

Anleitung XML-Kodierung

- Struktur
 - Blatt- und Seitenzählung
 - Überschriften
- Referenzierung
 - Bildung von eindeutigen xml:ids
 - Bibliographische Verweise
 - Verweise in das VD 16
 - Verlinkungen auf andere Dokumente der Edition
 - Verweise auf (Dokumente) ander e(r) Webseiten
 - Verlinkung auf andere Textstellen innerhalb der Edition
- Fußnoten
 - Fußnoten - allgemein
 - Fußnoten in der Druckfassung
 - Textkritischer Apparat
- Auszeichnung
 - Kurztitel
 - Kodierung von Sonderzeichen/Graphem- und Phonemwechsel
 - Neue Kodierung spraxis
 - Alte Kodierung spraxis
-
-
- Zeichensetzung
 - Neue Kodierung spraxis
 - Alte Kodierung spraxis
- Auflösung von Abkürzungen
- Kodierung unterschiedlicher Absätze oder Sinneinheiten
- Kodierung ausgelassener Textteile
- Hervorhebungen im Druck
- Probleme/Hinweise

Struktur

Blatt- und Seitenzählung

Zur Unterscheidung einer Blatt- und Seitenzählung wird folgendermaßen kodiert:

```
<!-- Paginierung -->
<pb n="3v"/>

<!-- Seitenzählung -->
<pb n="3" type="seite"/>
```

Die Ausgabe im ersten Fall:

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. [Blatt: 3v]

Die Ausgabe im zweiten Fall:

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. [Seite: 3]

Überschriften

Header können über @type unterschieden werden:

```
<head type="main">Hauptüberschrift Kapitel 1</head>
<head type="sub">Unterüberschrift Kapitel 1</head>
```

Bei der Verwendung von <head> muss darauf geachtet werden, dass das <head>-Element nur unmittelbar auf ein <div> Element folgen darf:

```
<div type="chapter" n="1">
  <head type="main">Hauptüberschrift Kapitel 1</head>
  <p>Text zur Hauptüberschrift</p>
  <div type="section" n="1">
    <head type="sub">Unterüberschrift Sektion 1</head>
    <p>Text Text zur Unterüberschrift Sektion 1</p>
  </div>
  <div type="section" n="2">
    <head type="sub">Unterüberschrift Sektion 2</head>
    <p>Text zur Unterüberschrift Sektion 2</p>
  </div>
</div>
```

Referenzierung

Bildung von eindeutigen xml:ids

die Benennung von xml:ids kann frei erfolgen; dabei ist darauf zu achten, dass:

- 1.) die id bisher nicht verwendet wurde (auch in anderen Registern)
- 2.) keine Sonderzeichen verwendet werden (z.B. Umlaute)

- 3.) auf Groß- und Kleinschreibung verzichtet wird
- 4.) keine Satzzeichen verwendet werden
- 5.) keine Spatien verwendet werden

Bibliographische Verweise

Im Text werden sämtliche Titelverweise in Kurzform gemacht, z.B. Schwan: Vermächtnis, S. 30 oder einfach Schwan: Vermächtnis.

In einem externen Literaturverzeichnis (listBibl.xml) wird der Langtitel aufgenommen:

```
<div type="bibliography">
  <head>Bibliographische Nachweise und
  Forschungsliteratur</head>
  <listBibl>
    <bibl xml:id="schwan_vermaechtnis">
      <author>
        <surname>Schwan</surname>
        <forename>Heribert</forename>
      </author>
      <author>
        <surname>Jens</surname>
        <forename>Tilman</forename>
      </author>
      <title>Vermächtnis. Die
  Kohl-Protokolle</title>
      <pubPlace>s.l.</pubPlace>
      <date>2014</date>
    </bibl>
  </listBibl>
</div>
```

Aus dem Text heraus wird auf die @xml:id verwiesen.

