

XML-Auszeichnung mit dem DTA-Basisformat (DTABf)

Susanne Haaf

3. Münchner Digital Humanities Summerschool
11.-15. September 2017

Das **<div>**-Element

Die Textenteilung (z.B. Kapitel, Abschnitte, Teile ...) eines Buches wird mittels verschachtelter **<div>**-Elemente dargestellt. Diese enthalten in der Regel ein Attribut **@n**, welches die Strukturebene angibt.

<http://www.deutschestextarchiv.de/doku/basisformat/div.html>

```
<div n="1">
```

```
  <head>[Titel Kapitel 1]</head>
```

```
  <div n="2">
```

```
    <head>[Titel Unterkapitel 1.1]</head>
```

```
    <p>[Text]</p>
```

```
    <p>[Text]</p> ...
```

```
  </div>
```

```
</div>
```

Das **<div>**-Element

Mit **@type** kann das **<div>**-Element näher bestimmt werden. Folgende Werte kann **@type** dabei z.B. annehmen:

- Textabschnitte:
 - **chapter** Kapitel
 - **dedication** Widmung
 - **diaryEntry** Tagebucheintrag
 - **edition** Abdruck einer externen Textquelle
 - **letter** Brief
 - **poem** Gedicht
 - **recipe** Rezept
- Textabschnitte im Drama:
 - **act** Akt im Drama
 - **scene** Szene im Drama

Das **<div>**-Element

Mit **@type** kann das **<div>**-Element näher bestimmt werden. Folgende Werte kann **@type** dabei z.B. annehmen:

- Register und Verzeichnisse:
 - **abbreviations** Abkürzungsverzeichnis
 - **appendix** Anhang
 - **bibliography** Bibliographie
 - **contents** Inhaltsverzeichnis
 - **figures** Verzeichnis der Abbildungen
 - **index** Register

☞ Darüber hinaus sind weitere Werte möglich (vgl. <http://www.deutschestextarchiv.de/doku/basisformat/div.html>)

DTABf: Weitere häufige Elemente

- Textarten
 - **<p>** Paragraph im Prosatext
 - **<lg>** "line group", i.e. Versgruppe (Gedicht, Strophe)
 - **<l>** Vers
 - Hervorhebungen:
 - **<hi>** (highlighting) mit **@rendition**
 - Folgende Werte kann **@rendition** z.B. haben:
 - **#b** Fettdruck
 - **#i** Kursivdruck
 - **#c** zentrierter Text
 - **#g** gesperrter Text
 - **#aq** Wechsel zur Antiqua-Schrift
- vgl. <http://www.deutschestextarchiv.de/doku/basisformat/typogrBspe.html>

DTABf: Häufige Elemente

- Erscheinungsbild der Seite
 - **<fw>** Orientierungshilfen im Buchblock, mit **@type**
 - **sig** Bogensignatur
 - **catch** Kustode (Catchword)
 - **header** Kolumnentitel

```
<fw type="sig" place="bottom">3a</fw>
```
 - **<pb/>** Seitenumbruch mit den Attributen **@n** (Seitenzahl) und **@facs** (Faksimilenummer)

```
<pb n="8" facs="#0018" />
```
 - **<lb/>** Zeilenumbruch
 - **<cb/>** Spaltenumbruch

DTABf: Häufige Elemente

- **<note>** Anmerkungen mit Attribut **@place** (Ort der Anmerkung):
 - **"foot"** Fußnote
 - **"end"** Endnote
 - **"right"** Marginalie rechts
 - **"left"** Marginalie links

... und **@type**:

 - **"editorial"** editorischer Sachkommentar

DTABf: Häufige Elemente

- Überschriften
 - **<head>** als Unterelement in:
 - **<div>** Kapitel-/Paragraphenüberschrift
 - **<lg>** Gedichttitel
 - **<list>** Titel einer Liste
 - **<table>** Titel einer Tabelle
- Zitate
 - **<cit>** Zitat insgesamt
 - **<quote>** zitierter Text
 - **<bibl>** bibliographische Angabe

<cit>
 <quote>[zitierter Text]**</quote>**
 <bibl>[bibliographische Angabe]**</bibl>**
</cit>

DTABf: Häufige Elemente

- **<table>** Tabelle
- **<row>** Tabellenzeile
- **<cell>** Tabellenzelle

<table>

<head>[ggf. Titel der Tabelle]**</head>**

<row>

<cell>[Text einer Tabellen-Zelle]**</cell>**

<cell>[Text einer Tabellen-Zelle]**</cell>**

</row> ...

