



XML – Die Lingua Franca der Digital Humanities

Ursula Welsch
29. bis 31. Juli 2015



Agenda

Teil 1: Was ist XML und wie funktioniert es?

- XML als Markupsprache
- XML-Anwendung

Teil 2: XML in den Digital Humanities

- Editionen mit TEI P5
- Wörterbücher

Teil 3: Arbeiten mit XML

- Workflows
- Software
- Qualitätssicherung



Teil 1: Was ist XML und wie funktioniert es?



XML ist ...

- software- und plattformunabhängiges Datenformat
- Datenaustauschformat bzw. Schnittstellenformat
- Markup- oder Auszeichnungssprache für elektronische Manuskripte
- Modellierungssprache für digitale Datenbestände



Typen von XML-Dokumenten

- Daten-zentriert
- Dokumenten-zentriert



XML ist ...

eine Markupsprache, die es erlaubt,

- die Inhalte eines Dokuments
- mit einem eigenen Vokabular
auszuzeichnen

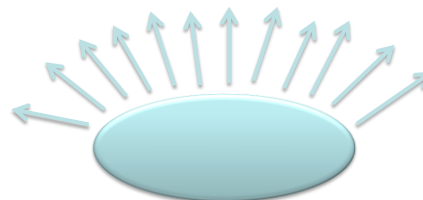
Markup-Sprache

Markup	Auszeichnung von (Papier-)Manuskripten
Markup-Sprache	Auszeichnung von Textdateien
~ proprietär	Auszeichnungen, die nur von bestimmten Softwares verstanden werden
~ neutral	textliches Markup
presentational	binäre Software-Codes, z.B. Word, InDesign
procedural	textliches Markup für die direkte Umsetzung durch bestimmte Softwares, z.B. TeX, LaTeX, RTF, MIF, PDF, Postscript
descriptive (logical, conceptual)	textliche Strukturbeschreibung des Dokuments – bis hin zu semantischen Auszeichnungen, z.B. XML, HTML

XML als Markup-Sprache

Wichtigste Vorteile:

- problemloser Dokumentenaustausch zwischen verschiedenen Rechnern und Softwares
- Explizite Kennzeichnung der Struktur eines Dokuments
- Beschreibung des Inhalts bis hin zur semantischen Benennung
- Medienneutrale Aufbereitung, da sie in jede medienspezifische Form umgesetzt werden können
- Investitionsschutz, da Daten längerfristig gelesen werden können





Einsatz von XML für Publikationen

- in der Buchproduktion für
 - Inhaltliche Metadaten:
d.h. Verlinkung, Kennzeichnung von Personen etc., Register
 - Crossmediale Produktion:
d.h. mehrere Ausgabeformen aus einer Substanz
 - Automatisierung:
d.h. verkürzte Produktionszeiten = geringstmöglicher Zeitraum zwischen letzter inhaltlicher Korrektur und Erscheinen



Einsatz von XML für Publikationen

- in der Archivierung
 - Nachhaltigkeit im Sinn der
Nachnutzbarkeit:
d.h. sich selbst beschreibendes Format, neutrales Format
 - Langzeitarchivierung:
d.h. neutrales Format unabhängig von Software und Plattform, selbst beschreibendes Format, offener Standard

Einsatz von XML für Publikationen

Körperschaften

Signatur

Ausgewählte Inkunabeln der Dombibliothek Freising und der Bibliothek des Metropolitankapitels München

Katalog

Die Sammlungen der Dombibliothek Freising und der Bibliothek des Metropolitankapitels München hüten in ihren Beständen insgesamt 285 Inkunabeln. Bei der Neukatalogisierung bzw. Einbringung der Inkunabelbestände in den Bibliotheksverbund Bayern wurde festgestellt, dass 18 Ausgaben bisher noch nicht in den Beständen der Bayerischen Staatsbibliothek nachgewiesen sind. Von einer Inkunabel (J 115) wird in Freising das einzige bekannte Exemplar aufbewahrt.