Soll später im Text aus dem Kurztitel automatisch ein Langtitel generiert werden, dann sollte das mittels @subtype markiert werden, z.B.:

```
Siehe hierzu das Kapitel bei <ref type="bibliography"
  subtype="long"
  target="#schwan_vermaechtnis">Schwan</ref>, S. 123.
```

Durch die Differenzierung über den @subtype kann ich später automatisch den Inhalt von <ref> durch die Angaben im Literaturverzeichnis ersetzen.

Wird das Element <ptr> innerhalb des Elements <ref> verwendet, ohne dass automatisiert ein ", in:" eingefügt werden soll, erhält das Element den type="notein".

Verweise in das VD 16

```
<ptr type="vd16" cRef="R+784"/>VD16 R 784).</ptr>
```

Durch das <ptr>-Element wird automatisch in der digitalen Editionen folgendes Aussehen

generiert:

4° [VD16 R 784] als anklickbarer Link. Der im bibliographische Verweis in runden Klammern kann entfallen, so dass optimalerweise der Quellcode so aussieht:

```
<bibl>Abschiet [...] <lb/>Wittenberg: Schirlentz,  
Nickel [1532], 11 Bl., 4°. <ptr type="vd16"  
cRef="R+784" /></bibl>
```

Achtung: bei der Bildung des @cRef-Inhalts zwischen Buchstaben und Zahlen unbedingt ein + verwenden!

Verlinkungen auf andere Dokumente der Edition

```
<note>Vergleiche hierzu <ref target="#xyz">Text  
XYZ</ref></note>
```



Die zu referenzierende id innerhalb von @target ist identisch mit dem Dateinamen in der Dropbox, z.B. <ref target="frankfurter_anstand.xml">

Dabei wird über das @type die Zuordnung geregelt. Handelt es sich um einen Einleitungstext, dann muss "einleitung" als Attributswert benutzt werden. Quellentexte werden mit "quelle" gekennzeichnet. Eine Verlinkung auf die Einleitung zum Frankfurter Anstand sähe demzufolge so aus:

```
<ref type="einleitung" target="frankfurter_anstand_einleitung.xml">
```

Verweise auf (Dokumente) andere(r) Webseiten

```
<note>Vergleiche hierzu <ref  
target="http://www.xzy.com/usw.pdf">Text  
XYZ</ref></note>
```

Die referenzierten Webseiten und Dokumente sollten stabil und möglichst persistent sein.

Verlinkung auf andere Textstellen innerhalb der Edition

Verweise auf Textstellen in anderen Dokumenten werden folgendermaßen kodiert:

1. Zunächst muss die zu referenzierende Textstelle eine @xml:id bekommen. Da dies ein globales Attribut ist, kann man @xml:id in jedem Element verwenden. Möchte man beispielsweise einen ganzen Abschnitt referenzieren, dann kann man dem entsprechenden Abschnitt einfach eine id vergeben, z.B. <p xml:id="frankfurter_anstand_p1">. Wenn im Text allerdings kein oder kein semantisch sinnvolles Element vorhanden ist, kann man die Textstelle mit <anchor/>, <seg> oder <ab> auszeichnen und eine @xml:id vergeben.

```

<div xml:id="frankfurter_anstand_div1">
  <p xml:id="frankfurter_anstand_p1">
    Lorem ipsum dolor sit amet, consetetur sadipscing
    elitr,
    sed diam nonumy eirmod tempor invidunt ut labore et
    dolore magna aliquyam erat,
    sed diam voluptua.
  </p>
  <ab xml:id="frankfurter_anstand_ab1">
    At vero eos et accusam et justo duo dolores et ea
    rebum.
  </ab>
  <p xml:id="frankfurter_anstand_p1">
    Stet clita kasd gubergren, no sea takimata sanctus est
    Lorem ipsum dolor sit amet.
    Lorem ipsum dolor sit amet, consetetur sadipscing
    elitr, sed diam nonumy eirmod tempor invidunt ut
    <seg xml:id="frankfurter_anstand_seg1">labore et
    dolore</seg> magna aliquyam erat, sed diam voluptua.
    At vero eos et accusam et justo duo dolores et ea
    rebum. Stet clita kasd gubergren, no sea takimata
    sanctus est
    Lorem ipsum dolor<anchor
    xml:id="frankfurter_anstand_anchor1"/> sit amet.
  </p>

```

Verwendung von `<anchor/>`: wenn kein bestimmter Textinhalt, sondern nur eine einzelne Stelle, bzw. ein Punkt verlinkt werden soll.

