

Summer School "Digitale Edition"

Erschließung geisteswissenschaftlicher Quellen
mit digitalen Methoden

5.-9. September 2016

Zentrum für Informationsmodellierung
Austrian Centre for Digital Humanities
Elisabethstraße 59/III, SR 81.31

Programm

Dienstag 6.9.

- 9:00-10:30 *TEI-Vertiefung 1 – Metadaten, Text und Transkription*
Roman Bleier
- 11:00-12:30 *TEI-Vertiefung 2 – <text> und Transkription; Text und Bild*
Wout Dillen
- 14:30-16:00 *Textnavigation mit XPath*
Patrick Sahle
- 16:30-18:00 *TEI-Vertiefung 2 – Fortsetzung*
Wout Dillen
- 18:30 *Stadtführung*
Treffpunkt: Universitätsplatz 1 (vor Hauptgebäude der Uni Graz); Dauer: ca. 1 Stunde

TEI-Vertiefung 1

<teiHeader> Bereich
Metadaten

<teiHeader> Übung

<text> Bereich

<text> Übung

Einfaches TEI Dokument

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <TEI version="5.0" xmlns="http://www.tei-c.org/ns/1.0">
3    <teiHeader>
4      <fileDesc>
5        <titleStmt>
6          <title>The shortest TEI Document Imaginable</title>
7        </titleStmt>
8        <publicationStmt>
9          <p>First published as part of TEI P2, this is the P5
10           version using a name space.</p>
11        </publicationStmt>
12        <sourceDesc>
13          <p>No source: this is an original work.</p>
14        </sourceDesc>
15      </fileDesc>
16    </teiHeader>
17    <text>
18      <body>
19        <p>This is about the shortest TEI document imaginable.</p>
20      </body>
21    </text>
22  </TEI>
```

Example from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-TEI.html>

<teiHeader>

‚Container element‘ für Metadaten

Metadaten:

- Strukturierte Daten über Daten
- Standardisierte
- Maschinenlesbar

Metadaten

Alice's adventures in Wonderland / by Lewis Carroll ; with forty-two illustrations by John Tenniel. / by Lewis Carroll ; with forty-two illustrations by John Tenniel

Additional actions:



Carroll, Lewis, 1832-1898.

Printed Book | Macmillan | 1888 | People's edition.

Available at Early Printed Books Stacks (use call slip) (OLS POL 741)

[Bookmark the link to this record](#)

Items

Location	Shelfmark	Status
Early Printed Books Stacks (use call slip)	OLS POL 741	IN

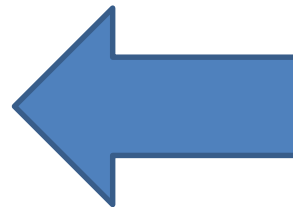
Details

Edition People's edition.

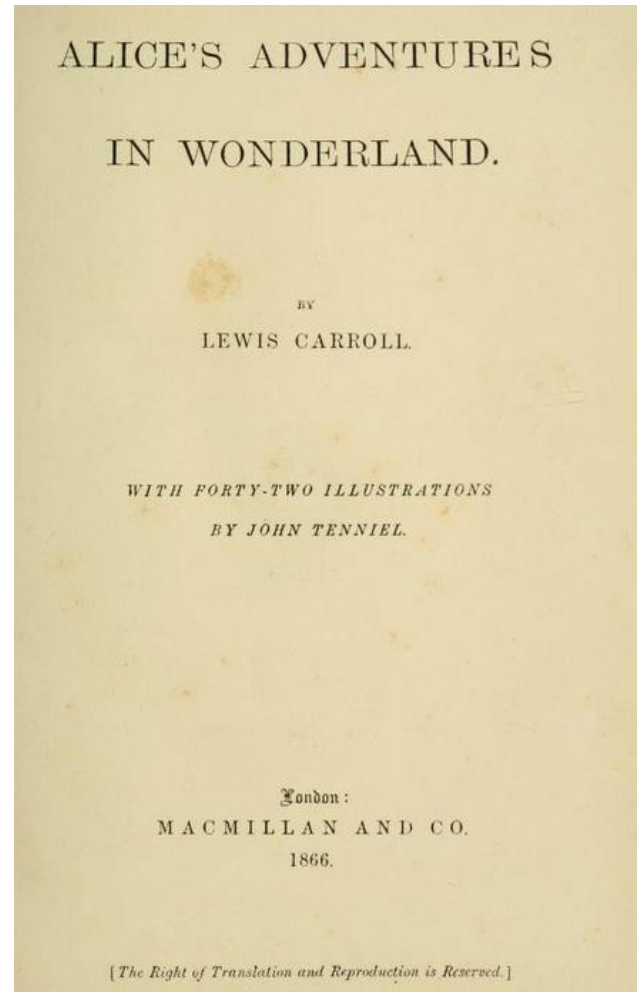
Description [12], 174, [14], 208, [4]



Struktur



Standards



Im (Online-) Bibliothekskatalog

<https://commons.wikimedia.org/wiki/File:Schlagwortkatalog.jpg>

© [GNU-Lizenz für freie Dokumentation](#), Version 1.2 oder einer späteren Version

<https://archive.org/details/alicesadventur00carr>

© Internet Archive (<https://archive.org/>)

Wozu Metadaten?

Beschreibung

Erweiterte Suchanfragen

Identifikation

Vernetzung

Darstellung

Metadaten in HTML

