXSS</p>
  </projectDesc>
  <editorialDecl>
    <correction>
      <p>Apparent errors have been marked as <gi>sic</gi>
      but correct readings not provided</p>
    </correction>
    <hyphenation>
      <p>Hyphens have been transcribed as they
      appear.</p>
    </hyphenation>
  </editorialDecl>
</encodingDesc>
```

Profile Description: "diverse Infos"

<creation>

- Informationen über den Ursprung des Texts, bspw. Zeit und Ort
- weitere Details in der <msDesc>

<langUsage>

- Informationen über die Sprachen, Register, Schriftsysteme, im Text

<textDesc> and <textClass>

- Mechanismen zur Klassifikation des Textes oder der Texte

<handNotes>:

- Informationen über die Schriften ("hands") in einem Manuskript

Beispiel für <textDesc>

```
<profileDesc>
  <textDesc n="novel">
    <channel mode="w">print; part issues</channel>
    <constitution type="single"/>
    <derivation type="original"/>
    <domain type="art"/>
    <factuality type="fiction"/>
    <interaction type="none"/>
    <preparedness type="prepared"/>
    <purpose type="entertain" degree="high"/>
    <purpose type="inform" degree="medium"/>
  </textDesc>
</profileDesc>
```

- Anmerkung: überwiegend leere Elemente!

Revision Description

- Zur Dokumentation von Überarbeitungen des TEI-Dokuments
- neuester Eintrag ganz oben

Struktur

- Liste von `<change>` Elementen
- Attribute: `@date` und `@who`
- kann in eine `<listChange>` gefasst werden

Achtung!

- Im Kontext direkt untergeordnet zu `<teiHeader>`: `<change>` bezieht sich auf das TEI-Dokument.
- Im Kontext von `<creation>` (in `<sourceDesc>`): `<change>` bezieht sich auf die kodierte Quelle

Beispiel <revisionDesc>

```
<teiHeader>
  <!-- andere Inhalte --!>

  <revisionDesc>
    <listChange>
      <change when="2012-07-03" who="#JC">
<persName>James
      Cummings</persName> improved the header.</change>
      <change when="2012-02" who="#RvB"> <persName>Renée
van Baalen</persName> transcribed the <title>Letter
to Leslie Gunston</title> document. </change>
    </listChange>
  </revisionDesc>
</teiHeader>
```

- Anmerkung: "#JC" und "#RvB" für @who beziehen sich auf @xml:ids, die anderswo vergeben werden, bspw. in den <respStmnt>s.

Noch nicht das Ende, aber:

Es gibt noch einiges mehr zum <teiHeader> zu sagen, aber wir belassen es hier einmal.

- der Header ist eine Wissenschaft für sich
- die Manuskript-Beschreibung lohnt eine eigene Session