

TEI: Der teiHeader



Prof. Dr. Christof Schöch

Modul Auszeichnungssprachen
MSc. Digital Humanities, Universität Trier





Überblick

1. Erinnerung: Der minimale `teiHeader`
2. Erweiterungen des `teiHeaders`

(1) Erinnerung: Der minimale
teiHeader

Der `teiHeader`

- Neben `<text>` der zweite, zentrale Teil des TEI-Dokuments
- Zentrale Funktion: dokumentbezogene Metadaten festhalten
- Aber: Nicht zu verwechseln mit der Kodierung der Titelseite (!)
- Dokumentation: <https://tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>

Purposes of the `teiHeader`

This chapter addresses the problems of describing an encoded work so that the text itself, its source, its encoding, and its revisions are all thoroughly documented. Such documentation is equally necessary for scholars using the texts, for software processing them, and for cataloguers in libraries and archives. Together these descriptions and declarations provide an electronic analogue to the title page attached to a printed work. They also constitute an equivalent for the content of the code books or introductory manuals customarily accompanying electronic data sets. ("The TEI Header")

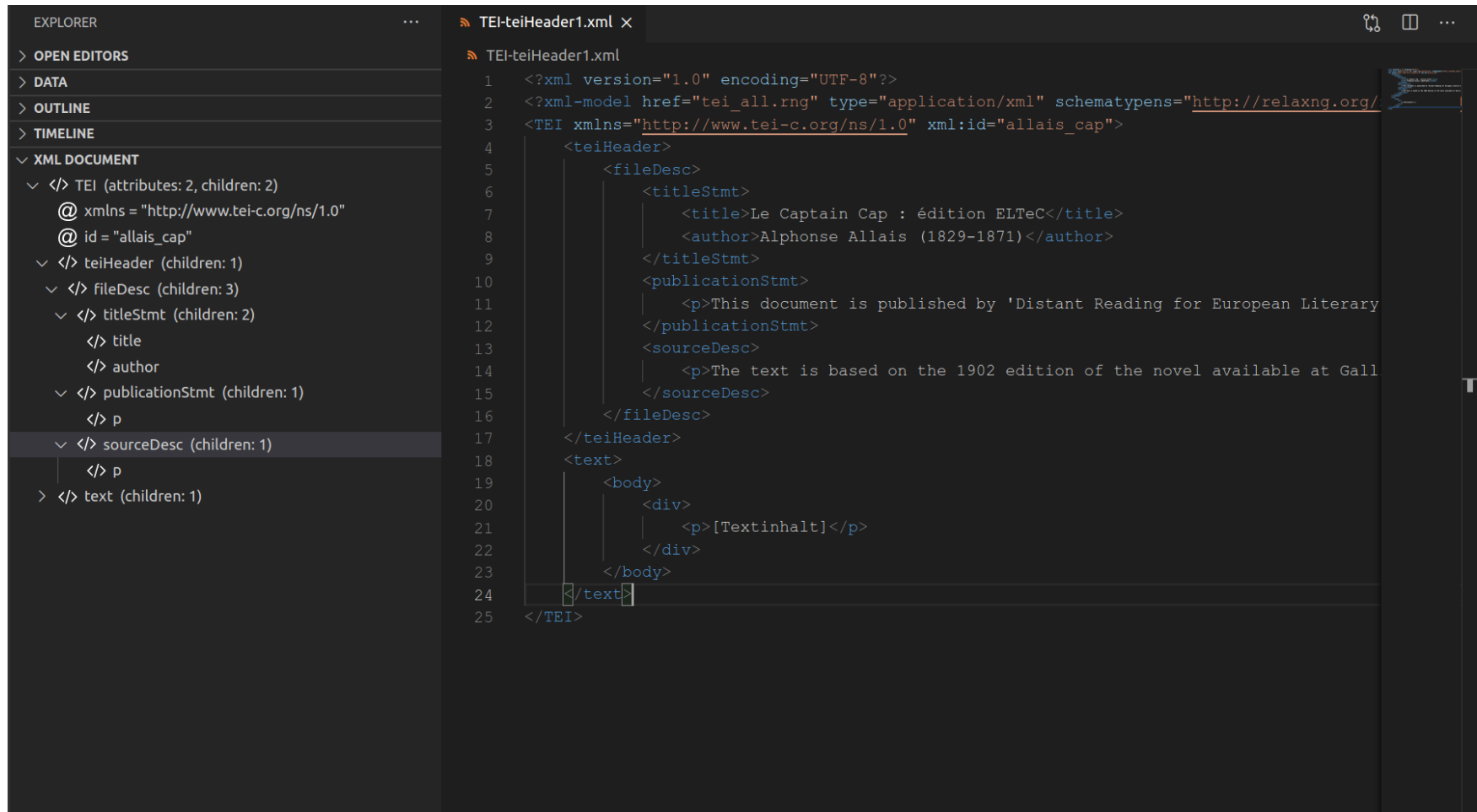
Komponenten des `teiHeaders`

- `<fileDesc>`: verpflichtend; enthält die wesentlichen Metadaten zum Dokument selbst
- `<encodingDesc>`: dokumentiert die Prinzipien der Textkodierung