Dieser besondere Teilbestand wird hier in Form von Volligitalisaten bereitgestellt und durch kurze Texte beschrieben.



Ort

Zur Geschichte der Dombibliothek Freising

Schon seit der Gründung des Bistums Freising 739 bestand auf dem Freisinger Domberg eine Bibliothek. Ein planmäßiger Aufbau von Bibliothek und Schreibschule begann jedoch erst unter Bischof Arbeo (764-784). Den Höhepunkt ihrer Produktivität erreichte die Schreibschule unter den Bischöfen Hitto (811-835) und Erchanbert (836-854). Insgesamt sind aus der Dombibliothek über 650 Codices mit antiken und mittelalterlichen Texten überliefert. Besonders bedeutend sind die "Freisinger Denkmäler", die ältesten Texte in altslawischer Sprache, enthalten in einem Handbuch für die Slawenmission, das unter Bischof Abraham (957-994) in Freising

Aus J 107: Exlibris Fürstbischof Johann Franz Eckher von Kapfing (1696)

Personen

Markupsprache XML

Drei Aspekte:

1. Plattformunabhängigkeit:
durch textliche Codierung
2. Flexible Nutzbarkeit:
durch frei benennbare Codierung
3. Automatisierbare crossmediale
Produktion:
durch Trennung von Inhalt und Form

Markupsprache XML

1. Plattformunabhängigkeit

- keine software-spezifischen (proprietären) Codierungen
- sondern textliche Zeichenfolgen, die vom übrigen Text durch sog. Delimiter getrennt werden

Markupsprache XML

Beispiel:

```
<H1>Giotto di Bondone</H1>
```

```
<P>&asterisk; 1266 in Vespignano  
&uuml;ber Florenz</P>
```

```
<P>&dagger; 1337 in Florenz</P>
```

```
<P>Schon seine Zeitgenossen feierten  
Giotto als den Neuerer der  
italienischen Malerei.</P>
```

Markupsprache XML

2. Flexible Nutzbarkeit

- selbst definierbare Codierungen
(bei Eigenentwicklung einer DTD)
- beschreibende bzw. inhalts-
orientierte Benennungen empfohlen
- als sog. „generic coding“,
d.h. im Text wiederkehrende strukturelle
Muster formaler oder inhaltlicher Natur

Markupsprache XML

Beispiel 1: semantische Benennung

```
<maler>
<name>Giotto di Bondone</name>
<lebensdaten>
  <geb-jahr>1266</geb-jahr>
  <geb-ort>Vespignano</geb-ort>
  <tod-jahr>1337</tod-jahr>
  <tod-ort>Florenz</tod-ort>
</lebensdaten>
<biografie>Schon seine Zeitgenossen feierten
  Giotto als den Neuerer der italienischen
  Malerei.</biografie>
</maler>
```



Markupsprache XML

Beispiel 1: semantische Benennung

Lexikon :

Giotto di Bondone

** 1266 in Vespignano, † 1337 in Florenz*

Schon seine Zeitgenossen feierten Giotto als den Neuerer der italienischen Malerei.

Register:

Giotto di Bondone (1266-1337)



Markupsprache XML

Beispiel 2: medienneutrale Verlinkung

Printausgabe:

Dies sind wichtige Ausführungen (siehe auch Seite 33),
und hier ist noch etwas Wichtigeres (siehe Kap. 3.3).

Elektronische Ausgabe:

Dies sind wichtige [Ausführungen](#),
und hier ist noch etwas [Wichtigeres](#).

