

Entitäten in Editionen: TEI names, dates, places und Normdaten

Entitäten in Editionen: TEI names, dates, places und Normdaten - Markus Schnöpf



Leopoldina
Nationale Akademie
der Wissenschaften



BERGISCHE
UNIVERSITÄT
WUPPERTAL

Auszeichnen von Personen, Orten & Daten

ermöglicht bei konsistenter Kodierung z.B.

- automatische Erstellung von Registern
- differenzierte Suchmöglichkeiten
- prosopographische Analysen
- Georeferenzierungen
- Anreicherung mit Normdaten
- Integration externer Informationen
- Vernetzung

TEI-Vokabular für Namen

<name/> (name or proper noun)

➤ enthält einen Eigennamen

➤ **<rs/>** (reference string)

➤ allgemeiner gehalten;

➤ kann auch indirekte Namensnennungen auszeichnen, z.B.:

➤ „GUZ“,

➤ „der Sänger der Aeronauten“

➤ „die Anwesenden“ usw.

➤ Spezifizierung durch **@type** (**person**, **place**)

<name/> und **<rs/>** sind Bestandteil des **TEI-core** moduls.

Namen vs. ‚Entitäten‘

- Namen können ausgezeichnet werden:
 - im Text selber: /TEI/text/body/p o.ä.
 - In der profile description des Headers (participant description oder setting description):
 - /TEI/teiHeader/profileDesc/**particDesc**/listPerson/**person**
 - /TEI/teiHeader/profileDesc/**particDesc**/org
 - /TEI/teiHeader/profileDesc/**settingDesc**/listPlace/**place**
- <person/>, <org/>, <place/> beschreiben ‚Entitäten‘, nicht Namen
- Vorteil bei Auszeichnung im Header:
 - Der Name wird nur einmal genau spezifiziert
 - Personen und Orte können mit einer ID versehen werden, auf die im Text verwiesen werden kann
- Namenslisten können in separate TEI-Dateien ausgelagert werden

<rs/> vs. <name/>

<p>

<q>My dear **<name>Mr .**

Bennet</name>, </q> said

<rs>his lady</rs> to him one day,

<q>have you heard that

<name>Netherfield Park</name> is
let at last?

</q>

</p>

Wichtige Attribute: @type und @role

@type

spezifiziert die Art des Names- bzw. Referenzstrings

- z.B. Personennamen, Geoname, Körperschaftsname
- ggf. weitere Spezifizierung per @subtype

➤ @role

spezifiziert die Rolle einer Person im jeweiligen Kontext

- z.B. „Sprecher“, „Autor“, „Editor“, „Schreiber“ etc.
- Berufsbezeichnungen
- etc.

Beispiel: @type und @role

```
<p><q>My dear <name
type="person">Mr. Bennet</name>,
</q> said <rs type="person"
role="speaker">his lady</rs> to
him one day, <q>have you heard
that <name type="place">
Netherfield Park</name> is let at
last?</q></p>
```

Wichtige Attribute: @key und @ref

(siehe [att.canonical](#))

- **@key** zur Identifizierung der genannten Person, Stadt etc.
 - Werte für **@key** werden durch die Editoren (anderswo) definiert
 - sollten konsistent und einigermaßen verständlich sein
- **@ref** zur eindeutigen Referenzierung der genannten Person, Stadt etc.
 - verweist auf eine Liste, Datei, Ressource, in der die „Entität“ beschrieben wird
 - intern z.B. im TEI Header
 - extern z.B. in einer Normdatei (GND o.ä.)
 - in Form einer URL oder URI (Uniform Resource Identifier)
- **@key** => weniger streng formalisiert
- **@ref** => eindeutiger bei Datenaustausch

Beispiel: @key und @ref bei Namen

```
<p>  
<q>My dear  
<name type="person" key="BENM1"  
ref="http://janeausten.wikia.com/wiki/Mr._Benn  
et">Mr. Bennet</name>,  
</q> said  
<rs type="person" key="BENM2"  
ref="http://janeausten.wikia.com/wiki/Mrs._Ben  
net">his lady</rs>  
to him one day, <q>have you heard that  
<name type="place" key="NETP1"  
ref="http://janeausten.wikia.com/wiki/Netherfi  
eld">Netherfield Park</name>  
is let at last?</q>  
</p>
```