</table>

<http://www.deutschestextarchiv.de/doku/basisformat/tabAllg.html>

DTABf: Häufige Elemente

- **<list>** Liste, mit Unterelement:
- **<item>** Listeneintrag

<list>

<head>[ggf. Titel der Liste]**</head>**

<item>[Inhalt eines Listenpunkts]**</item>**

<item>[Inhalt eines Listenpunkts]**</item>**

</list>

<http://www.deutschestextarchiv.de/doku/basisformat/liAllg.html>

DTABf-Dokumentation ...

DTABf-Schema

1. Schema (RNG):

<http://www.deutschestextarchiv.de/basisformat.rng>

*Welche Elemente & Attribute der TEI sollen Teil des DTABf sein?
Welche Werte werden festgelegt?*

2. Schematron-Regeln: ~/basisformat.sch

*Welche besonderen Regeln sollen für den Elementinhalt gelten?
Welche Beziehungen zwischen den Elementen sollen festgelegt
werden? ...*

Schema einbinden im oXygen-Editor

The screenshot shows the oXygen-Editor interface with the 'Dokument' menu open. The 'Schema' option is highlighted, and a submenu is displayed with the following items: 'Definition zeigen' (Strg+Umschalt+Eingabe), 'Schema zuweisen...', 'Externes Schema öffnen', and 'Schema erzeugen/konvertieren...'. The 'Schema zuweisen...' option is selected. In the background, the XML editor shows a document named 'Unbenannt1.xml*' with the following content:

```
1 <?xml version="1.0" encoding="UTF-8"
2 <
3 <TEI xmlns="http://www.tei-c.org/ns/
4 <teiHeader>
5 <fileDesc>
6 <titleStmt>
7 <title>Title</title>
8 </titleStmt>
9 <publicationStmt>
10 <p>Publication Information</p>
11 </publicationStmt>
12 <sourceDesc>
13 <p>Information about the source</p>
14 </sourceDesc>
15 </fileDesc>
16 </teiHeader>
17 <text>
18 <body>
19 <p>Some text here</p>
20 </body>
21 </text>
22 </TEI>
```

The 'Schema zuweisen' dialog box is open in the foreground, showing the following fields and options:

- URL:
- ☒ Einen Pfad benutzen, der zum Dateipfad relativ ist
- Schematyp:
- ☐ Zusätzliche Assoziation für eingebettete Schematron-Regeln hinzufügen
- Öffentliche ID:
- Buttons: OK, Abbrechen

A large blue arrow points from the 'Schema zuweisen...' option in the menu to the 'Schema zuweisen' dialog box.

Des Autoris Gedancken in seiner letzten Krankheit.

Wenn Blut und Lüste schäumen/
So stärke meinen Geist/
Daß er sich auch im Träumen
Aus Satans Neze reiße.

Hilff für mein Bestes sorgen/
Verändere meinen Sinn/
Und mache/ daß ich morgen
Ein neu Geschöpfe bin.

Ich seh' das Licht verschwinden/
Die trübe Nacht bricht ein/
Ach Herr/ laß meine Sünden
Auch mit verschwunden seyn/
Streich' sie aus deinem Buche/
Das mich zum Schuldner macht/
Und rette mich vom Fluche/
Der mir war zugebracht.

