



Metadata: `teiHeader`, `msDesc`

[Helmut W. Klug, Elisabeth Raunig]

CLARIAH-AT

TEI Introductory School

10.-13. September 2019

Venue:

Centre for Information Modelling (ZIM)
Elisabethstraße 59/III, SR 81.31

Programm

Tuesday 10.9.

9:00-10:30 *Digital Scholarly Editing*

Roman Bleier, Martina Scholger

11:00-12:30 *Textencoding with XML*

Hans Clausen, Gerlinde Schneider

14:00-15:30 *TEI – General Introduction*

Hans Clausen, Gerlinde Schneider

16:00-17:30 *Practice: XML/TEI*

Hans Clausen, Gerlinde Schneider

Programm

Wednesday 11.9.

- 9:00-10:30 *Metadata:teiHeader, msDesc*
Helmut W. Klug, Elisabeth Raunig
- 11:00-12:30 *Text and transcription*
Sean M. Winslow
- 14:00-15:30 *Practice*
Helmut W. Klug, Elisabeth Raunig, Sean M. Winslow
- 16:00-18:00 *Presentations and social event*

Programm

Thursday 12.9.

9:00-10:30 *TEI genre-specific encoding: letters*
Roman Bleier

11:00-12:30 *TEI Customization*
Martina Scholger

14:00-15:30 *Names, people, places*
Christian Steiner

16:00-17:30 *Practice: letters*
Roman Bleier, and Martina Scholger

Programm

Friday 13.9.

9:00-10:30 *XPath*

Syd Bauman

11:00-12:30 *Schematron, practice*

Syd Bauman

14:00-15:30 *Easy TEI publication (CETEEIcean)*

Raffaele Viglianti

16:00-17:30 *Case Studies: TEI for data-rich documents*

James Cummings

TEI header

- metadata on the TEI-XML-file, the source, persons involved, publication information, encoding rules etc.

Placement of <teiHeader>

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">  
  <teiHeader>  
    <fileDesc>  
      <titleStmt></titleStmt>  
      <publicationStmt></publicationStmt>  
      <sourceDesc></sourceDesc>  
    </fileDesc>  
  </teiHeader>  
  <text>  
    <body>  
      <p>Some text here.</p>  
    </body>  
  </text>  
</TEI>
```

metadata

transcription

<teiHeader>, parts of

<fileDesc>

<encodingDesc>

<profileDesc>

<xenoData>

<revisionDesc>

Prose, Decl., Desc., Stmt.

Types of content in the <teiHeader>

- Declaration: encoding practices, structured
- Statement: grouping elements, structured
- Description: Prose (paragraphs, phrases) or structured

<teiHeader>, minimal

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>
        <!-- title of the resource -->
      </title>
    </titleStmt>
    <publicationStmt>
      <p>
        <!-- Information about distribution of the resource -->
      </p>
    </publicationStmt>
    <sourceDesc>
      <p>
        <!-- Information about source from which the resource derives -->
      </p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

<teiHeader>, corpus

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">  
  <teiHeader>  
    <!-- corpus-level metadata here -->  
  </teiHeader>  
  <TEI>  
    <teiHeader>  
      <!-- metadata specific to this text here -->  
    </teiHeader>  
    <text>  
      <!-- ... -->  
    </text>  
  </TEI>  
  <TEI>  
    <teiHeader>  
      <!-- metadata specific to this text here -->  
    </teiHeader>  
    <text>  
      <!-- ... -->  
    </text>  
  </TEI>  
</teiCorpus>
```

<revisionDesc>

- summarizes the revision history for a file

```
<revisionDesc>
  <list>
    <item>
      <date when="1989-04-12">12 Apr 89</date> Last checked by CAC</item>
    <item>
      <date when="1989-03-01">1 Mar 89</date> LB made new file</item>
    </list>
  </revisionDesc>
```

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

<xenoData>

- provides a container element for metadata in non-TEI formats

```
<xenoData>
  <oai_dc:dc>
    <dc:title>The colonial despatches of Vancouver Island and British
      Columbia 1846-1871: 11566, CO 60/2, p. 291; received 13 November.
      Trevelyan to Merivale (Permanent Under-Secretary)</dc:title>
    <dc:date>1858-11-12</dc:date>
    <dc:creator>Trevelyan</dc:creator>
    <dc:publisher>University of Victoria Humanities Computing and Media
      Centre, and UVic Libraries</dc:publisher>
    <dc:type>InteractiveResource</dc:type>
    <dc:format>application/xhtml+xml</dc:format>
```