Exkurs: URI

- URI besteht aus 5 Bestandteilen (nur scheme und path obligatorisch)
 - **scheme (Schema)**
 - authority (Anbieter)
 - **path (Pfad)**
 - query (Abfrage)
 - fragment (Teil)
- Allgemeiner Aufbau: **scheme://authoritypath?query#fragment**
 - http://de.wikipedia.org/wiki/Uniform_Resource_Identifier#Aufbau
 - <https://www.google.de/?q=URI>
 - urn:nbn:de:bvb:12-bsb00036610-5
- Bei internen Verweisen kann auch nur auf das fragment verwiesen werden
 - ref="#person01"
- Fragmente werden in der Quelle durch das Attribut @id oder @xml:id spezifiziert

Exkurs: Wichtige Normdateien / Authority Files

➤ GND (Gemeinsame Normdatei)

Personen, Körperschaften, Orte

- Beschreibung: http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html
- Abfrage über OGND: <http://www.ub.uni-freiburg.de/index.php?id=3343>

➤ LCCN (Library of Congress Control Number)

Amerikanisches Pendant zur GND

- <http://authorities.loc.gov/>

➤ VIAF (Virtual International Authority File)

Mapping verschiedener Normdateien

- <http://viaf.org/>

➤ Geonames

Offenes Projekt für Ortsnamen

- <http://www.geonames.org/>

➤ Getty Thesaurus of Geographic Names (TGN)

- <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>
- z.B.: [ID 7004717](#)

Modul namesdates

Für eine genauere und strukturierte Beschreibungen von Namen stellt das Modul *namesdates* ([Kapitel 13](#)) zahlreiche weitere Elemente zur Verfügung, z.B.:

- Personennamen: `<persName />`
- Körperschaftsnamen : `<orgName />`
(organizational names)
- Ortsnamen: `<placeName />`, `<geogName />`

Personennamen

`<persName />` kann u.a. enthalten

- `<surname />`
- `<forename />`
- `<addName />` (Zusätze wie ‚Dr.‘, ‚der Große‘ usw.)
- `<genName />` (‚generational name‘, z.B. „Sen.“, „Jun.“, „VII.“)
- `<nameLink />` (‚van der‘, ‚of‘)
- `<roleName />` (Titel u.ä., z.B. ‚president‘, ‚König‘)
- jeweils spezifizierbar durch `@type`

Könnte theoretisch auch durch verschachtelte `<name />`-Elemente realisiert werden.

Beispiel: Personennamen

```
<p>  
  <persName>  
    <forename type="given">Margaret</forename>  
    <forename type="unused">Hilda</forename>  
    <surname type="maiden">Roberts</surname>  
    <surname type="married">Thatcher</surname>  
  </persName>  
</p>
```

```
<p>  
  <persName ref="http://d-nb.info/gnd/118560034">  
    <roleName>König</roleName>  
    <forename>Karl</forename>  
    <genName>I</genName>  
    <addName>der Große</addName>  
  </persName>  
</p>
```

Ortsnamen

- **<placeName/>**: ‚normaler‘ Ortsname, z.B.:
 - Rostock
 - USA
 - Atlantis
- **<geogName/>**: geographischer Name, z.B.:
 - Rhein
 - Schwarzwald
 - Die Gefilde der Glückseligen
- mögliche Unterelemente:
 - **<district/>**, **<settlement/>**, **<region/>**, **<country/>**, **<bloc/>**, **<geogFeature/>**
- **<orgName/>** für Körperschaften, kann Unterelemente aus Personen und Ortsnamen enthalten, z.B.:
 - Uwe Johnson-Forschungsstelle Rostock

Beispiel: Orts- und Körperschaftsnamen

```
<p>
  <placeName>
    <district>Harlem</district>
    <settlement type="city">New York</settlement>
    <region type="state">N.Y.</region>
    <country>USA</country>
  </placeName>

  <orgName>
    <orgName>Uwe Johnson-Forschungsstelle</orgName>
    <orgName>
      <name type="role">Universität</name>
      <name type="city">Rostock</name>
    </orgName>
  </orgName>
</p>
```

Namen vs. ‚Entitäten‘

- **Namen** können ausgezeichnet werden:
 - im Text selber: `/TEI/text/body/p` (o.ä.)
 - als Teil einer Personen- bzw. Ortsliste in der profile description des Headers (z.B. participant description oder setting description):
 - `/TEI/teiHeader/profileDesc/particDesc/listPerson/person`
 - `/TEI/teiHeader/profileDesc/particDesc/listOrg/org`
 - `/TEI/teiHeader/profileDesc/settingDesc/listPlace/place`
- `<person/>`, `<org/>`, `<place/>` beschreiben **Entitäten**, nicht Namen
- Vorteil bei Auszeichnung im Header:
 - Der Name wird nur einmal genau spezifiziert
 - Personen und Orte können mit einer ID versehen werden, auf die im Text verwiesen werden kann
- Namenslisten können in separate TEI-Dateien ausgelagert werden