Wenn heut mein Ziel der Jahre
Mein letzter Abend ist/
Wolan ! wann ich nur fahre
Wo deine Wohnung ist.
(Wo du/ mein Vater/ bist.)
Doch soll ich länger leben/
So laß den festen Schluß
Mir stets für Augen schweben/
Daß ich einst scheiden muß!

A

Das

www.deutschestextarchiv.de/canitz_gedichte_1700/14

→ 1. Textanalyse → Gedicht!

2. Recherche in der DTABf-Doku:



Home / Inhaltlich / Textkörper / Gedichte und gebundene Sprache / Gedichte in Gedichtbänden

Gedichte in Gedichtbänden

In Gedichtbänden können innerhalb von Gedichten neben dem Gedichttitel und den einzelnen Strophen zusätzliche Textbestandteile vorkommen, z.B. Datumszeilen, Sprecher oder zusätzliche Angaben zur Entstehungsgeschichte des Gedichts. In Gedichtbänden mit solcherlei Gedichten ebenso wie in literarischen Sammlungen verschiedener Textsorten wird die Grundstrukturierung mittels `<div>`-Elementen vorgenommen. Ein Gedicht wird dabei durch ein `<div>`-Element umschlossen; der Gedichttitel steht im `<head>` des umschließenden `<div>`-Elements. Datumszeilen, Sprecher, zusätzliche Angaben in einfacher Prosa werden wie gewohnt ausgezeichnet. Die Gedichtauszeichnung mittels `<lg>`s beginnt auf Strophenebene.

Auszeichnung von Gedichten in Gedichtbänden (beispielhaft):

```
<div type="poem">
  <head>[Titel]</head><!-- sofern vorhanden -->
  <dateline>[Datumsangabe]</dateline>
  <lg type="poem">
    <lg n="[Strophennummer]"><!-- sofern kein einstrophiges Gedicht -->
      <1>[Vers]</1>
      <1>[Vers]</1>
    </lg>
  </lg>
</div>
```

3. Annotation mithilfe des Schemas



Validierung der XML-Instanzen im oXygen

121 <lb/>
122 <l
123 <l> lg (line group) #x017F;chwinden/</l>
124 <lb> l contains one or more
125 <l> lb verse lines l>
126 <lb> functioning as a
127 <l> formal unit, e.g. a >
128 <lb> sterss-refrein

129 <l>
130 <lb/>
131 <l>Streich&#
132 <lb/>
133 <l>Das mich
134 <lb/>
135 <l>Und rette
136 <lb/>
137 <l>Der mir w
138
139 <lb/>
140 <lg n="3">

122 <lg type
123 <l>Ich type characterizes the x017F;chwinden/</l>
124 <lb/> corresp element in some
125 <l>Die an sense, using any
126 <lb/> next convenient
127 <l>Ach prev classification
128 <lb/> rend scheme or typology.
129 <l>Auch rend
130 <lb/>
131 <l>Streich&
132 <lb/>
133 <l>Das mich
134 <lb/>
135 <l>Und rett
136 <lb/>
137 <l>Der mir
138
139 <lb/>
140 <lg n="3">

122 <lg type=""
123 <l>Ich ſchwinden/</l>
124 <lb/>
125 <l>Die ͤbe Nacht bricht ein/</l>
126 <lb/>
127 <l>Ach Herr/ laß meine Suͤnden/</l>
128 <lb/>
129 <l>Auch mit verſchwunden ſeyn/</l>
130 <lb/>
131 <l>Streich’ ſie aus deinem Buche/</l>
132 <lb/>
133 <l>Das mich zum Schuldner macht/</l>
134 <lb/>
135 <l>Und rette mich vom Fluche/</l>

Annotationsübung II: Tagging von Lyrik/Drama nach DTABf

Übung I: Blumenbach (leicht)

1. Im oXygen das DTABf-Schema einbinden:
> Dokument > Schema > Schema zuweisen
2. Der Text des Kapitels wird mit <div> umschlossen. Dieses <div> befindet sich auf der ersten Hierarchieebene, @n ist somit "1".
Haben wir einen @type, der hier passt? (vgl. <http://www.deutschestextarchiv.de/doku/basisformat/div.html>)
3. Der Titel des Kapitels steht in <head>, das direkt auf das öffnende <div> folgt.
4. Jeder Paragraph inkl. Paragraphenzeichen wird wiederum durch ein <div> umschlossen. Wir befinden uns dabei auf der zweiten Hierarchieebene, ergo: @n="2".
5. Jedes Paragraphenzeichen steht in <head>.

