

TEI: TEIHEADER AND MANUSCRIPT DESCRIPTION
TORSTEN SCHAßAN
FEBRUARY 21, 2024
MASTER CLASS 'DIGITAL SCHOLARLY EDITING'

H E R Z O G
A U G U S T
B I B L I O
T H E K

TEI teiHeader and Manuscript Description



Handschriftenportal

DFG Deutsche
Forschungsgemeinschaft



Credits

- The contents of the presentation is inspired by and uses some texts and examples from a presentation by Helmut W. Klug and Elisabeth Raunig, 2019.
- Examples are taken from various resources, amongst them
 - TEI Guidelines
 - Handschriftenportal
 - ZfdG
 - various HAB editions

HERZOG
AUGUST
BIBLIO
THEK

teiHeader

Overall structure

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">  
  <teiHeader></teiHeader>  
  <facsimile></facsimile>  
  <text></text>  
  <sourceDoc></sourceDoc>  
</TEI>
```

<teiHeader>

<teiHeader> (TEI header) supplies descriptive and declarative metadata associated with a digital resource or set of resources.

- One of the few elements unconditionally required in any TEI document.
- First child within <TEI> or <teiCorpus>
- Information about
 - the file, its publication, its source(s),
 - the encoding rules,
 - detailed description of non-bibliographic aspects of a text
 - the revision history of the file

<teiHeader>, minimal

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title><!-- Title of the resource --></title>
    </titleStmt>
    <publicationStmt>
      <p><!-- Publication information --></p>
    </publicationStmt>
    <sourceDesc>
      <p><!-- Information about the source --></p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

TEI element naming conventions

- **<xyStmt>** = Statement, grouping elements in a structured way
- **<xyDecl>** = Declaration, describing encoding practices in a structured way
- **<xyDesc>** = Description, contains elements in pure prose (<p>) or in a structured way

<fileDesc>

<fileDesc> contains a full bibliographic description of an electronic file.

- **<titleStmt>** information about the title of a work and those responsible for its content
- **<editionStmt>** information relating to one edition of a text.
- **<extent>** size of a text stored on some carrier medium
- **<publicationStmt>** information concerning the publication or distribution of an electronic or other text
- **<seriesStmt>** information about the series, if any
- **<notesStmt>** any notes additional to that recorded in other parts of the bibliographic description.
- **<sourceDesc>** describes the source(s) from which an electronic text was derived or generated

<titleStmt>, example

```

<titleStmt>
  <title>Shakespeare: the first folio (1623) in electronic form</title>
  <author>Shakespeare, William (1564–1616)</author>
  <respStmt>
    <resp>Originally prepared by</resp>
    <name>Trevor Howard-Hill</name>
  </respStmt>
  <respStmt>
    <resp>Revised and edited by</resp>
    <name>Christine Avern-Carr</name>
  </respStmt>
</titleStmt>

```

(cf. TEI Guidelines, “Show all”)

Assigning roles

- Digital editions are usually work of many contributors. But often, many of the contributors weren't made visible, at least not in a computable and/or searchable way.
- “CRediT (Contributor Roles Taxonomy) is a high-level taxonomy, [...] that can be used to represent the roles typically played by contributors to research outputs. The roles describe each contributor's specific contribution to the scholarly output.” (<https://credit.niso.org/>)
- Typical DH-roles include: Data curation, Methodology, Software, Visualization
- Additionally, the MARC Relators vocabulary might be applied (<https://id.loc.gov/vocabulary/relators.html>)

Assigning roles, cont.

- It is equally possible to
 - list all contributors and assign them all contributions (ZfdG example)
 - list all contributions and assign them the contributors (example „Netzwerke der Nonnen“)
- The choice might depend on the intended primary presentation output, e.g. bibliographical entry vs. website

Rights declarations

- In the digital sphere, the rights declaration is crucial for the (re)use of data.
- The rights are declared in the publicationStmt/availability, best with explicit licencing
- In a TEI file might be as many licences declared as needed, e.g. für transcription, translation, and digital facsimile

<availability>, example