Verwendung von `<ab>` (anonymous block): auf struktureller Ebene, z.B. auf `<div>`-Ebene; außerhalb von `<p>`.

Verwendung von `<seg>` (arbitrary segment): Textspanne auf unterster Ebene, z.B. innerhalb von `<p>`.

2. Analog zur [Verlinkung auf andere Dokumente](#) muss jetzt ein Link auf die id gesetzt werden.

Z.B. `<ref type="quelle" target="frankfurter_anstand_einleitung.xml#frankfurter_anstand_anchor1">`Verweis auf erste anchor-id in der Einleitung zum Frankfurter Anstand.

Fußnoten

Fußnoten - allgemein

Jede Anmerkung bekommt das Element

```
<note type="annotation">Anmerkungstext</note>
```

Soll im Verlaufe der Kommentierung eine `<note>` referenziert werden, dann bekommt die zu referenzierende `<note>` ein `@xml:id`. Die Benennung der `@xml:id` steht völlig frei, sollte aber (meiner Meinung nach halbwegs) logisch sein, z.B.

```
<note type="annotation"
xml:id="001">Anmerkungstext</note>
```

D.h. im Text gibt es Anmerkungen mit @xml:id, wenn diese auch referenziert werden. Alle anderen bleiben einfach unberücksichtigt.

Anschließend kann auf diese Anmerkung folgendermaßen verwiesen werden:

```
<note type="annotation">Vgl. hierzu <ref
target="#001">Anm. Nr#</ref></note>
```

Fußnoten in der Druckfassung

Sollen Fußnoten o.ä. in der Druckfassung anders ausgegeben werden, dann erfolgt die Differenzierung ebenfalls durch das @subtype.

```
<ref target="#notel" subtype="print">(wie Anm.
Nr.#)</ref>
```

Textkritischer Apparat

Bei einfachen Fällen sieht die Kodierung des textkritischen Apparates folgendermaßen aus:

```
Er hat viel getrunken. <note type="crit_app">In
Handschrift B: gegessen</note>
```

Bei größeren Textpassagen wird folgendermaßen kodiert:

```
Er hat <seg type="crit_app">viel getrunken</seg>.
<note type="crit_app">In Handschrift B: wenig
gegessen</note>
```

In komplexen Fällen wird folgendermaßen kodiert:

```
Er hat <milestone xml:id="crit1"/>viel
getrunken<milestone xml:id="crit2"/>. <span
type="crit_app" from="#crit1" to="#crit2">In
Handschrift B: wenig gegessen</span>
```

Der Vorteil dieser Kodierung ist, dass sich in größeren Textpassagen beliebig viele <milestone>-Elemente unterbringen lassen, die durch -Elemente und @from und @to gezielt angesprochen werden können. Verschiedene Lesarten ließen sich so exakt zuordnen.

Bibliographische Angaben innerhalb von können nicht mit <bibl> angegeben werden. Stattdessen ist die (derzeitige) Lösung mit <milestone> zu verwenden:

```
Er hat <milestone xml:id="crit1"/>viel
getrunken<milestone xml:id="crit2"/>. <span
type="crit_app" from="#crit1" to="#crit2">Bei <rs
type="bibl"><ref target="#mueller_titel">Müller:
Titel</ref>, S. 34</rs>: wenig gegessen</span>
```



Achtung

Das Element <milestone> benötigt ein @unit. Je nach Bedeutung des <milestone> sollte man einen der folgenden Werte aus der TEI-Übersicht benutzen:

page -> physical page breaks (synonymous with the **pb** element).

column -> column breaks.

line -> line breaks (synonymous with the **lb** element).

book -> any units termed book, liber, etc.

poem -> individual poems in a collection.

canto -> cantos or other major sections of a poem.

speaker -> changes of speaker or narrator.

stanza -> stanzas within a poem, book, or canto.

act -> acts within a play.

scene -> scenes within a play or act.

section -> sections of any kind.

absent -> passages not present in the reference edition.

unnumbered -> passages present in the text, but not to be included as part of the reference.