Markupsprache XML

Beispiel 2: medienneutrale Verlinkung

```
<absatz>Dies sind wichtige  
<verweis art="siehe auch" typ="Seite"  
  idref="1234">  
Ausführungen</verweis>, und hier ist  
  noch etwas  
<verweis art="siehe" typ="Kap"  
  idref="1234">  
Wichtigeres</verweis>.</absatz>
```

Markupsprache XML

3. Automatisierbare crossmediale Produktion

durch Trennung von Struktur und Form

- d.h. keinerlei Format- bzw. Darstellungsinformationen im Dokument
- sondern nur der eigentliche Inhalt und die Strukturinformationen
- Layoutbeschreibung extern bzw. in der benutzten Software

Markupsprache XML

Trennung von Struktur und Form

konventionelles
Vorgehen



alle drei Komponenten
untrennbar miteinander
verbunden

➤ Ergebnis ist eine einzige
Ausgabeform

inhaltsorientier-
tes Vorgehen



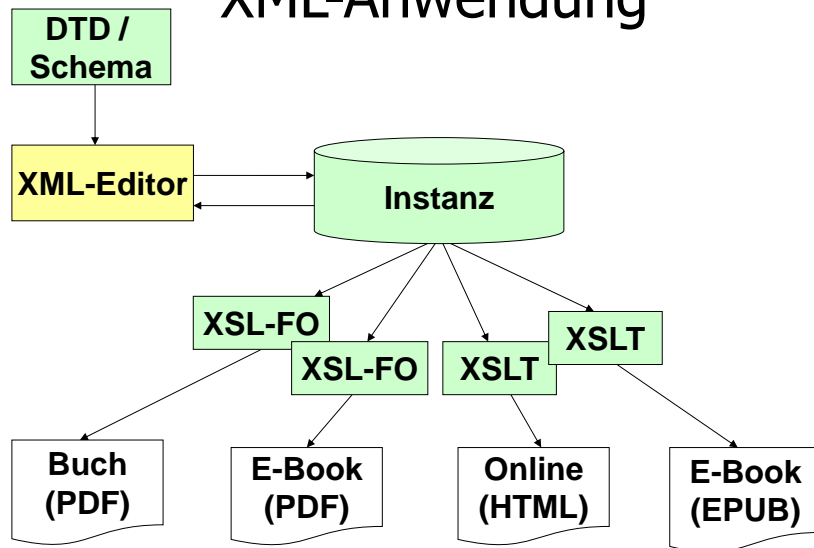
nur Inhalt und Struktur
miteinander verbunden,
Design gesondert abgelegt

➤ Ergebnis sind viele
verschieden gestaltete
Ausgabeformen

XML-Anwendung

- **Strukturbeschreibung:**
DTD (Dokumenttyp-Definition) oder
XML-Schema
- **Instanz:**
als XML-Datenbestand
- **Datenbearbeitung:**
mit XML-Editor
- **verschiedene Ausgabeformen:**
XSL-Stylesheets für Buch, Online, E-Books

XML-Anwendung



Datengrundlage XML

DTD für Rezepte

```
<!ELEMENT kochbuch (rezept+)>
<!ELEMENT rezept (titel, zutaten, zubereitung,
beilagen?)>
<!ELEMENT titel (#PCDATA)>
<!ELEMENT zutaten (zutat+)>
<!ELEMENT zutat (#PCDATA | verweis)*>
<!ATTLIST zutat menge CDATA #IMPLIED
             einheit CDATA #IMPLIED >
<!ELEMENT zubereitung (arbeitsschritt+)>
<!ELEMENT arbeitsschritt (#PCDATA | verweis)*>
<!ELEMENT beilagen (#PCDATA | verweis)*>
<!ELEMENT verweis (#PCDATA)>
<!ATTLIST verweis ziel CDATA #REQUIRED>
```



Datengrundlage XML XML-Instanz (Rezepte)

```
<kochbuch>
  <rezept id="laiberl">
    <titel>Laiberl von Lachsforelle und Hecht</titel>
    <zutaten>
      <zutat menge="300" einheit="g">Hechtfilet</zutat>
      ...
      <zutat menge="1" einheit="Rezept">
        <verweis ziel="pfeffersosse">weiße Pfeffersosse
          </verweis>
      </zutat>
    </zutaten>
    <zubereitung>
      <arbeitsschritt>Hechtfilet würfeln, einsalzen, kalt
stellen, Forellenfilets in 8 Stücke teilen, salzen,
pfeffern und mit Zitronensaft beträufeln.</arbeitsschritt>
      ...
    </zubereitung> </rezept> </kochbuch>
```