- `<profileDesc>`: enthält Informationen, die den Text beschreiben und kontextualisieren
- `<revisionDesc>`: dokumentiert die Versionsgeschichte des Dokuments
- `<xenodata>`: Metadaten, die in einem anderen Format als TEI vorgehalten werden

Obligatorische Teile der

- `<titleStmt>`
 - Informationen zum Titel
 - `<title>` Titel des digitalen Dokuments (nicht der Quelle)
- `<publicationStmt>`
 - Aussage zum Status des Dokuments bzgl. Publikation
 - Mindestens ein unstrukturierter Absatz `<p>`
- `<sourceDesc>`
 - Beschreibung der Quelle, die dem digitalen Dokument zugrunde liegt
 - Mindestens ein unstrukturierter Absatz `<p>`

Ein minimaler teiHeader (1)



The screenshot shows an IDE with two panels. The left panel, titled 'EXPLORER', displays the XML document's structure as a tree. The right panel shows the XML code for 'TEI-teiHeader1.xml'.

EXPLORER Panel:

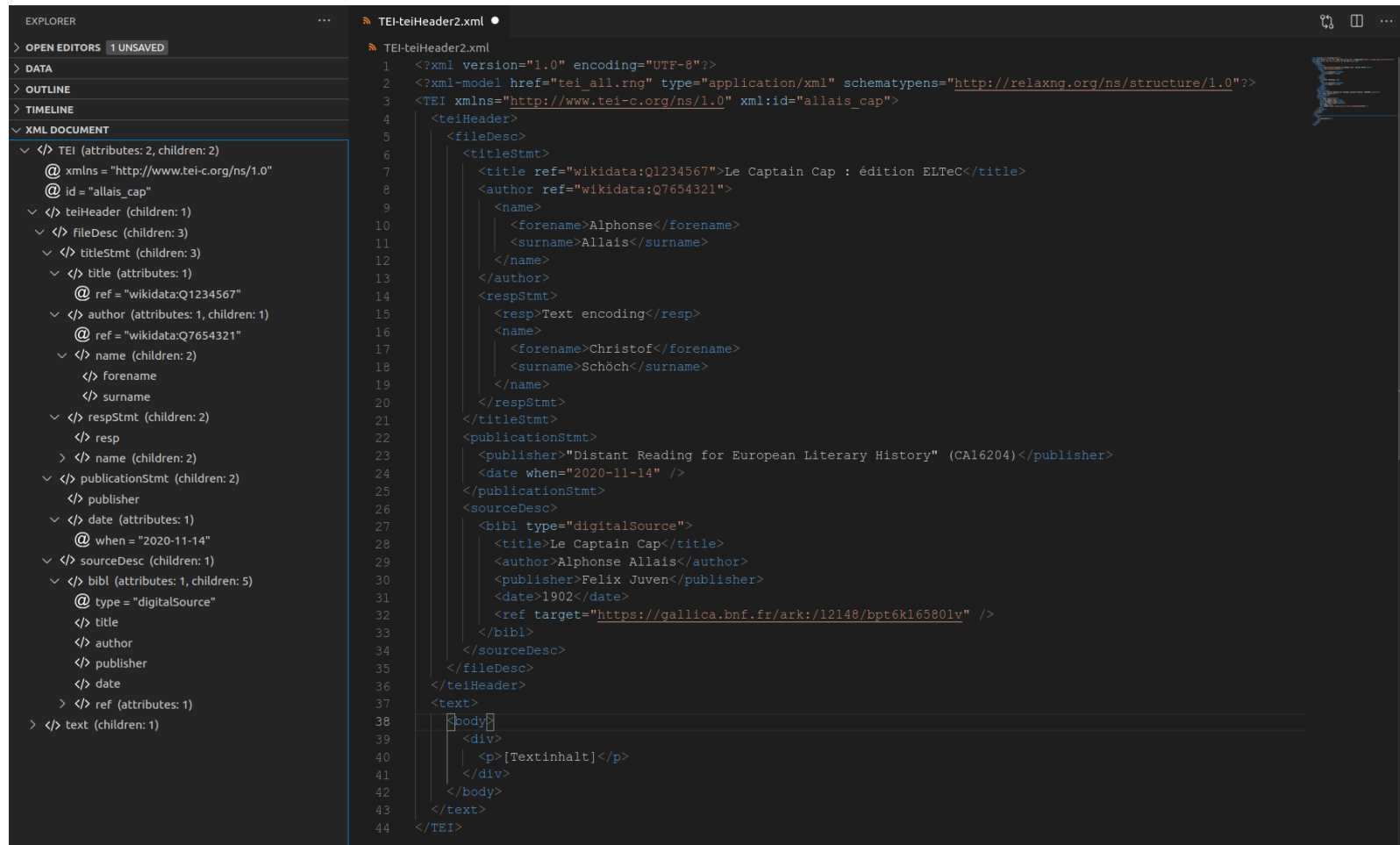
- > OPEN EDITORS
- > DATA
- > OUTLINE
- > TIMELINE
- ✓ XML DOCUMENT
 - ✓ </> TEI (attributes: 2, children: 2)
 - @ xmlns = "http://www.tei-c.org/ns/1.0"
 - @ id = "allais_cap"
 - ✓ </> teiHeader (children: 1)
 - ✓ </> fileDesc (children: 3)
 - ✓ </> titleStmt (children: 2)
 - </> title
 - </> author
 - ✓ </> publicationStmt (children: 1)
 - </> p
 - ✓ </> sourceDesc (children: 1)
 - </> p
 - > </> text (children: 1)