Auszeichnung

Kurztitel

Kurztitel lassen sich im Literaturverzeichnis folgendermaßen auszeichnen:

```
<bibl>
  <title type="short">ZGO</title>
  <title type="main">Zeitschrift für die Geschichte
des Oberrheins</title>
  [...]
</bibl>
```

Kodierung von Sonderzeichen/Graphem- und Phonemwechsel

Neue Kodierungspraxis



Geschützte Leerzeichen

Geschützte Leerzeichen bitte mit dem hexadezimalen Unicode-Zeichen kodieren. Also z.B. Friedrich II.

Bei Graphem- und Phonemwechsel wird das **gesamte Wort**, das das bzw. die betroffenen Zeichen enthält, in den zwei verschiedenen Schreibungen kodiert. Das Wort in diplomatischer Schreibweise (Beispiel: dier) wird eingefärbt und mithilfe der **Codevorlage <w>** transformiert. Das folgende Beispiel veranschaulicht das Ergebnis:

```
<w n="dier">dier</w>
```

Anschließend wird händisch die Schreibweise des mit <w> getaggtten Wortes modernisiert:

```
<w n="dier">diser</w>
```

Dasselbe gilt auch bei der Verwendung von Unicodes, z. B.:

```
<w n="Ra&#x0364;tt">Rätt</w>
```

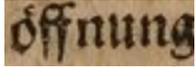
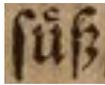
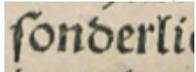
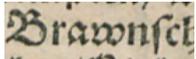
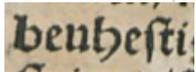
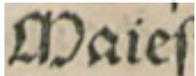
Ausnahme:

Ist das zu modernisierende Wort in der Editionsvorlage **getrennt**, darf das oben beschriebene Verfahren nicht angewendet werden. Stattdessen muss händisch wie folgt getaggt werden:

```
<choice><orig><w>Friedens  
<lb/>schlu</w></orig><reg><w>Friedens<lb/>schluss</w></reg></choice>
```

Alte Kodierungspraxis

Sonderzeichen sollten wie folgt kodiert werden:

Graphem	Phonem	Unicode	Transkription in XML	Bildbeispiel
o	ö	0364	(ö)	
u	ü	0364	(ü)	
c.	etc.	A75B	(etc.)	
ů	u		#u#	
	s	017F		
ᷣ	r		(r)	
v	u		(u)	
w	u		[u]	
w	uu		[uu]	
u	v		(v)	
	un		(un)	
j	i		(i)	
i	j		(j)	

	en		[en]	
	ém		[ém]	> im tschechischen Druck
	em		[em]	
	en		(en)	
n̄	nn		(nn)	
ñ	nd		(nd)	u(nd)
m	mm		(mm)	
ã	am		(am)	gehor(am)
ã	an		(an)	L(an)dchafft
	am		[am]	J[am]merlijcken
	an		[an]	v[an]de
aͤ	ä		(ä)	
õ	on		(on)	v(on)
	on		[on]	
	as		(as)	d(as), w(as)
m	mb		(mb)	wideru(mb)

Mit den dazugehörigen Großbuchstaben wird analog dazu verfahren.