Datengrundlage XML XSL-Stylesheet (in HTML)

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet
xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
version="1.0">
<!-- Regel fuer das root-Element -->
<xsl:template match="/">
  <html><body>
    <xsl:apply-templates/>
  </body></html>
</xsl:template>
<!-- Regel fuer das Element 'Rezept' -->
<xsl:template match="rezept">
  <h1 style="background-color:#CCCCFF">
    <xsl:apply-templates select="titel"/>
  </h1>
  ...
</xsl:template>
```

Datengrundlage XML

Darstellung in HTML

Laiberl von Lachsforelle und Hecht

Besonders dekorativ: Klößchen aus heller Hechtmousse mit einer grünen Hülle aus Spinatblättern und einem rosa Kern aus saftigem Lachsforellenfleisch.

Zutaten

Für 4 Portionen

- 300 g Hechtfilet
- 250 g Lachsforellenfilet
- Salz
- Pfeffer aus der Mühle
- Zitronensaft
- 200 g große Spinatblätter
- 1 Ei
- 1/4 l Schlagsahne
- Muskat
- 1 EL geschlagene Sahne
- 1 Rezept **weiße Pfeffersoße**

Zubereitung

1. Hechtfilet würfeln, einsalzen, kalt stellen, Forellenfilets in 8 Stücke teilen, salzen, pfeffern und mit Zitronensaft beträufeln.
2. Spinat putzen. Kurz in reichlich kochendes Wasser geben und sofort kalt abschrecken.
3. Auf Frischhaltefolie 8 gleichgroße, geschlossene Spinatflächen auslegen.
4. Abgetropftes Hechtfleisch mit Ei und Sahne im Mixer pürieren. Durch ein Haarsieb streichen, um restliche Gräten zu entfernen.
5. Mit Muskat, Salz und Pfeffer abschmecken. Geschlagene Sahne unterheben. Die Menge achteln.
6. Jeweils ein Folienteil mit Spinatblättern in einen Schöpflöffel (8 cm Durchmesser) legen. Die

Datengrundlage XML

XSL-FO-Stylesheet (Layout)

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet
  version="1.0"
  xmlns:fo="http://www.w3.org/1999/XSL/Format"
  xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
  <xsl:output method="xml" version="1.0"/>
  <!-- Blattgroessen/Satzspiegel -->
  <xsl:param name="Papierbreite">148mm</xsl:param>
  <xsl:param name="Papierhoehe">210mm</xsl:param>
  <xsl:param name="Seitenraender-rechts">20mm 21mm 14mm 21mm</xsl:param>
  <xsl:param name="Seitenraender-links">20mm 21mm 14mm 21mm</xsl:param>
  <!-- Musterseiten/Seitenvorlagen -->
  <xsl:template match="/">
    <fo:root><fo:layout-master-set>
      <fo:simple-page-master margin="0mm 0mm 0mm 0mm"
        page-height="{ $Papierhoehe }"
        page-width="{ $Papierbreite }"
        master-name="Musterseite.U1">
        <fo:region-body margin="0mm 0mm 0mm 0mm"/>
      </fo:simple-page-master> ... </fo:layout-master-set>
    </fo:root></xsl:template>
```



Datengrundlage XML

Darstellung im Satz

95

Laiberl von Lachsforelle und Hecht

FÜR 4 PORTIONEN
300 G HECHTFILET
250 G LACHSFORELLENFILET
SALZ, PFEFFER AUS DER MÜHLE
ZITRONENSAFT
200 G GROSSE SPINATBLÄTTER
1 EI, ¼ L SCHLAGSAHNE
MUSKAT, 1 ESSL. GESCHLAGENE SAHNE
1 REZEPT BUTTERMILCHSOSSE (SEITE 99)

