

# Referat 04: Die Guidelines von 1 bis 23

**DARIAH-DE Tutorials *Digitale Textkodierung mit TEI***

Redaktion: Christof Schöch (Univ. Würzburg)

Version 1.0, 02/2014

Grundlage der Folien: DH@Oxford 2012

Lizenz: [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) (CC-BY)



# Eckdaten zu den Guidelines

## Ein paar Daten

- Umfang: 1600 Seiten
- Vokabular: 521 Elemente
- Organisation: 146 Klassen und 22 Module
- Formalisierte Regeln (in Relax NG): 7185 Zeilen Code

## Außerdem:

- Sehr viel erklärende Prosa
- Mindestens ein Beispiel für jedes Element

# Aufbau der Guidelines (1)

1. The TEI Infrastructure
2. The TEI Header
3. Elements Available in All TEI Documents
4. Default Text Structure
5. Representation of Non-standard Characters and Glyphs
6. Verse
7. Performance Texts
8. Transcriptions of Speech
9. Dictionaries
10. Manuscript Description
11. Representation of Primary Sources
12. Critical Apparatus

# Aufbau der Guidelines (2)

13. Names, Dates, People, and Places
14. Tables, Formulæ, and Graphics
15. Language Corpora
16. Linking, Segmentation, and Alignment
17. Simple Analytic Mechanisms
18. Feature Structures
19. Graphs, Networks, and Trees
20. Non-hierarchical Structures
21. Certainty and Responsibility
22. Documentation Elements
23. Using the TEI

# Kapitel mit grundlegenden Informationen

## v. A Gentle Introduction to XML

- Fasst die Grundlagen von XML zusammen, die zur Nutzung der Guidelines notwendig sind. (Pflichtlektüre!)
- Prinzipien des Markup, seine Repräsentation in XML, die Möglichkeiten der Weiterverarbeitung

## 1. The TEI Infrastructure

- Definition und Erklärungen der Funktionsweise von Klassen und globalen Attributen
- Präsentiert die “Informations-Architektur” der TEI
- Relevant, um zu verstehen, wie der TEI-Standard genau funktioniert

# Basis-Kapitel (cf. Input 02)

Kapitel 2: der TEI Header

Kapitel 3: "Elements available in all documents" ("Core")

Kapitel 4: "Default Text Structure"

# 5. Nicht-Standard Zeichen und Glyphen

## Unicode

- beinhaltet fast 100.000 Zeichen
- fast alles ist dabei
- Verweis auf Unicode: `&x[hexadez. Wert];`  
`&xU+0119;`

## Wenn nicht:

- `<g>` (glyph) – erlaubt die Kodierung von Glyphen und Zeichen, die nicht in Unicode sind
- `<charDecl>` mit `<char>` und `<glyph>` - erlaubt es, die Glyphen und besondere Zeichen zu definieren bzw. zu beschreiben
- Das Attribut `@ref` in `<g>` verweist auf `<charDecl>`

## 5 – Beispiel für <g> (gaiji)

### Im Text

Letters in this manuscript are sometimes written in a really `<g ref="#curlyF">f</g>`unny way.

### In der <encodingDesc> (im Header):

```
<charDecl>
  <glyph xml:id="curlyF">
    <glyphName>LATIN SMALL F WITH CURLY ENDS</glyphName>
    <desc>Small latin letter f with curly ends.</desc>
    <mapping type="equivalent">f</mapping>
    <figure>
      <graphic url="chars/curlyF.png"/>
    </figure>
  </glyph>
</charDecl>
```



# 6 – Verse

In “Core” gibt es bereits:

- `<lg>` (line-groups) - Strophen
- `<l>` (line) - Zeilen

Dieses Modul fügt hinzu

- `<rhyme>` (Reim), `<caesura>` (Zäsur)
- `<metDecl>` und `<metSym>` (Deklaration des metrischen Systems)
- Attribute `met` und `real`, um diese zu markieren

# 6 – Beispiel: Verse

```
<lg type="quatrain" rhyme="aabccb">
  <l>Les étoiles, points d'or, percent les branches
n<rhyme label="a">oires</rhyme> ; </l>
  <l>Le flot huileux et lourd décompose ses
m<rhyme label="a">oires</rhyme></l>
  <l>Sur l'océan blê<rhyme label="b">mi</rhyme> ;</l>
  <l>Les nuages ont l'air d'oiseaux prenant la
f<rhyme label="c">uite</rhyme>,</l>
  <l>Par moments le vent parle, et dit des mots sans
s<rhyme label="c">uite</rhyme> </l>
  <l>Comme un homme endor<rhyme label="b">mi</rhyme>.</l>
</lg>
```