TEI-teiHeader1.xml Panel:

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="tei_all.rng" type="application/xml" schematypens="http://relaxng.org/
3 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="allais_cap">
4   <teiHeader>
5     <fileDesc>
6       <titleStmt>
7         <title>Le Capitain Cap : édition ELTeC</title>
8         <author>Alphonse Allais (1829-1871)</author>
9       </titleStmt>
10      <publicationStmt>
11        <p>This document is published by 'Distant Reading for European Literary
12      </publicationStmt>
13      <sourceDesc>
14        <p>The text is based on the 1902 edition of the novel available at Gall
15      </sourceDesc>
16    </fileDesc>
17  </teiHeader>
18  <text>
19    <body>
20      <div>
21        <p>[Textinhalt]</p>
22      </div>
23    </body>
24  </text>
25</TEI>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader1.xml>

Ein minimaler teiHeader (2)



The screenshot displays a code editor with two panels. The left panel shows the XML DOM tree for 'TEI-teiHeader2.xml'. The right panel shows the raw XML code.

XML DOM Tree (Left Panel):

- TEI (attributes: 2, children: 2)
 - @ xmlns = "http://www.tei-c.org/ns/1.0"
 - @ id = "allais_cap"
 - teiHeader (children: 1)
 - fileDesc (children: 3)
 - titleStmt (children: 3)
 - title (attributes: 1)
 - @ ref = "wikidata:Q1234567"
 - author (attributes: 1, children: 1)
 - @ ref = "wikidata:Q7654321"
 - name (children: 2)
 - forename
 - surname
 - respStmt (children: 2)
 - resp
 - name (children: 2)
 - forename
 - surname
 - publicationStmt (children: 2)
 - publisher
 - date (attributes: 1)
 - @ when = "2020-11-14"
 - sourceDesc (children: 1)
 - bibl (attributes: 1, children: 5)
 - @ type = "digitalSource"
 - title
 - author
 - publisher
 - date
 - text (children: 1)
 - ref (attributes: 1)

Raw XML Code (Right Panel):