Übung I: Blumenbach (leicht)

6. Der Seitenumbruch am Beginn der Seite wird mit `<pb>` umschlossen. Die Seitenzahl steht im Attribut `@n`, die Faksimilenummer im Attribut `@facs`. Achtung: Das `<pb>`-Element ist ein leeres Element (d.h. vor der schließenden spitzen Klammer steht der Slash: `<pb/>`).
7. Die Fußnoten stehen in `<note>`-Elementen, die `@place="foot"` enthalten (für "Fußnote"). Das Fußnotenzeichen steht in `@n`. Die Fußnoten werden an der Stelle im Text wiedergegeben, an der sie im Text eingewiesen sind.
8. Am Ende jeder Zeile steht `<lb/>`, um den Zeilenumbruch anzuzeigen.
9. Eine der Fußnoten enthält ein Zitat. Wir zeichnen dieses mit `<cit>`, `<quote>` und `<bibl>` aus, vgl.
<http://www.deutschestextarchiv.de/doku/basisformat/zeMitAutor.html>

Übung II: Gedicht (leicht bis mittel)

1. Im oXygen das DTABf-Schema einbinden:
> Dokument > Schema > Schema zuweisen
2. Wir beginnen mit dem Titel "Der 139. Psalm".
3. Das Gedicht wird mit `<lg type="poem">` umschlossen.
4. Der Gedichtstitel steht in `<head>`
5. Jede Strophe steht in `<lg>`. Die jeweilige Strophennummer wird in `@n` wiedergegeben.
6. Jede Zeile im Text wird durch den Zeilenumbruch abgeschlossen (`<lb/>`)
7. Jeder Vers im Text wird mit `<l>` umschlossen. Achtung: Verse können über einen Zeilenschluss hinausreichen.

Übung II: Gedicht (leicht bis mittel)

8. Die Kustode steht in `<fw type="catch" place="bottom">`
9. Der Seitenumbruch am Beginn der Seite wird mit `<pb>` umschlossen. Die Seitenzahl steht im Attribut `@n`, die Faksimilenummer im Attribut `@facs`.
10. Die Seite 31 wollen wir nicht taggen. Wie gebe ich an, dass in meiner Edition Text fehlt? vgl.:
<http://www.deutschestextarchiv.de/doku/basisformat/gapSupplied.html>
11. Wir beenden das Gedicht auf Seite 31 (mit `</lg>`), taggen den Seitenumbruch (wie in 9.) und taggen das Gedicht auf Seite 32
12. Wie kann ich die Initiale im ersten Vers des jeweiligen Gedichts auszeichnen?

Übung III: Drama (für Fortgeschrittene)

1. Die Abbildungen zeigen die Figurenaufstellung im Drama, den Beginn eines Aktes, zweier Auftritte, Bühnenanweisungen, Sprecher und Sprechakte.
2. Das DTABf dokumentiert hier:
<http://www.deutschestextarchiv.de/doku/basisformat/drama.html>
das Tagging in Dramen. Bitte lesen und das gegebene Dokument entsprechend taggen.
3. Die Seiten folgen nicht direkt aufeinander. Mit `<gap>` kann ich auszeichnen, dass Text fehlt. Wie genau das geht, steht im DTABf-Doku-Kapitel zu `<gap>`, welches z.B. über die Übersicht zu allen Elementen erreichbar ist.
4. Sprecher sind vom übrigen Text durch Frakturwechsel abgehoben. Wie kann ich diesen Frakturwechsel auszeichnen?
5. Sprecherangaben in Bühnenanweisungen sind gesperrt gedruckt.