# 7 – Performance Texts

## Im Modul “Core”

- `<sp>` (speech), `<stage>` (Bühnenanweisungen), `<speaker>`

## Ce module rajoute :

- `<castList>`, `<castItem>`, `<castGroup>`, `<actor>`, `<role>`,  
`<roleDesc>`
- `<camera>`, `<caption>`, `<move>`, `<performance>`,
- `<prologue>`, `<epilogue>`,
- `<set>`, `<sound>`, `<tech>`, `<view>`

## 7 – Beispiel: <performance>

```
<performance>
  <p>
    <name type="theatre">Théâtre national du Petit-Opéra,
    Paris</name>, <date>du 10 Décembre 1974 au 12 Janvier
    1975</date>.
    <castList>
      <castItem>
        <role>Paul Valéry</role>
        <actor>Michel Duchaussoy</actor>
      </castItem>
      <castItem>
        <role>Joseph</role>
        <actor>Gérard Caillaud</actor>
      </castItem>
      <!-- ... -->
    </castList></p>
  <p>Adaptation et mise en scène de <name>Pierre
  Franck</name></p>
</performance>
```

# 8 – Transcription of Speech

Dies ist ein Modul speziell für die Transkription von gesprochener Sprache

Zusätzliche Elemente für den Header

- `<broadcast>`, `<equipment>`, `<recording>`, `<recordingStmt>`,  
`<scriptStmt>`

Elemente für die Transkription:

- `<incident>`, `<kinesic>`, `<pause>`, `<shift>`, `<u>`, `<vocal>`,  
`<writing>`

# 8 – Beispiel: Transkription

```
<vocal iterated="true">  
  <desc>regelmäßiges Zischen</desc>  
</vocal>  
  
<u who="#spkr1">si tu te déplaçais</u>  
<u trans="latching" who="#spkr2">Joe et moi l'aurions mis  
entre nous</u>  
  
<list type="speakers">  
  <item xml:id="spkr1">Markus</item>  
  <item xml:id="spkr2">Julia</item>  
</list>
```

# 9 – Dictionaries

## Spezielle Elemente für Wörterbücher / Einträge

### Strukturiert oder unstrukturiert

- `<entry>` (strukturiert) und `<entryFree>` (unstrukturiert)

### Einige grundlegende Elemente:

- `<def>`, `<form>`, `<gen>`, `<hom>`, `<etym>`, `<lang>`, `<mood>`,  
`<number>`, `<orth>`, `<pos>`, `<sense>`, `<syll>`, `<usg>`

### Weitere Elemente

- `<case>`, `<colloc>`, `<dictScrap>`, `<gram>`, `<gramGrp>`, `<hyph>`,  
`<iType>`, `<lbl>`, `<oRef>`, `<oVar>`, `<pRef>`, `<pVar>`, `<per>`,  
`<pron>`, `<re>`, `<stress>`, `<subc>`, `<superEntry>`, `<tns>`, `<xr>`

# 9 – Beispiel Wörterbuch-Eintrag

```
<entry>
  <form>
    <orth>poussin</orth>
    <pron>[pusë]</pron>
  </form>
  <gramGrp>
    <pos>n.</pos>
    <gen>m.</gen>
  </gramGrp>
  <sense n="1">Jeune poulet, nouvellement sorti de l'œuf,
  encore couvert de duvet.</sense>
  <sense n="2">Zool. Jeune oiseau (par rapport aux
  adultes, aux parents). </sense>
  <sense n="3"> (êtres humains)
    <def n="1">Fam. Terme d'affection (enfant). </def>
    <def n="2"> Sports. Catégorie d'âge (9 ans) qui
    précède celle des benjamins.</def>
  </sense>
</entry>
```



# 10 – Modul “Manuscript Description”

Dieses Modul stellt das Element `<msDesc>` bereit

- zur Beschreibung von Manuskripten
- im Header oder im Text
- strukturiert und/oder in Prosa
- auch für Inkunabeln, moderne Manuskripte, etc.