ZUBEREITUNG

1. Hechtfilet würfeln, einsalzen, kalt stellen. Forellenfilets in 8 Stücke teilen, salzen, pfeffern und mit Zitronensaft beträufeln.
2. Spinat putzen. Kurz in reichlich kochendes Wasser geben und sofort kalt abschrecken.
3. Auf Frischhaltefolie 8 gleichgroße, geschlossene Spinatflächen auslegen.
4. Abgetropftes Hechtfleisch mit Ei und Sahne im Mixer pürieren. Durch ein Haarsieb streichen, um eventuelle Gebeine zu entfernen.

Besonders dekorativ: Klößchen aus heller Hechtmousse mit einer

Gesottene Forellen

FÜR 4 PORTIONEN
4 RÜCHENFERTIGE BACH-
ODER SEEFORELLEN À ETWA 300 G
1 KAROTTE
1 PETERSILIENWURZEL
¼ SELLERIEWURZEL
¼ L WEINSSIG
½ L WEISSWEIN
2 ZWIEBELN, 3 WACHOLDERBEEREN
1 LORBEERBLATT, 5 PFEFFERKÖRNER
1 KNOBLAUCHZEHE, 1 NELKE
SALZ, ZUCKER
PFEFFER AUS DER MÜHLE

ZUBEREITUNG

1. Forellen unter fließendem Wasser innen abspülen. Dabei darauf achten, daß die Schleimschicht auf der Haut nicht weggewaschen wird. Sie gibt dem Fisch beim Garen den blauen Schimmer.
2. Karotte, Petersilienwurzel und Sellerie putzen und in feine Streifen schneiden.
3. In einem großen ovalen Topf (am besten aus Edelstahl oder emailliert) Wasser, Pfefferkörner, Wacholderbeeren, Lorbeerblatt, Knoblauchzehe, Nelke, Karotte, Petersilienwurzel, Sellerie, Weisswein, Weinessig, Zucker, Salz und Pfeffer aus der Mühle kochen lassen.

*Die bayrische
Version der
Forelle blau.*



Datengrundlage XML

Übung am Rechner:

- Software: XML-Editor epcEdit
- Beispieldaten: medwerk_TN



Teil 2: XML in den Digital Humanities



Das Herzstück einer XML- Anwendung ist die Struktur!

- Struktur muss die Inhalte abbilden – nicht das Printwerk
- Semantik so viel wie notwendig für geplante Anwendungen
- Inhaltsmodelle so eng wie möglich für hohe Datenqualität
- Unterstützung von Autor, Redaktion und Produktion



Das Herzstück: das Schema

Das Schema fixiert die Struktur des Inhalts

- in sinnvolle Einheiten gegliedert
- in Vorkommensregeln gebunden

Je nach Strukturlogik ergeben sich verschiedene Dokumenttypen, z.B.

- Monografie
- Lexikon / Wörterbuch
- Edition
- Zeitschrift ...



Auswahl der Struktur

Prüfen der Einsetzbarkeit:

- Welche Grundstruktur braucht mein (Text-)Objekt?
- Welche strukturellen Besonderheiten muss ich abdecken?
- Welche Textauszeichnungen soll es geben?
- Welche Verlinkungen?
- ...



Standard-Schemas

- DocBook:
www.docbook.org
- NLM – National Library of Medicine:
dtd.nlm.nih.gov/3.0/
 - JATS = Journal Article Tag Suite
 - NCBI BTL = National Center of Biotechnology Information Book Tag Library
- TEI – Text Encoding Initiative:
www.tei-c.org



Entstehung

Text Encoding Initiative (TEI)

- erste Ideen seit Ende der 80er Jahre (Vassar College)
- erste Ausgabe der "Guidelines" Mai 1994
- Entwicklung, Nutzung und Pflege bis 2000 als Gemeinschaftsaktivität mehrerer wissenschaftlicher Disziplinen
- seit 2001 TEI Consortium
- Aktuelle Version: TEI P5