Zeichensetzung

Neue Kodierungspraxis

Satzzeichen, die im modernisierten Text **ersatzlos wegfallen** sollen, werden (einschließlich der Leerzeichen, die wegfallen sollen) mit <orig> getaggt:

```
Ein<orig> /</orig> Friedensbringer
```

Vom Editor **ergänzte** Satzzeichen, die nur im modernisierten Text erscheinen sollen, werden mit <supplied> getaggt:

```
Mietten<supplied>,</supplied> gaben
```

Satzzeichen, die im modernisierten Text **ersetzt** werden sollen, werden mitsamt zu ersetzender Leerzeichen markiert, unter Verwendung der w-Codevorlage transformiert und dann innerhalb von <w> modernisiert (analog zu Wörtern, in denen Buchstaben zu modernisieren sind):

```
Ein Friede<w n=" /">,</w> der
```

Alte Kodierungspraxis



Durch die Verwechslungsmöglichkeit mit dem Backslash muss bei der Kodierung

darauf geachtet werden, dass unmittelbar vor und nach der Virgel ein Spatium steht, unabhängig von der Textvorlage. Ausnahme: unmittelbar anschließende <note>

Transkription im XML-Text	Anmerkung
/ oder /<note	Virgel soll in modernisierter Fassung durch Komma ersetzt werden
(/)	Virgel soll in modernisierter Fassung durch Doppelpunkt ersetzt werden
(/)	Virgel soll in modernisierter Fassung durch Strichpunkt ersetzt werden
(/)	Virgel soll in modernisierter Fassung ersatzlos wegfallen
(.)	Punkt soll in modernisierter Fassung durch Komma ersetzt werden
(.)	Punkt soll in modernisierter Fassung durch Strichpunkt ersetzt werden
(:)	Doppelpunkt soll in modernisierter Fassung durch Punkt ersetzt werden
(:)	Doppelpunkt soll in modernisierter Fassung durch Komma ersetzt werden
(:)	Doppelpunkt soll in modernisierter Fassung durch Strichpunkt ersetzt werden
(.)	Punkt entfällt in modernisierter Fassung
(,)	Komma entfällt in modernisierter Fassung
[,]	zusätzliches Komma in der modernisierten Fassung
[:]	zusätzlicher Strichpunkt in der modernisierten Fassung
[.]	zusätzlicher Punkt in der modernisierten Fassung
· oder (.) oder (·.)	"Mittelpunkt" wird durch Punkt in der modernisierten Fassung ersetzt.
(·) oder (-)	"Mittelpunkt" entfällt.
(,) oder (·,)	"Mittelpunkt" wird durch Komma in der modernisierten Fassung ersetzt.

Auflösung von Abkürzungen

Abkürzungen werden mittels <ex>-Element aufgelöst

```
Röm<ex>isch</ex> Kay<ex>serliche</ex>  
Maj<ex>estät</ex>
```

Anzeige im

a) diplomatischen Druck: Röm Kay Maj

b) modernisierten Druck: Röm[isch] Kay[serliche] Maj[estät]

Abkürzungszeichen müssen mit <am> ausgezeichnet werden, damit sie je nach Ausgabe angezeigt oder nicht angezeigt werden können:

```
Röm<am> .</am><ex>isch</ex>  
Kay<am> .</am><ex>serliche</ex>  
Maj<am> .</am><ex>estät</ex>
```

Anzeige im

- a) diplomatischen Druck: Röm. Kay. Maj.
- b) modernisierten Druck: Röm[jisch] Kay[serliche] Maj[estät]

Abkürzungen innerhalb von Wörtern können ebenso aufgelöst werden:

```
M<ex>ajes</ex>t<am> .</am><ex>ät</ex>
```

Anzeige im

- a) diplomatischen Druck: Mt.
- b) modernisierten Druck: M[ajes]t[ät]

Kodierung unterschiedlicher Absätze oder Sinneinheiten

In Fällen, in denen Absätze innerhalb eines Satzes auftauchen, werden diese Absätze mit `<p style="in_sentence">` kodiert.

Vom Editor ergänzte Absätze, die nur in der modernisierten Fassung angezeigt werden sollen, werden mit `<p resp="XY">` kodiert. Als Attributwert werden die Initialen des jeweiligen Editors verwendet.

Kodierung ausgelassener Textteile

Zum Teil kann es vorkommen, dass einzelne Textpassagen des ursprünglichen Textes nicht in die Edition aufgenommen werden (z.B. Augsburgsburger Religionsfrieden).