Die `msDesc` hat 4 Teile:

- `<msIdentifier>`, `<msContents>`, `<physDesc>`, `<history>`
- außerdem: `<additional>`
- und beliebig viele `<msPart>`

# 10 – Beispiel für <msDesc>

```
<msDesc>
  <msIdentifier>
    <settlement>Oxford</settlement>
    <repository>Bodleian Library</repository>
    <idno>MS. Add. A. 61</idno>
    <altIdentifier type="SC">
      <idno>28843</idno>
    </altIdentifier>
  </msIdentifier>
  <p>'Hic incipit Bruitus Anglie,' the De origine et
gestis Regum Angliae of Geoffrey of Monmouth (Galfridus
Monumetensis: beg. 'Cum mecum multa & de multis.' </p>
  <p>On fol. 54v very faint is 'Iste liber est fratris
guillelmi de buria de ... Roberti ordinis fratrum
Pred[icatorum],' 14th cent. (?): 'hanauilla' is written
at the foot of the page (15th cent.). Bought from the
rev. W. D. Macray on March 17, 1863, for £1 10s. </p>
</msDesc>
```

# 11 – Representation of Primary Sources

Spezielle Elemente für physische Aspekte eines Dokuments, zusätzlich zu denen aus dem Modul “Core”.

## Elemente zur Transkription

- `<addSpan>`, `<damageSpan>`, `<delSpan>`,
- `<am>`, `<ex>`, `<fw>`,
- `<handNotes>`, `<handShift>`,
- `<restore>`, `<subst>`, `<supplied>`
- `<damage>`, `<space>`,

## Elemente für digitale Faksimiles

- `<facsimile>`
- `<surface>`
- `<zone>`

# 11 – Beispiel <facsimile>

```
<TEI>
  <teiHeader>
    <!-- Infos im Header) --!>
  </teiHeader>
  <facsimile>
    <graphic url="page1.png"/>
    <graphic url="page2.png"/>
    <graphic url="page3.png"/>
    <graphic url="page4.png"/>
  </facsimile>
  <text>
    <!-- Transkription --!>
  </text>
</TEI>
```

# 12 – Critical Apparatus

Spezielle Elemente für die Repräsentation eines klassischen kritischen Apparats.

## Apparat-Einträge

- `<app>`, `<lem>`, `<rdg>`, `<rdgGrp>`

## Textzeugen

- `<wit>`, `<witDetail>`, `<witEnd>`, `<witStart>`, `<witness>`,  
`<listWit>`,

## Sonstiges

- `<lacunaEnd>`, `<lacunaStart>`, `<variantEncoding>`

# 12 – Beispiel: Apparat-Eintrag

```
<app>
  <lem wit="wit1">comte</lem>
  <rdgGrp type="orthographic">
    <rdg wit="#wit2">cante</rdg>
    <rdg wit="#wit3">contes</rdg>
  </rdgGrp>
</app>
```

# 13 – Names, Dates, Places and People

Spezielle Elemente für die Repräsentation “benannter Entitäten” (named entities)

Für Personen

- `<persName>`, `<forename>`, `<surname>`,
- `<birth>`, `<death>`, `<nationality>`, `<socecStatus>`, `<occupation>`

Für Orte

- `<country>`, `<region>`, `<settlement>`,
- `<placeName>`, `<geogName>`,

Organisiert in Listen

- `<listEvent>`, `<listNym>`, `<listOrg>`, `<listPerson>`, `<listPlace>`

# 13 – Beispiel für <person> und <name>

```
<person xml:id="Ovi01" role="Dichter">
  <persName xml:lang="la">Publius Ovidius Naso</persName>
  <birth when="-0044-03-20">20. März 43 BC
    <placeName>
      <settlement type="city">Sulmona</settlement>
      <country key="IT">Italien</country>
    </placeName>
  </birth>
  <!-- ... -->
</person>
<!-- ... -->
<text>
  <body>
    <p>Die Metamorphosen sind ein langes Gedicht in 15
    Büchern, das von <name ref="#Ovi01">Ovid</name>
    geschrieben wurde. </p>
  </body>
</text>
```