In Arbeit:

- SIG Correspondence

Ergänzung

- Music Encoding Initiative (MEI)



Einsetzbarkeit

- Ziel:
alle Arten von wissenschaftlichem Text
- Dokument orientiert:
teiCorpus, teiHeader, text
- Anpassungskonzept:
Einzelmodule, die je nach Bedarf zu einer
Anwendung compiliert werden können (ROMA)
- Ausnahme:
TEI lite
- Einsatz:
heute vor allem in den Digital Humanities
- Lizenz:
Creative Commons Lizenz (CC BY 3.0)



Grundkonzept

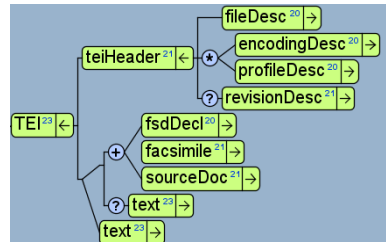
- Sprache:
englisch
- Dokumentation:
ausführlich (1613 Seiten), englisch
- klare Trennung von Metadaten und Text
- Semantische Strukturen:
z.T. in den Elementnamen, z.T. in Attributwerten
- Konfigurierbarkeit:
via Online-Tool ROMA
=> DTD, XML-Schema, Relax NG

Metadaten

Element: `teiHeader`

Teile:

- `fileDesc`: bibliografische Beschreibung
- `encodingDesc`: Besonderheiten der verwendeten Struktur
- `profileDesc`: textliche Beschreibung des enthaltenen Korpus
- `revisionDesc`: Entstehungs- und Bearbeitungsgeschichte der Datei



Metadaten

Beispiel:

- Tag-Nutzung
- Wichtig für Langzeitarchivierung

```
encodingDesc
tagsDecl
namespace name=http://www.tei-c.org/ns/1.0
-- Auswahl: --
tagUsage gi=head @type="editorhead": editorischer Briefkopf tagUsage
tagUsage gi=name @type="from": Absender im Briefkopf, nur als Kind von head tagUsage
tagUsage gi=name @type="to": Adressat im Briefkopf tagUsage
tagUsage gi=opener Datumszeile/Anrede (Autortext) im Briefkopf tagUsage
tagUsage gi=closer Gruß und ggf. Postskriptum (Autortext) tagUsage
tagUsage gi=p @rend: kodiert den Befund des Originals, nicht das Zielformat tagUsage
tagUsage gi=lb Zeilenwechsel im Original; @type="sh": mit Silbentrennung tagUsage
tagUsage gi=note @type="comment": editorischer Kommentar, Erläuterung tagUsage
tagUsage gi=note @type="transcript": interner Transkriptionskommentar tagUsage
tagUsage gi=rs @type="person": Belegstelle für Personenregister tagUsage
tagUsage gi=rs @type="place": Belegstelle für das Ortsregister tagUsage
tagUsage gi=ref @type="letter": Querverweis auf Briefe tagUsage
tagUsage gi=ref @type="sum": Querverweis auf Überblickskommentar tagUsage
tagUsage gi=ref @type="lit": Querverweis auf Literaturverzeichnis tagUsage
tagUsage gi=add Textinzufügung des Editors, z. B. wegen Textverlust des
Lieferungsträgers tagUsage
tagUsage gi=hi Hervorhebung im Original durch Unterstreichung (Autortext) tagUsage
namespace
tagsDecl
encodingDesc
```

Allgemeine Textelemente

Hierarchische Struktur

- text (front, body, back)
- div und div1, div2 etc.