Diese Textstellen werden wie folgt ausgezeichnet:

```
<app>  
  <lem>  
    <gap reason="irrelevant" />  
  </lem>  
  <rdg>  
    Ausgelassener Text  
  </rdg>  
</app>
```

Der ausgelassene Text wird in `<rdg>` als Lesart ausgezeichnet und in der html-Ausgabe nicht angegeben. Das `<gap/>` als maßgebliche Lesart des editierten Textes (nämlich die Textauslassung) erzeugt in der Ausgabe ein Auslassungszeichen, z.B. eckige Klammer mit drei Punkten, das wiederum anklickbar ist. Nach dem Anklicken klappt der ausgelassene Text als eigener Absatz unterhalb des Auslassungszeichens auf.

Hervorhebungen im Druck

Hervorhebungen in der Druckvorlage wie Kursivsetzungen, Sperrungen und sonstige Hervorhebungen werden in der diplomatischen Fassung des Textes wiedergegeben, nicht aber in der modernisierten Fassung. In diesen Fällen muss die w-Kodierung durch choice-Kodierung ersetzt werden.

Bei einer Kursivsetzung sähe die Kodierung dann bsp.weise so aus: `<choice><orig><hi rend="italic">Roͤmichen</hi></orig><reg>Römischen</reg></choice>`

Wechsel zwischen Fraktur und Antiqua, wie sie häufig bei Mischtexten in Vernakularsprache und Latein stehen, oder auch Wechsel verschiedenen Fraktur/Antiqua-Schriften sollen ebenfalls in der diplomatischen Fassung des Textes wiedergegeben werden, aber nicht in der modernisierten Fassung.

Die Kodierung sieht dann wie folgt aus:

```
<choice><orig><hi rend="antiqua">Ad mandatum domini</hi></orig><reg>Ad mandatum domini</reg></choice>
```

In Einzelfällen, in denen der Schriftartwechsel eine besondere inhaltliche Hervorhebung markiert, sollte auch in der modernisierten Fassung des Textes dieser kodiert werden.

Hier ist die korrekte Auszeichnung:

```
<hi rend="antiqua">...</hi>
```

```
<hi rend="fraktur">...</hi>
```

Probleme/Hinweise

Bei der Kodierung der Texte kommt es immer wieder zu Darstellungsproblemen durch die Verwendung der Funktion "Automatisch formatieren".



Dadurch entstehen im Text ungewollt Zeilenumbrüche, die in der Ausgabe Leerzeichen produzieren, die sich nicht automatisch über ein Transformationskript abfangen lassen, weil sie nicht im Text selbst kodiert sind, sondern willkürlich durch Oxygen hinzugefügt werden. (Bild einfügen)

Lösungsansätze:

1. Verwendung der Oxygen-Funktion unterlassen.
2. Bei der Verwendung der Auto-Formatierung den Zeilenumbruch vor bestimmten Elementen untersagen (Einstellungen/Formatierung)

Zu 2. Um die Verwendung der gleichen Einstellungen von allen Teammitgliedern zu gewährleisten, sollte ein Projekt erstellt werden, das dann alle mit der Option "Globale (oder lokale) Einstellungen verwenden" sollen.

Unleserliches Digitalisat

Wenn in dem für die Edition verwendeten Digitalisat Stellen unleserlich sind, können die Buchstaben stillschweigend ergänzt werden, wenn durch Autopsie des Drucks, von dem das Digitalisat gefertigt wurde, oder durch ein Digitalisat von einem anderen Exemplar der Buchstabenbestand geprüft wurde.

Zusätzlich muss in der Einleitung unter "Textvorlage" darauf hingewiesen werden, woher die Ergänzungen der im Digitalisat nicht lesbaren Buchstaben stammen (Autopsie bzw. Digitalisat von anderem Exemplar). Wurde ein anderes Exemplar herangezogen, ist dieses unter Angabe der Signatur zu benennen.