```
<publicationStmt>  
  <publisher></publisher>  
  <date></date>  
  <availability status="free | restricted">  
    <licence target="https://creativecommons.org/...">  
      <p></p>  
    </licence>  
  </availability>  
  <pubPlace></pubPlace>  
</publicationStmt>
```

Licences

- Free and open licences (for creative works) are:
 - Public Domain Mark
 - CC0
 - CC BY
 - CC BY-SA
- other CC licences are not considered as „free“ licences
- Common software licences are:
 - MIT licence
 - GNU General Public License

<sourceDesc>

<sourceDesc> describes the source(s) from which an electronic text was derived or generated, typically a bibliographic description in the case of a digitized text, or a phrase such as "born digital" for a text which has no previous existence.

- recording information (of spoken text)
- script information (used for spoken text)
- authority lists (listXY)
- manuscript description(s)

<sourceDesc>, example

```
<sourceDesc>
```

```
  <p>Born digital: no previous source exists.</p>  
</sourceDesc>
```

```
<sourceDesc>
```

```
  <p>OCR generierter Text, TEI-Rohkodierung durch die BSB,  
Konversion nach TEI-P5 durch Torsten Schaßan.</p>  
</sourceDesc>
```


<encodingDesc>

<encodingDesc> documents the relationship between an electronic text and the source or sources from which it was derived.

- **<transcriptionDesc>** describes the set of transcription conventions used, particularly for spoken material.
- **<projectDesc>** describes in detail the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.
- **<variantEncoding>** declares the method used to encode text-critical variants.

<encodingDesc>, cont.

- **<editorialDecl>** provides details of editorial principles and practices applied during the encoding of a text.
- **<classDecl>** contains one or more taxonomies defining any classificatory codes used elsewhere in the text.
- **<refsDecl>** specifies how canonical references are constructed for this text.
- **<tagsDecl>** provides detailed information about the tagging applied to a document.

The editing guidelines would have been declared in the introduction of the edition in print, but can be made available as explicit, and more importantly: machine-readable, computationally useful rules!

<encodingDesc>

```

<encodingDesc>
  <classDecl>
    <taxonomy xml:id="gnd">
      <bibl>Gemeinsame Normdatei</bibl>
    </taxonomy>
  </classDecl>
  <projectDesc>
    <p>Beschrieben im Rahmen des Projektes
      <name type="project"></name>.
    </p>
  </projectDesc>
</encodingDesc>

```

(cf. Hainhofer edition, Kaempfer edition)

<profileDesc>

<profileDesc> provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.

- **<correspDesc>** contains a description of the actions related to one act of correspondence.
- **<textClass>** groups information which describes the nature or topic of a text in terms of a standard classification scheme, thesaurus, etc.

<correspDesc>, example

```

<profileDesc>
  <correspDesc>
    <correspAction type="sent">
      <rs type="person" ref="#wilde_mechtild">Priorin Mechtild Wilde</rs>
      <rs type="place">Kloster Lüne</rs>
      <date when-custom="1529-08-04" calendar="#julian">4. August 1529</date>
    </correspAction>
    <correspAction type="received">
      <rs type="person">Hartwig Stöterogge</rs>
      <rs type="place">Lüneburg</rs>
    </correspAction>
  </correspDesc>
</profileDesc>

```

<textClass>, example

```
<profileDesc>  
  <textClass>  
    <keywords scheme="#gnd">  
      <term xml:lang="de">Musik</term>  
      <term xml:lang="de">Arithmetik</term>  
    </keywords>  
  </textClass>  
</profileDesc>
```

<revisionDesc>

<revisionDesc> summarizes the revision history of a file.

- **<listChange>** groups a number of change descriptions associated with either the creation of a source text or the revision of an encoded text.
- **<change>** documents a change or set of changes made during the production of a source document, or during the revision of an electronic file.

When to use <revisionDesc>?

- As a general rule of thumb, only revisions of the (textual) content of the file are described verbally. Technical changes to the file might rather be documented by the file history, e.g. in git.

<revisionDesc>, example