Absatzformate

- Generell: p
- wissenschaftliche, z.B. argument, note, eg
- editionsspezifische, z.B. del, add, rdg
- ansonsten mit dem Attribut @type bei nahezu jedem Element näher spezifizierbar

- <caesura/>
- <calendar>
- <calendarDesc>
- <camera>
- <caption>
- <case>
- <castGroup>
- <castItem>
- <castList>
- <catDesc>
- <catRef/>
- <catchwords>
- <category>
- <cb/>
- <cell>
- <certainty>
- <change>
- <channel>
- <char>
- <charDecl>
- <charName>

Allgemeine Textelemente

Zeichenformate

- eine reiche Palette von Elementen zur linguistischen und interpretatorischen Beschreibung des Texts, z.B. s, seg, hi, emph, foreign, mentioned, term, title etc.

Verweise

- ref, ptr, anchor, xr

Allgemeine Textelemente

Abbildungen

- Figure: gruppiert alles, was zu der Abbildung gehört
- Graphic: ist die Abbildung selbst

```
figure
  graphic url=pullman.png graphic
  head Above: head
  p The drawing room of the Pullman house, the white and gold saloon
  where the magnate delighted in giving receptions for several
  hundred people. p
  figDesc The figure shows an elaborately decorated room, at least
  twenty-five feet side to side and fifty feet long, with ornate
  mouldings and Corinthian columns on the walls, overstuffed
  armchairs and loveseats arranged in several conversational
  groupings, and two large chandeliers. figDesc
figure
```

Allgemeine Textelemente

Tabellen

- sehr simple eigene Struktur
- nur row und cell
- kein Header-Element
- für komplexere Tabellen HTML- oder CALS-Tabellenmodell integrierbar via Namensräume

```
table rend=boxed rows=2 cols=2
  head rend=it Report of the conduct and progress
  of Ernest Pontifex. Upper Vth form — half
  term ending Midsummer 1851 head
  row
  cell role=label Classics cell
  cell Idle listless and unimproving cell
  row
  cell role=label Mathematics cell
  cell ditto cell
  row
  cell role=label Divinity cell
  cell ditto cell
  row
  cell role=label Conduct in house cell
  cell Orderly cell
  row
  cell role=label General conduct cell
  cell Not satisfactory, on
  account of his great unpunctuality and inattention to
  duties cell
```

Bayerische Akademie der Wissenschaften 45

Strukturkonfiguration

TEI Roma: generating validators for the TEI

You are currently working on Test XUG Muc

TEI Roma: generating validators for the TEI

You are currently working on Test

Time to give you a schema

Which format do you prefer?

- RELAX NG schema (compact syntax)
- RELAX NG schema (compact syntax)
- RELAX NG schema (XML syntax)
- ISO Schematron
- Schematron
- W3C schema (in ZIP archive)
- DTD

Generate

Use RELAX NG

www.w3.org/wiki/XML_Schema_Language_comparison

Modules

New Customization Language Sanity Checker

List of TEI Modules

Module name
add analysis
add certainty
add core
add corpus
add dictionaries
add drama
add figures
add gajij
add header
add iso-fs

Character and glyph documentation

The TEI Header

Feature structures

Bayerische Akademie der Wissenschaften 46

Weiterverarbeitung

- Verfügbare Stylesheets, u.a.:

GitHub This repository Search or type a command Explore Feat

TEIC / Stylesheets

TEI XSL Stylesheets

4,994 commits 1 branch 0 releases

branch: master Stylesheets / +

updated LaTeX output files

sebastianrahtz authored 15 hours ago

File	Description
Test	updated LaTeX output files
bibtex	change from use of common2 to common in ant tasks
bin	add html2tei script, and make html2tei code work with HTML not in XH...
common	Merge branch 'lambdalu/hypersetup' of https://github.com/SemanticsOS/...
css	sort out where .css files are
csv	scripts for nlm and csv conversions; tweak docbook
debian-tei-xsl	docx to tei empty w:hyperlinks shouldn't be converted to refs
doc	fix bad link in customisation form, starting new version
docbook	scripts for nlm and csv conversions; tweak docbook
docx	add chance of passing a title with an external link



Fazit zur TEI-Struktur

- Für nahezu alle wissenschaftlichen Textanforderungen gibt es semantische Elemente.
- Weitere Semantisierung über allseits verfügbare Attribute wie z.B. @type und @subtype.
- In der Regel Verzicht auf ganz enge Inhaltsmodelle – jedoch Konfiguration via ROMA möglich.
- Es sind umfangreiche und präzise Erläuterungen im Redaktionshandbuch notwendig, um einen konsistenten Einsatz der XML-Strukturen zu erreichen.