<profileDesc>

- abstract
- detailed description of non-bibliographic aspects of a text
- the languages and sublanguages used
- the situation in which it was produced
- the participants and the setting

<profileDesc>

```
<profileDesc>
  <langUsage>
    <language ident="fr">français</language>
  </langUsage>
  <textDesc n="roman">
    <channel mode="w">copie; extraits </channel>
    <constitution type="single"/>
    <derivation type="original"/>
    <domain type="art"/>
    <factuality type="fiction"/>
    <interaction type="none"/>
    <preparedness type="prepare"/>
    <purpose type="distraction" degree="high"/>
    <purpose type="information"
      degree="medium"/>
  </textDesc>
  <settingDesc>
    <setting>
      <name>Paris, France</name>
      <time>Fin 19e</time>
    </setting>
  </settingDesc>
</profileDesc>
```

<profileDesc>

```
<profileDesc>
  <abstract>
    <p>This paper is a draft studying
      various aspects of using the TEI
      as a reference serialization framework
      for LMF. Comments are welcome to bring
      this to a useful document for the
      community.
    </p>
  </abstract>
</profileDesc>
```

<encodingDesc>

- documents the relationship between an electronic text and the source

<projectDesc> description of the TEI project

<samplingDecl> details what text(s) have been selected, eg. corpus?

<editorialDecl> details of editorial principles and practices applied during the encoding of a text

<tagsDecl> information about the tagging applied to a document

...

<encodingDesc>

```
<encodingDesc>
```

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

```
</encodingDesc>
```

```
<encodingDesc>
```

```
  <projectDesc>
```

```
    <p>Corpus de  
      textes sélectionnés pour la formation MISAT, Frejus,  
      juillet 2010.
```

```
    </p>
```

```
  </projectDesc>
```

```
  <samplingDecl>
```

```
    <p>Corpus  
      d'échantillons de 2000 mots pris au début de chaque texte. </p>
```

```
  </samplingDecl>
```

```
</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>
      <p>Turned letters are silently corrected.</p>
    </correction>
    <normalization>
      <p>Original spelling and typography is retained, except that long s and ligatured
        forms are not encoded.</p>
    </normalization>
  </editorialDecl>
  <refsDecl xml:id="ASLREF">
    <cRefPattern matchPattern="(\\S+) ([^.]+)\\.(.*)"
      replacementPattern="#xpath(//div1[@n='$1']/div2/[@n='$2']//lb[@n='$3'])">
      <p>A reference is created by assembling the following, in the reverse order as that
        listed here: <list>
          <item>the <att>n</att> value of the preceding <gi>lb</gi>
          </item>
          <item>a period</item>
          <item>the <att>n</att> value of the ancestor <gi>div2</gi>
          </item>
          <item>a space</item>
          <item>the <att>n</att> value of the parent <gi>div1</gi>
          </item>
        </list>
      </p>
    </cRefPattern>
  </refsDecl>
</encodingDesc>
```

<fileDesc>

- full bibliographic description of an electronic file

<editionStmt> information relating to one edition of a text

<extent> size of a text stored on some carrier medium

<notesStmt> notes additional to bibliographic information

<publicationStmt> information concerning publication or distribution

<seriesStmt> information about the series, if any

<sourceDesc> describes the source from which an electronic text was derived or generated

<titleStmt> information about the title of a work and those responsible for its content

<fileDesc>

```
<fileDesc>
  <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>
  <publicationStmt>
    <distributor>Oxford Text Archive</distributor>
    <address>
      <addrLine>13 Banbury Road, Oxford OX2 6NN, UK</addrLine>
    </address>
    <idno type="OTA">119</idno>
    <availability>
      <p>Freely available on a non-commercial basis.</p>
    </availability>
    <date when="1968">1968</date>
  </publicationStmt>
  <sourceDesc>
    <bibl>The first folio of Shakespeare, prepared by Charlton Hinman (The Norton Facsimile,
      1968)</bibl>
  </sourceDesc>
</fileDesc>
```

<sourceDesc>, general

- Recording Information (spoken text)
- Authority lists

<sourceDesc>, bibliogr.

- either: prose description
- or: bibliographical information
<bibl>, <biblStruct>, <biblFull>
- or: manuscript description
<msDesc>

Exercise

Your project is the transcription of a historic cookbook.

=> Prepare the TEI header as detailed as possible!

<https://bit.ly/2kaqZx4>

<sourceDesc> <msDesc>

Why Manuscript Descriptions?