# 14 – Tables, Formulae and Graphics

Elemente, um Tabellen, Formeln und Grafiken zu repräsentieren

Tabellen

- `<table>`, `<row>`, `<cell>`,

Grafiken

- `<figure>`, `<figDesc>`

Formeln

- `<formula>`

# 14 – Beispiel: Tabelle

```
<table rows="4" cols="4">
  <head>Schnee-Persistenz in den schweizer Alpen</head>
  <row>
    <cell role="label">In einer Höhe von </cell>
    <cell role="data">650 m.</cell>
    <cell role="data">1300m.</cell>
    <cell role="data">1950m.</cell>
    <cell role="data">2700m.</cell>
  </row>
  <row>
    <cell role="label">bleibt der Schnee </cell>
    <cell role="data">77 Tage</cell>
    <cell role="data">200 Tage</cell>
    <cell role="data">245 Tage</cell>
    <cell role="data">365 Tage</cell>
  </row>
</table>
```

# 15 – Language Corpora

Dieses Kapitel handelt von der Bildung umfangreicher linguistischer Korpora

Es nutzt Elemente aus vielen anderen Kapiteln, und fügt einige spezielle Elemente hinzu, insbesondere für Korpus-Metadaten

- `<activity>`, `<channel>`,
- `<constitution>`, `<derivation>`, `<domain>`, `<factuality>`,  
`<interaction>`, `<locale>`, `<particDesc>`, `<preparedness>`,  
`<purpose>`, `<setting>`, `<settingDesc>`, `<textDesc>`

# 15 – Beispiel: <setting>

```
<settingDesc>  
  <setting who="#p1 #p2">  
    <name>Clermont-Ferrand</name>  
    <date>Mitte Dezember 2011</date>  
    <locale>Ceyrat</locale>  
    <activity>Das Auto zum Laufen bringen. </activity>  
  </setting>  
</settingDesc>
```

# 16 – Linking, Segmentation, Alignment

## Strukturierung / Segmentierung:

- `<ab>`, `<seg>`

## Links (Hyperlinks):

- `<alt>`, `<altGrp>`, `<anchor>`, `<join>`, `<joinGrp>`, `<link>`,  
`<linkGrp>`

## Zeitliche Verbindung / Synchronisation:

- `<timeline>`, `<when>`
- Attribute: `corresp`, `synch`, `sameAs`, `next`, `prev`, `exclude`, `select`.

# 17 – Simple Analytic Mechanisms

## Spezielle Elemente für Textanalyse

### Linguistische Analyse

- Segmenttation: `<c>`, `<cl>`, `<m>`, `<phr>`, `<s>`, `<w>`
- Analyse: Attribut: `@ana`

### Interpretationen

- `<interp>`, `<interpGrp>`, `<span>`, `<spanGrp>`

# 17 – Beispiel: Linguistische Annotation

```
<p>  
  <w ana="#03">Le </w>  
  <w ana="#SM">vin </w>  
  <w ana="#V">sait </w>  
  <w ana="#V">revêtir </w>  
  <w ana="#A">le plus </w>  
  <w ana="#J">sordide </w>  
  <w ana="#SM">lieux </w>  
  <w ana="#DF">d' </w>  
  <w ana="#DF">un </w>  
  <w ana="#SM">luxe </w>  
  <w ana="#J">miraculeux </w>  
</p>
```

# 21 – Certainty and Responsibility

Elemente und weitere Mechanismen, eine Aussage über eine bestimmte Kodierung auszudrücken

- Sicherheit: `<certainty>` - ist das gewählte Element korrekt?
- Präzision: `<precision>` - wie genau ist ein angegebener Wert/Zahl?
- Verantwortung: `<respons>` - wer ist für die Angabe, die Kodierung, etc. verantwortlich

Attribute für `<certainty>`

- `@locus` – auf welchen Aspekt des Markup bezieht sich die Unsicherheit? (Name/Position des Elements, Wert des Attributs)
- `@degree` – welches Maß an Sicherheit besteht? (0-1)



# 21 – Beispiel: "certainty"

## Einfachste Variante

```
...Elizabeth went to <placeName>Essex<certainty  
degree="0.6"/></placeName>...
```

## Genauere Variante

```
...Elizabeth went to <placeName xml:id="PN12">  
Essex</placeName>...
```

```
<certainty target="#PN12" locus="name" degree="0.6">  
  <desc>May also refer to a person, the Earl of  
  Essex.</desc>  
</certainty>
```

## Anmerkung

- Die genauere Variante erlaubt zahlreiche Erweiterungen

# 18-23 – Weitere Kapitel

18 – Feature Structures

19 – Graphs, Networks and Trees

20 – Non-hierarchical Structure

22 – Documentation Elements

23 – Using the TEI