```
<revisionDesc>  
  <change who="schassan" when="2023-02-23">  
    Kategorie zum Element textClass hinzugefügt: Arithmetik  
  </change>  
  <change who="schassan" when="2016-04-06">  
    Kategorie zum Element textClass hinzugefügt: Musik  
  </change>  
</revisionDesc>
```


Exercise

Enhance bits of the `teiHeader` your project data.

- Start with an empty TEI file (or use any of your project files!)
- Enhance the `<titleStmt>`
 - Add some responsibilities
- Declare the rights
- If applicable, add `<correspDesc>`

Manuscript Description

<msDesc>

<msDesc> contains a description of a single identifiable manuscript or other text-bearing object such as an early printed book.

- Similar: <object> (Note: not <objectDesc>!)
- TEI definition of the overall structures was oriented at the cataloguing practice (as of 2006, with preference for “non-German” cataloguing rules)

<msDesc>

TEI element is intended for several different kinds of applications:

- standalone database of library records (finding aid)
- discursive text collecting many records (catalogue raisonné)
- metadata component within a digital surrogate (electronic edition)
- tool for ‘quantitative codicology’

Two approaches:

- Use larger text containers and enhance the full-text for searching
- Use the (deep!) substructures the TEI offers

<msDesc>, placement

- <msDesc> can be used everywhere where a bibliographical description (<bibl>) can appear!
- In an edition, it is likely to have it in <sourceDesc>, or the introduction
- In a catalogue, the description is not metadata but data, and therefore to be placed in <text>.

msDesc

```
<msDesc xml:lang="de">  
  <msIdentifier></msIdentifier>  
  <head></head>  
  <msContents></msContents>  
  <physDesc></physDesc>  
  <history></history>  
  <additional></additional>  
  <msPart></msPart>  
</msDesc>
```

Minimal and Core information

Minimal data:

- holding institution → settlement, repository, shelfmark

Core data:

- content → title, textual languages, decoration, musical notation
- materiality → object type, material, extent, dimensions, format; status
- origin → place of origin, date of origin, type of dating
- In the descriptions core fields are defined which contain norm(alised) data

With manuscript descriptions (25767) With digital images (17514) IIF (13940)

89762 OBJECT/S

HOLDING INSTITUTION

Settlement

- München (49068)
- Wolfenbüttel (10124)
- Leipzig (2671)
- Augsburg (2322)
- Bamberg (2262)
- Eichstätt (2230)
- Lübeck (2025)
- Erlangen (1981)

+ More

Institution

MATERIALITY

Object type

Material

Extent

1 to 2142

(Empty)

Use filter

Format

Height

1 to 786 cm

(Empty)

Use filter

Width

ORIGIN

Place of origin

Date of origin

401 to 2000

(Empty)

Period entirely within range

Use filter

Type of date

(Empty) (64763)

Datable (19076)

Dated (7129)

CONTENT

Language

Decoration

(Empty) (84826)

Included (4936)

Musical notation

(Empty) (87560)

Included (2202)

APPLY FILTERS

FEEDBACK

<msIdentifier>

```

<msIdentifier>
  <settlement>Wolfenbüttel</settlement>
  <repository>Herzog August Bibliothek</repository>
  <collection>Novissimi Handschriften in folio</collection>
  <idno>Cod. Guelf. 105 Noviss. 2°</idno>
  <altIdentifier type="alternative">
    <settlement>München</settlement>
    <repository>Bayerische Staatsbibliothek</repository>
    <idno>Clm 30055</idno>
  </altIdentifier>
  <msName>
    Evangeliar Heinrichs d. Löwen und Mathildes von England
  </msName>
</msIdentifier>

```

<head>

<head> contains any type of heading, for example the title of a section, or the heading of a list, glossary, manuscript description, etc.

```
<head>
  <title>
    Evangeliar Heinrichs d. Löwen und Mathildes von England
  </title>
  <origDate notBefore="1101" notAfter="1">12. Jh.</origDate>
  <origPlace>Benediktinerkloster Helmarshausen</origPlace>
</head>
```

Description core fields

```
<head>  
  <index indexName="norm_dimensions">  
    <term type="dimensions">26,5 × 21</term>  
    <term type="height">26,5</term>  
    <term type="width">21</term>  
    <term type="depth"/>  
    <term type="dimensions_typeOfInformation">factual</term>  
  </index>  
  <index indexName="norm_format">  
    <term type="format">quarto</term>  
    <term type="format_typeOfInformation">computed</term>  
  </index>  
  ...  
</head>
```

Description core fields

```

<head>
  <index indexName="norm_origPlace">
    <term type="origPlace">Weißenburg (Elsass) / Bobbio</term>
    <term type="origPlace_norm" ref="http://d-nb.info/gnd/4079134-8">Weißenburg</term>
    <term type="origPlace_norm" ref="http://d-nb.info/gnd/4080441-0">Bobbio</term>
  </index>
  <index indexName="norm_origDate">
    <term type="origDate">8. Jh., 1. Hälfte</term>
    <term type="origDate_notBefore">0701</term>
    <term type="origDate_notAfter">0750</term>
    <term type="origDate_type">datable</term>
  </index>
  ....
</head>

```

<msContents>

<msContents> describes the intellectual content of a manuscript, manuscript part, or other object either as a series of paragraphs or as a series of structured manuscript items.