```
<meta charset="UTF-8"/>
<meta name="keywords" content="your, tags"/>
<meta name="description" content="150 words"/>
<meta name="subject" content="your website's subject"/>
<meta name="copyright" content="company name"/>
<meta name="language" content="ES"/>
<meta name="robots" content="index, follow"/>
<meta name="revised" content="Sunday, July 18th, 2010, 5:15 pm"/>
<meta name="abstract" content=""/>
<meta name="topic" content=""/>
<meta name="summary" content=""/>
<meta name="Classification" content="Business"/>
<meta name="author" content="name, email@hotmail.com"/>
<meta name="designer" content=""/>
<meta name="reply-to" content="email@hotmail.com"/>
<meta name="owner" content=""/>
<meta name="url" content="http://www.websiteaddress.com"/>
<meta name="identifier-URL" content="http://www.websiteaddress.com"/>
<meta name='DC.title' content='Unstoppable Robot Ninja'/>
```

<https://gist.github.com/lancejpollard/1978404>

© lancejpollard

Metadaten im Internet



Alice's adventures in wonderland

by [Carroll, Lewis, 1832-1898](#)

SHOW MORE

With forty-two illustrations by John Tenniel; bound in red cloth; stamped in gold; all edges gilt; blue coated endpapers; hinges repaired

Publisher [Macmillan and Co.](#)

Pages [232](#)

Language [English](#)

Digitizing sponsor [Lyrisis Members and Sloan Foundation](#)

Book contributor [Gettysburg College, Musselman Library, Special Collections](#)

Collection [gettysburgcollege](#); [americana](#)

Beispiel Internetarchive

<https://archive.org/details/alicesadventur00carr>

© Internet Archive (<https://archive.org/>)

Metadaten im Internet

The Internet Archive API:

<https://archive.org/details/alicesadventur00carr>

<https://archive.org/metadata/alicesadventur00carr/metadata>

<https://archive.org/metadata/alicesadventur00carr/metadata/creator>

Metadaten Standards

MARC MACHine-Readable Cataloging

Dublin Core

MODS Metadata Object Description Schema

METS Metadata Encoding and Transmission Standard

TEI <teiHeader>
 <msDesc>

Struktur des <teiHeader>

<fileDesc>	Dateibeschreibung
<encodingDesc>	Beschreibung der Kodierung
<profileDesc>	Beschreibung des Textprofils
<revisionDesc>	Beschreibung der Textrevision

<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 type="main">The Confession of St Patrick</title>
  <title type="sub">A machine-readable edition</title>
  <title xml:lang="de">Das Bekenntnis des Heiligen Patrick</title>
  <author>Patrick of Armagh</author>
  <funder>Volkswagen Stiftung</funder>
  <principal>Roman Bleier</principal>
  <respStmt>
    <name>Conor Neilan</name>
    <resp>data entry and proof correction</resp>
  </respStmt>
</titleStmt>
```