- Used to describe Manuscripts AND other text-bearing objects.
- Anything written that is not a print publication.

additional information

Additional Information:

- Manuscripts are unique objects, sometimes of cultural or political value.
- ancient or modern manuscripts
- it benefits us to use the same forms of description
- no formalised bibliographic conventions for manuscripts.

The 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'

msDesc

One of the Description elements. It can contain a prose description, possibly, but not necessarily, organised under some specific headings by suggested sub-elements.

Placement: where you can put a `<p>`, `<msDesc>` is valid too, but the best place for us is the `teiHeader` as a child-element of the `<sourceDesc>`.

Guidelines:

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>

Main Elements in <msDesc>

<msIdentifier>

<head>

<msContents>

<physDesc>

<history>

<additional>

other Elements:

<msPart>

<msFrag>

About <msIdentifier>

Only Element that is required and has to be structured
Can be followed by just <p> Elements

Example Description of Ms. 1609

```
<msIdentifier>  
  Place { <country>Austria</country>  
           <region>Styria</region>  
           <settlement>Graz</settlement>  
  Repository { <institution>Karl-Franzens-Universität Graz</institution>  
                 <repository>Universitätsbibliothek</repository>  
                 (<collection>Abteilung für Sondersammlungen</collection>)  
  Identifier { <idno>1609</idno>  
                 <altIdentifier>  
                   <idno>34/8</idno>  
                 </altIdentifier>  
  <msName>Graz, UB, Ms. 1609</msName>  
</msIdentifier>
```

About <head>

most important information like:

<title>

and e.g.:

<origDate>

<origPlace>

<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-head.html>

About `<msContent>`

TEI Definition: (manuscript contents) 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.

Can include a description of the language(s) used in the Manuscript.

Example

```
<msContents>  
  <textLang mainLang="gmh" otherLangs="lat-med-deu">  
    <!-- prose description -->  
  </textLang>  
  <msItem></item>  
</msContents>
```

or:

```
<msContents>  
  <p></p>  
</msContents>
```

full example

```
<msContents>
  <textLang mainLang="bar-vmf" otherLangs="">
    Nordbairisch/Ostfränkisch (Bavarian)</textLang>

  <msItem xml:id="_1v">
    <locus>1v</locus>
    <title>empty</title>
    <incipit>empty</incipit>
  </msItem>
  <msItem xml:id="_2r">
    <locus>2r</locus>
    <title>recipes and Notae</title>
    <textLang>Nordbairisch/bavarian</textLang>
    <incipit><locus></locus> </incipit>
    <note>(recipes for gingerbread, sauces)</note>
  </msItem>
</msContents>
```

About <physDesc>

Any physical descriptions/information about:

<objectDesc> (see next slide)

<handDesc> writer hands

<scriptDesc> which type of script

<musicNotation> music

<decoDesc> decoration

<bindingDesc> binding of the MS

<sealDesc> added seals

<accMat> accompanying/additional material

About <objectDesc>

physical components of described object:

<layoutDesc> layout

<supportDesc> physical support/material

<objectDesc> example Ms 1609 (1)

```
<layoutDesc>  
  <layout columns="1">  
    <p><!-- prose description about columns and lines --></p>  
  </layout>  
</layoutDesc>
```

```
<supportDesc material="paper">  
  <support><p>paper</p></support>  
  <extent>  
    <measure type="leavesCount" unit="leaf">469</measure>  
    <dimensions type="block" unit="mm">  
      <height>140</height>  
      <width>110</width>  
    </dimensions>  
  </extent>  
<foliation>  
  <p><!-- prose description about folio counting and if any folios are missing --></p>  
</foliation>
```

<objectDesc> example Ms 1609 (2)

<collation>

```
<formula style="chroust">2.V + 5.VI + IX + 2.VI + V +IV + 2.VI + V + (VI+3) + IX+  
+ 3.VI + V + VI + V + (VI+1) + VII + VI + (VIII-1) + VIII-2) +  
2.VI + (VIII-1) + 4.VI + (VI+1) + V + VI  
(Kiewisch 1994, 118)</formula>
```

```
<catchwords><!-- when catchwords are used  
(help during the binding process) --></catchwords>
```

</collation>

<condition>

```
<desc> <!-- prose description --> </desc>
```

other possible elements:

```
<material ana="folio">Paper</material>
```

```
<material ana="binding">wood, leather, parchment</material>
```

```
<watermark xml:id="WaageKreis"> Waage in Kreis
```

```
<locusGrp><!-- place of the watermark -->
```

```
<locus from="f1" to="f50"></locus>
```

```
</locusGrp>
```

```
</watermark>
```