Details zu Personen, Orten und Körperschaften

- `<person/>`, `<place/>`, `<org/>`
 - sind übergeordnete Elemente
 - die Entitäten beschreiben
 - und nicht nur Namen enthalten können, sondern auch z.B.:
 - alternative Namen
 - Informationen zur Biographie, z.B. `<birth/>`, `<death/>`
 - Informationen zu Eigenschaften `<trait/>`, Geschlecht `<sex/>`, etc.
 - Informationen zur geographischen Lage: `<location>`, `<geo>`
- `<person/>`, `<place/>`, `<org/>` können nur im TEI-Header oder in entsprechenden Listen (`<listPerson/>`, `<listPlace/>`, `<listOrg/>`) benutzt werden!
- `<persName/>`, `<placeName/>`, `<orgName/>` können auch im Text benutzt werden
- Referenzen auf Normdateien (z.B. mit `@ref` auf GND) sind nur bei Namen, nicht bei Entitäten möglich

Beispiel: Ausführlicher Personendatensatz

```
<person xml:id="KARL01">
  <persName xml:lang="de" >
    <roleName>König</roleName>
    <forename>Karl</forename>
    <genName>I</genName>
    <addName>der Große</addName>
  </persName>
  <persName xml:lang="la">
    <forename>Carolus</forename>
    <addName>Magnus</addName>
  </persName>
  <birth>2. April 747 oder 748</birth>
  <death>28. Januar 814</death>
  <sex>männlich</sex>
  <trait type="physical">
    <label>Augenfarbe</label>
    <desc>blau</desc>
  </trait>
</person>
```

Beispiel: Ausführliche Ortsbeschreibung

```
<settingDesc>
  <listPlace>
    <place xml:id="NEWYORK">
      <placeName>
        <settlement type="city">New York
City</settlement>
        <region type="state">New York</region>
        <country>USA</country>
      </placeName>
      <location>
        <geogName>Nordamerika</geogName>
        <geo>40° 43' N, 74° 0' W</geo>
        <geo>40.712778°, -74.005833°</geo>
      </location>
    </place>
  </listPlace>
</settingDesc>
```

Beispiel: Verweise auf Personen und Orte

<p>

```
<rs ref="#MAGGIE">Die ehemalige  
englische Ministerpräsidentin</rs>  
äußerte sich bei ihrem Besuch im  
<name ref="#NEWYORK">Big  
Apple</name>  
mehrfach bewundernd über den  
<rs ref="#KARL01">ersten  
Karolingerkaiser</rs>.
```

</p>

Elemente

- `<date/>` zeichnet eine Datumsangabe aus
- `<time/>` zeichnet eine Zeitangabe aus
- normalisierte Zeitangaben (W3C-Stil, gregorianischer Kalender) durch
 - `@when`, `@notBefore`, `@notAfter`, `@from`, `@to`, `@dur`
- alternative Normierungen durch
 - `@when-custom`, `@from-custom` usw.
- ggf. wichtig
 - `@calendar` und `<calendarDescription/>`
- Bestandteile von TEI core

Beispiele für Datums- und Zeitangaben

<p>

Heute ist der

<date when="2014-03-13">13. März</date> und es ist jetzt
<time when="15:30:00">halb vier</time>.

Das **<date from="1401" to="1500">15. Jahrhundert</date>**
dauerte **<date dur="P100Y">100 Jahre</date>**.

<name type="person">Lenin</name> wurde am
**<date calendar= "#julianisch" when="1870-04-22"
when-custom="1870-04-10">10. April 1870</date>**
in **<name type="place">Simbirsk</name>** geboren.
**<date notBefore="1870-04-22">Bald nach seiner
Geburt</date>** lernte er reden.

</p>

calendarDescription im TEI Header

```
<TEI>
  <teiHeader>
    <profileDesc>
      <calendarDesc>
        <calendar xml:id="julianisch">
          <p>Julianischer Kalender</p>
        </calendar>
      </calendarDesc>
    </profileDesc>
  </teiHeader>
</TEI>
```

Übung

- Zeichnen Sie in ihrem Übungs-XML die vorkommenden Namen und Datumsangaben aus
- Erstellen Sie dazu eine Liste der im Text vorkommenden Personen (nicht alle, nur 3-4) und kodieren Sie diese im TEI-Header
- Ergänzen Sie Normdaten