<publicationStmt>

umfasst Angaben zu Veröffentlichung
oder Vertrieb des TEI Dokuments

```
<publicationStmt>
  <publisher>
    Centre for Information Modelling - Austrian Centre for Digital Humanities, University of Graz</publisher>
  <pubPlace>Graz</pubPlace>
  <date when="2011-08-04">19 August 2011</date>
  <idno type="PID">o:epistles.1535</idno>
  <availability status="free">
    <p>Copyright 2011. This text is freely available provided the text is distributed
      with the header information provided. Full copyright statement: <ref
        target="http://creativecommons.org/licenses/by/4.0/">Creative
        Commons 4.0</ref>.</p>
  </availability>
</publicationStmt>
```

<sourceDesc>

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

```
<sourceDesc>
  <bibl>
    <editor>
      <surname>Bieler</surname>, <forename>Ludwig</forename>
    </editor>
    <title level="m">Clavis Patricii II: Libri Epistolarum Sancti Patricii Episcopi</title>
    <date>1993</date>
    <pubPlace>Dublin</pubPlace>
  </bibl>
</sourceDesc>
```


<sourceDesc>

```
<sourceDesc>
  <msDesc>
    <msIdentifier>
      <settlement>Oxford</settlement>
      <repository>Bodleian Library</repository>
      <idno type="Bod">MS Poet. Rawl. D. 169.</idno>
    </msIdentifier>
    <msContents>
      <msItem>
        <author>Geoffrey Chaucer</author>
        <title>The Canterbury Tales</title>
      </msItem>
    </msContents>
    <physDesc>
      <objectDesc>
        <p>A parchment codex of 136 folios, measuring approx
          28 by 19 inches, and containing 24 quires.</p>
        <p>The pages are margined and ruled throughout.</p>
        <p>Four hands have been identified in the manuscript: the first 44
          folios being written in two cursive anglicana scripts, while the
          remainder is for the most part in a mixed secretary hand.</p>
      </objectDesc>
    </physDesc>
  </msDesc>
</sourceDesc>
```

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

Struktur des <teiHeader>

<fileDesc> Dateibeschreibung

**<encodingDesc> Beschreibung der
Kodierung**

<profileDesc> Beschreibung des
Textprofils

<revisionDesc> Beschreibung der
Textrevision

<encodingDesc>

dokumentiert das Verhältnis zwischen elektro. 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>
```

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

<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>
</encodingDesc>
```

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-editorialDecl.html>

Struktur des <teiHeader>

<fileDesc>	Dateibeschreibung
<encodingDesc>	Beschreibung der Kodierung
<profileDesc>	Beschreibung des Textprofils
<revisionDesc>	Beschreibung der Textrevision

<profileDesc>

Inhaltliche Informationen über einen Text: z.B.

Sprachen und Subsprachen,
Entstehungszeit,
Entstehungsbedingungen, Teilnehmer

<profileDesc>

```
<profileDesc>
  <creation>The letter was written in <date when="1916">1916</date> in
    <name type="city">Dublin</name>.
  </creation>
  <langUsage>
    <language ident="en-GB" usage="70">British English</language>
    <language ident="ga-GA" usage="30">Irish</language>
  </langUsage>
  <handNotes>
    <handNote xml:id="rm">Robert Monteith</handNote>
    <handNote xml:id="sm">Stanislaus Murphy</handNote>
  </handNotes>
  <!-- correspDesc -->
</profileDesc>
```

Struktur des <teiHeader>

<fileDesc>	Dateibeschreibung
<encodingDesc>	Beschreibung der Kodierung
<profileDesc>	Beschreibung des Textprofils
<revisionDesc>	Beschreibung der Textrevision

<revisionDesc>

enthält alle Revisionschritte, die an der Datei vorgenommen wurden

```
<revisionDesc>  
  <listChange>  
    <change when="1991-11-11" who="#LB"> deleted chapter 10 </change>  
    <change when="1991-11-02" who="#MSM"> completed first draft </change>  
  </listChange>  
</revisionDesc>
```

Übung <teiHeader>

Ergänze den <teiHeader> im XML

Ersetze die Kommentare <!-- ??? -->
mit Daten aus dem Word Dokument

Übungsdateien:

My_name_is_Patrick.docx

Bleier_Übung.xml

Einfaches TEI Dokument

```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI version="5.0" xmlns="http://www.tei-c.org/ns/1.0">
3   <teiHeader>
4     <fileDesc>
5       <titleStmt>
6         <title>The shortest TEI Document Imaginable</title>
7       </titleStmt>
8       <publicationStmt>
9         <p>First published as part of TEI P2, this is the P5
10          version using a name space.</p>
11       </publicationStmt>
12       <sourceDesc>
13         <p>No source: this is an original work.</p>
14       </sourceDesc>
15     </fileDesc>
16   </teiHeader>
17   <text>
18     <body>
19       <p>This is about the shortest TEI document imaginable.</p>
20     </body>
21   </text>
22 </TEI>
```