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <?xml-model href="tei_all.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
3 <TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="allais_cap">
4   <teiHeader>
5     <fileDesc>
6       <titleStmt>
7         <title ref="wikidata:Q1234567">Le Captain Cap : édition ELTeC</title>
8         <author ref="wikidata:Q7654321">
9           <name>
10             <forename>Alphonse</forename>
11             <surname>Allais</surname>
12           </name>
13         </author>
14         <respStmt>
15           <resp>Text encoding</resp>
16           <name>
17             <forename>Christof</forename>
18             <surname>Schöch</surname>
19           </name>
20         </respStmt>
21       </titleStmt>
22       <publicationStmt>
23         <publisher>"Distant Reading for European Literary History" (CA16204)</publisher>
24         <date when="2020-11-14" />
25       </publicationStmt>
26       <sourceDesc>
27         <bibl type="digitalSource">
28           <title>Le Captain Cap</title>
29           <author>Alphonse Allais</author>
30           <publisher>Felix Juven</publisher>
31           <date>1902</date>
32           <ref target="https://gallica.bnf.fr/ark:/12148/bpt6k165801v" />
33         </bibl>
34       </sourceDesc>
35     </fileDesc>
36   </teiHeader>
37   <text>
38     <body>
39       <div>
40         <p>[Textinhalt]</p>
41       </div>
42     </body>
43   </text>
44 </TEI>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader2.xml>

(2) Erweiterungen der <fileDesc>

Erweitertes <titleStmt>

- Standard-Rollen

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

- Für spezifischere Rollen

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

Beispiel <titleStmt>

```
TEI-teiHeader3.xml
5      <fileDesc>
6          <titleStmt>
7              <title ref="viaf:123456789">Le Captain Cap : édition ELTeC</
              title>
8              <author ref="viaf:302147306">
9                  <persName>
10                     <forename>Alphone</forename>
11                     <surname>Allais</surname>
12                 </persName>
13             </author>
14             <editor>
15                 <persName>
16                     <forename>Christof</forename>
17                     <surname>Schöch</surname>
18                 </persName>
19             </editor>
20             <respStmt>
21                 <resp>Encoding</resp>
22                 <persName>
23                     <forename>Pia</forename>
24                     <surname>Geißel</surname>
25                 </persName>
26             </respStmt>
27         </titleStmt>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader3.xml>

Erweiterte `<sourceDesc>`

- Drei Ansätze
 - Unstrukturiert: freie Prosa in einem `<p>`
 - Leicht strukturiert: Prosa mit einzelnen bibliographischen Elementen: `<author>`, `<publisher>` in `<bibl>`
 - Stark strukturiert mit `<biblStruct>`: nur spezielle Elemente
- Außerdem spezielle Elemente
 - für gesprochene Sprache: `<recordingStmt>`
 - für Manuskripte: `<msDesc>`
 - Referenz-Listen: `<listPerson>`, `<listPlace>`, `<listOrg>`

Beispiel <sourceDesc>

```
TEI-teiHeader3.xml
74      </bibl>
75      <listPerson>
76        <person>
77          <persName>Capitain Cap</persName>
78        </person>
79        <person>
80          <persName>Capitain Cap</persName>
81        </person>
82        <person>
83          <persName>Paul Fabre</persName>
84        </person>
85      </listPerson>
86    </sourceDesc>
87  </fileDesc>
88  <encodingDesc n="eltec-1">
89    <projectDesc>
90      <p>The ELTeC project aims to provide sets of 100 novels encoded
91      in XML-TEI for at least 10 European languages.</p>
92    </projectDesc>
93    <samplingDecl>
94      <p>The texts were selected according to the corpus composition
95      principles of ELTeC.</p>
96    </samplingDecl>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader3.xml>

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 (@unit: kb, GB, Wörter, Seiten)
- <notesStmt> - Anmerkungen zu Aspekten, die anderswo nicht abgedeckt sind

Beispiel <fileDesc>

```
TEI-teiHeader3.xml
27      </titleStmt>
28      <editionStmt>
29      |   <edition>Version 1.0, November 2020</edition>
30      </editionStmt>
31      <extent>
32      |   <measure unit="words">41424</measure>
33      |   <measure unit="pages">124</measure>
34      </extent>
35      <publicationStmt>
36      |   <publisher>COST Action "Distant Reading for European Literary
37      |   History" (CA16204)</publisher>
38      |   <distributor>Zenodo.org</distributor>
39      |   <date when="2020-11-14" />
40      |   <idno type="doi">10.5281/zenodo.4271445</idno>
41      |   <availability>
42      |   |   <p>Creative Commons Attribution 4.0 International</p>
43      |   </availability>
44      </publicationStmt>
45      <seriesStmt>
46      |   <title>European Literary Text Collection</title>
47      |   <idno type="doi">10.5281/zenodo.4271467</idno>
48      |   <respStmt>
49      |   |   <resp>Editors</resp>
50      |   |   <persName>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader3.xml>

(3) Die `<encodingDesc>`

<encodingDesc>

- Enthält Informationen zum Kodierungs-Vorgang
- Als Prosa in einem <p>
- Oder mit speziellen Elementen, u.a.
 - <projectDesc>: Ziele des Projekts
 - <samplingDecl>: Prinzipien der Auswahl der Texte
 - <editorialDecl>: Editionsprinzipien, bspw. <normalization> oder <hyphenation>.

Beispiel <encodingDesc>