</condition>

</supportDesc>

About <bindingDesc>

```
<binding>  
  <condition><!-- prose description --></condition>  
  <decoNote><!-- prose description --></decoNote>  
</binding>
```

very detailed example with attributes:

```
<binding>  
  <condition><!-- prose description --></condition>  
  <decoNote type="bindingMaterial">  
    <material>Holzdeckel</material>  
    <material>Ledereinband</material>  
    <material>Pergamentrücken</material>  
  </decoNote>  
  <decoNote type="decoration">Blindpressungen</decoNote>  
  <decoNote type="clasps">eine MessingschlieÙe mit Lederband</decoNote>  
  <decoNote type="spine">Pergament</decoNote>  
  <decoNote type="added"></decoNote>  
  <decoNote type="wastepaper"><!-- prose description --></decoNote>  
</binding>
```

About <history>

<origin> Where it came from

<provenance> everything in between

<acquisition> how it got to its current place, or how you got it

```
<origin notAfter="1500" notBefore="1400">
  <origDate>1488</origDate>
  <origDate>1451</origDate>
  <origPlace cert="low">Kloster Tegernsee or Kloster am Mondsee</origPlace>
  <orgName role="scriptorium">Mondseer Koch- und Haushaltsbuch</orgName>
  <note type="purpose">Gebrauchshandschrift</note>
</origin>
<provenance>
  <p xml:id="prov1"> ..... </p>
  <p xml:id="prov2">.....</p>
  <p xml:id="prov3"> ..... </p>
</provenance>
<acquisition>
  <ab>University library Graz</ab>
  <signatures>seal of University library Graz
    <locus>page 1v</locus>
  </signatures>
</acquisition>
```

msDesc example so far

```
<msDesc>  
  <msIdentifier> </msIdentifier>  
  <head> </head>  
  <msContents> </msContents>  
  <physDesc>  
    <objectDesc>  
      <layoutDesc> </layoutDesc>  
      <supportDesc> </supportDesc>  
    </objectDesc>  
    <bindingDesc> </bindingDesc>  
  </physDesc>  
  <history> </history>  
</msDesc>
```

About `<msPart>`

Definition: contains information about an originally distinct manuscript or part of a manuscript, which is now part of a composite manuscript

```
<msDesc>  
  <msIdentifier></msIdentifier>  
  <head></head>  
  <msContents></msContents>  
  <physDesc></physDesc>  
  <history></history>  
  <additional></additional>  
  <msPart>  
    <msIdentifier></msIdentifier>  
  </msPart>  
  <msPart>  
    <msIdentifier></msIdentifier>  
  </msPart>  
  <msPart>  
    <msIdentifier></msIdentifier>  
  </msPart>  
</msDesc>
```

<msPart> Example

For composite Manuscripts

```
<msPart xml:id="_011r">
  <msIdentifier>
    <idno></idno> <!-- if the Part has its own Identifier-->
  </msIdentifier>
  <msContents>
    <msItem xml:id="ms16090011r">
      <locus ana="quire">Q2-Q8</locus>
      <locus ana="folio" corresp="#MS1609_0011r">11r-94v</locus>
      <title>Cooking recipes</title>
      <titlePart>Cooking recipes</titlePart>
      <textLang>Frühneuhochdeutsch/early new high german</textLang>
      <incipit corresp="#11r" xml:lang="deu-enh">Item wildw machen ein guet air von mandel...</incipit>
    </msItem>
    <msItem xml:id="ms16090095r">
      <locus ana="quire">Q9-Q11</locus>
      <locus ana="folio" corresp="#MS1609_0095r">95r-134v</locus>
      <title type="leer">empty</title>
      <titlePart>empty</titlePart>
    </msItem>
  </msContents>
</msPart>
```

About `<msFrag>`

Used when there is only a Fragment remaining of a manuscript or there is a Fragment within a Manuscript, added or used for the binding

```
<msFrag>  
  <msIdentifier>  
    <settlement>Sankt-Peterburg</settlement>  
    <repository>Rossiiskaia natsional'naia biblioteka</repository>  
    <idno>Q.p.I.72</idno>  
  </msIdentifier>  
</msFrag>
```

About the attributes used in the examples

- they are not necessary
- prose description is enough
- attributes in examples were an effort of a librarian in cooperation with a digital humanist.

Exercise

Your project is the transcription of a historic cookbook.

=> Prepare the TEI header as detailed as possible!

<https://bit.ly/2lLdtjC>