```
<msContents>  
  <textLang mainLang="la" otherLangs="goh">  
    Latein, mit althochdeutschen Glossen  
  </textLang>  
  <msItem>  
    <locus></locus>  
    <author></author>  
    <title></title>  
    <bibl></bibl>  
  </msItem>  
</msContents>
```

<physDesc>

<physDesc> contains a full physical description of a manuscript, manuscript part, or other object optionally subdivided using more specialized elements

- <objectDesc> contains a description of the physical components making up the object which is being described.
- <handDesc> contains a description of all the different hands used
- <scriptDesc> contains a description of the scripts used
- <musicNotation> contains description of type of musical notation.
- <decoDesc> contains a description of the decoration
- <bindingDesc> describes the present and former bindings
- <additions> contains a description of any significant additions found, such as marginalia or other annotations.

<objectDesc>

```
<objectDesc form="codex">
  <supportDesc material="parch">
    <support>
      <p><material>Pergament</material></p>
    </support>
    <extent>
      <measure type="leavesCount" quantity="225">225 Bl.</measure>
      <dimensions type="leaf" unit="cm">
        <height quantity="34.2">34,2</height>
        <width quantity="25.5">25,5</width>
      </dimensions>
    </extent>
  </supportDesc>
</objectDesc>
```

<history>

<history> groups elements describing the full history of a manuscript, manuscript part, or other object.

- **<origin>** contains any descriptive or other information concerning the origin of a manuscript, manuscript part, or other object.
- **<provenance>** contains any descriptive or other information concerning a single identifiable episode during the history of a manuscript, manuscript part, or other object after its creation but before its acquisition.
- **<acquisition>** contains any descriptive or other information concerning the process by which a manuscript or manuscript part or other object entered the holding institution.

<history>

```
<history>
  <origin>
    <p>
      <origDate notBefore="1101" notAfter="1">12. Jh.</origDate>
      <origPlace>Benediktinerkloster Helmarshausen</origPlace>
    </p>
  </origin>
  <provenance><p></p></provenance>
  <acquisition><p></p></acquisition>
</history>
```

<additional>

<additional> groups additional information, combining bibliographic information about a manuscript or other object, or surrogate copies of it, with curatorial or administrative information.

- **<adminInfo>** contains information about the present custody and availability of the manuscript or other object, and also about the record description itself.
- **<surrogates>** contains information about any representations of the manuscript or other object being described which may exist in the holding institution or elsewhere.
- **<listBibl>** contains a list of bibliographic citations of any kind.

<additional>

```
<additional>
  <adminInfo>
    <recordHist>
      <source><bibl></bibl></source>
    </recordHist>
    <custodialHist><p></p></custodialHist>
  </adminInfo>
  <surrogates></surrogates>
  <listBibl>
    <bibl></bibl>
  </listBibl>
</additional>
```

Digital facsimile

```
<surrogates>
```

```
<bibl>
```

```
<!--<ref type="manifest" target=""/>-->
```

```
<ref type="other" target="https://diglib.hab.de/mss/64-weiss/start.htm"/>
```

```
<ref type="thumbnail" target="http://diglib.hab.de/mss/64-weiss/thumbs/00554.jpg"/>
```

```
<date when="2011-03-31" type="digitized"/>
```

```
<date when="2011-03-31" type="published"/>
```

```
<orgName ref="8989-8" type="digitizing">Herzog August Bibliothek Wolfenbüttel</orgName>
```

```
<placeName ref="4066832-0" type="digitizing">Wolfenbüttel</placeName>
```

```
</bibl>
```

```
</surrogates>
```

<msPart>

<msPart> contains information about an originally distinct manuscript or part of a manuscript, which is now part of a composite manuscript.

- <msPart> has the same overall structure as <msDesc>
- Discussion, whether other parts of a manuscript that might have its own history etc may be called 'msPart' as well, e.g. bindings, fragments.

HERZOG
AUGUST
BIBLIO
THEK