```
TEI-teiHeader3.xml
88      <encodingDesc n="eltec-1">
89          <projectDesc>
90              <p>The ELTeC project aims to provide sets of 100 novels encoded
91                  in XML-TEI for at least 10 European languages.</p>
92          </projectDesc>
93          <samplingDecl>
94              <p>The texts were selected according to the corpus composition
95                  principles of ELTeC.</p>
96          </samplingDecl>
97          <editorialDecl>
98              <p>The encoding of this resource follows the schema defined as
99                  ELTeC level 1 encoding.</p>
100              <correction>
101                  <p>Obvious errors in spelling were corrected and marked
102                      with<gi>corr</gi>.</p>
103              </correction>
104              <normalization>
105                  <p>Accentuation was tacitly normalized.</p>
106              </normalization>
107              <hyphenation>
108                  <p>In-paragraph line breaks have not been preserved and
109                      hyphens tacitly removed.</p>
110              </hyphenation>
111          </editorialDecl>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader3.xml>

(4) Die `<profileDesc>`

Wesentliche Elemente

- `<creation>` - Informationen über den Ursprung des Texts, bspw. Zeit und Ort
- `<langUsage>` - Informationen über die Sprachen, Register, Schriftsysteme, im Text
- `<textDesc>` und `<textClass>` - Mechanismen zur Klassifikation des Textes oder der Texte
- `<handNotes>` – Informationen über die Schriften ("hands") in einem Manuskript

Beispiel <profileDesc>

```
TEI-teiHeader3.xml
106     </editorialDecl>
107   </encodingDesc>
108   <profileDesc>
109     <langUsage>
110       <language ident="fr">French</language>
111     </langUsage>
112     <textClass>
113       <keywords scheme="literary-ontology">
114         <term type="narrative-perspective">homodiegetic</term>
115         <term type="setting">urban</term>
116         <term type="time">1890s</term>
117       </keywords>
118     </textClass>
119     <textDesc>
120       <channel mode="w">Written</channel>
121       <constitution></constitution>
122       <derivation></derivation>
123       <domain>Novel</domain>
124       <factuality>Fictional</factuality>
125       <interaction></interaction>
126       <preparedness></preparedness>
127       <purpose>Entertainment</purpose>
128     </textDesc>
129   </profileDesc>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader3.xml>

(5) Die `<revisionDesc>`

Wesentliche Elemente

- Zur Dokumentation von Überarbeitungen des TEI-Dokuments
- neuester Eintrag ganz oben
- Struktur
 - Liste von `<change>` Elementen
 - Attribute: `@date` und `@who`
- Vorsicht
 - `<change>` hier: bezieht sich auf das TEI-Dokument
 - `<change>` in `//sourceDesc/creation` bezieht sich auf die kodierte Quelle

Beispiel <revisionDesc>

```
TEI-teiHeader3.xml
128     </textDesc>
129   </profileDesc>
130   <revisionDesc>
131     <change when="2020-12-18" who="#cs">Adapted teiHeader for demo
132     purposes</change>
133     <change when="2020-11-14" who="#lb">Converted by checkUp script for
134     new release</change>
135     <change when="2020-11-14" who="#lb">Checked by checkup script</
136     change>
137     <change when="2019-11-04" who="#lb">Checked by releaseChecker
138     script</change>
139     <change when="2019-06-15" who="#cs">Conversion to ELTeC level-1)</
140     change>
141   </revisionDesc>
142 </teiHeader>
143 <text>
144   <body>
145     <div>
146       <p>[Textinhalt]</p>
147     </div>
148   </body>
149 </text>
150 </TEI>
```

<https://github.com/dh-trier/markup/blob/main/data/TEI-teiHeader3.xml>

Abschluss

Lektürehinweise

Referenzlektüre

- "The TEI Header", in: *Guidelines of the Text Encoding Initiative*, P5, Version 4.1.0, 2020. <https://tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>

Danke!

Lizenz: [Creative Commons Attribution \(CC BY\)](#), 2020.