Example from: <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-TEI.html>

TEI <text>

enthält eigenständigen oder aus mehreren Teilen bestehenden Text

Kindelemente:

<front>

<body>

<back>

<group>

TEI <text>

```
<text>
  <front>
    <!-- front matter of copy text, if any, goes here -->
  </front>
  <body>
    <!-- body of copy text goes here -->
  </body>
  <back>
    <!-- back matter of copy text, if any, goes here -->
  </back>
</text>
```

TEI <front>

enthält alle dem Kerntext vorangestellten Texte (Titelseite, Vorworte, Widmungen, usw.)

```
<front>
  <docTitle>
    <titlePart> The Adventures of Sherlock Holmes
  </titlePart>
</docTitle>
<docImprint>First published in <title>The Strand</title>
  between July 1891 and December 1892</docImprint>
  <!-- any other front matter specific to this collection -->
</front>
```

<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/examples-group.html>

TEI <body>

enthält den gesamten, eigenständigen Text, außer Vorspann (front) und Nachspann (back)

```
<text>
  <body>
    <p>
      <!-- Text goes here -->
    </p>
  </body>
</text>
```

<div>: Struktur im <body>

<body>

```
<div type="Kapitel" n="1">  
  <div type="Unterkapitel" n="1">  
    <!-- Hier kommt der Text des ersten Unterkapitels im ersten Kapitel -->  
  </div>  
  <div type="Unterkapitel" n="2">  
    <!-- Hier kommt der Text des zweiten Unterkapitels im ersten Kapitel -->  
  </div>  
</div>  
<div type="Kapitel" n="2">  
  <div n="1" type="Unterkapitel">  
    <!-- Hier kommt der Text des ersten Unterkapitels im zweiten Kapitel -->  
  </div>  
  <div n="2" type="Unterkapitel">  
    <!-- Hier kommt der Text des zweiten Unterkapitels im zweiten Kapitel -->  
  </div>  
</div>
```

</body>

TEI <back>

enthält Anhänge jeglicher Art, die auf den Hauptteil eines Textes folgen

```
<back>
  <div type="appendix">
    <head>Bibliography</head>
    <listBibl>
      <bibl>Bieler, Ludwig, The Works of St. Patrick,
        in Ancient Christian Writers (Westminster, Maryland/ London, 1953).</bibl>
      <bibl>Bieler, Ludwig, "Irish Manuscripts in Medieval Germania",
        Irish Ecclesiastical Record, 87, 5th ser. (1957), 161-169.</bibl>
      <bibl> ... </bibl>
    </listBibl>
  </div>
</back>
```

Some core module elements

```
<div>
  <head>St Patrick's Confessio</head>
  <div n="1" type="chapter">
    <head>St Patrick is kidnapped</head>
    <p>...We would not listen to our priests,
    who advised us about how we could be
    saved<note n="1" place="bottom">
      <ref>Daniel 9:5-6</ref>: <quote>We have
        gone aside from your commandments
        ... we have not listened to your servants
        the prophets.</quote></note>. The Lord
        brought his strong anger upon us,...</p>
  </div>
</div>
```

Übung <text>

Ergänze den <text> Bereich im TEI Dokument.

Jeder von St Patrick's Texten soll in einem eigenen <text> kodiert werden.

Übungsdateien:

My_name_is_Patrick.docx

Bleier_Übung.xml

TEI to PDF, HTML, etc.

oXygen hat TEI-Default-Stylesheets

Wandelt TEI zu PDF, HTML, etc.

Übung:

Transformiere Bleier_Übung.xml zu HTML mit den

Fragen, Anregungen?

Roman Bleier

roman.bleier@uni-graz.at

Centre for Information Modelling

Austrian Centre for Digital Humanities

University of Graz



Dieses Werk ist lizenziert unter einer [Creative Commons Namensnennung 4.0 International Lizenz](https://creativecommons.org/licenses/by/4.0/).



Alle darin verwendeten Werke anderer Urheber sind Zitate zu wissenschaftlichem Gebrauch.