Beispiele

- Nachlass-Editionen
- Urkunden-Editionen
- Zeitschriften
- Lexika und Wörterbücher

Die Aufgabe von Editionen

- Bereitstellung einer zuverlässigen Textgrundlage für jedwede historische oder interpretatorische Betrachtung
 - systematische Sichtung sämtlicher zugehöriger Überlieferungsträger
 - zeichengetreue Transkription der Textzeugen
 - genetische Dokumentation und Textkonstitution
- Verbreitung der Buch- bzw. Online-Publikation, je nach Quellenlage – zu dem Zweck:
 - die Autographen und alten Urkunden zu schonen (z.B. Vermeidung von direktem Hautkontakt oder ungünstigen klimatischen Bedingungen)
 - Lücken im Bestand zu schließen, die z.B. durch Verluste und Aufteilungen in der Folge der beiden Weltkriege des 20. Jhs. entstanden sind

Nachlass-Editionen



Bibliographie des Editionsobjekts

Dazu gehören z.B.:

- Lagerort der Originale (Bibliothek, Nachlass ...)
- genaue Bezeichnung (Titel, Sigle, Aktenzeichen ...)
- bis auf Einzel-Dokumentbasis

Strukturanforderung:

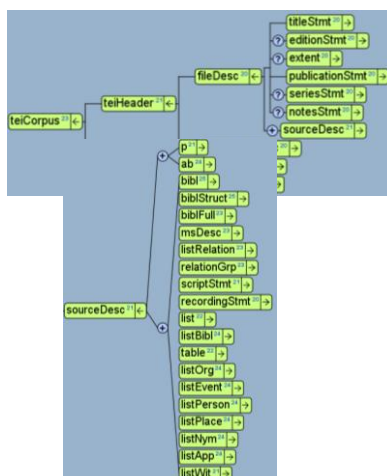
- klare Trennung von Metadaten und Textbeschreibung
- Metadatenbereich mit „Feldern“ für alle notwendigen Angaben

TEI:

- `teiHeader/fileDesc`
 - `titleStmt`
 - `publicationStmt`
 - `sourceDesc`

Bibliographie des Editionsobjekts

Strukturdiagramm:



Strukturbeispiel:



Beschreibung der Materialität

Dazu gehören:

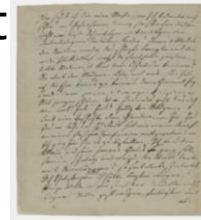
- Beschreibung der äußeren Gestalt
- ggf. aufgeteilt in Blätter, Seiten, Einzelobjekte ...

Strukturanforderung:

- textliche Beschreibung des Textträgers
- ggf. Zeilen-, Seiten-, Spaltenwechsel ...
- ggf. Zuordnung von Sub-Dokumenten

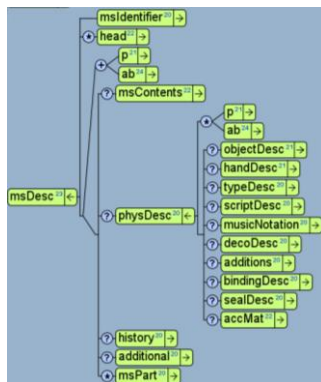
TEI:

- physDesc
- lb, pb, cb
- @rend



Beschreibung der Materialität

Strukturdiagramm:



Strukturbeispiel:



Repräsentation des Textes

Dazu gehören:

- verschiedene Dokumenttypen, z.B. Briefe, Urkunden, Notizbücher, Prosatexte, Gedichte, szenische Texte ...
- mit allen spezifischen Anforderungen an die Aufbereitung

Strukturanforderung:

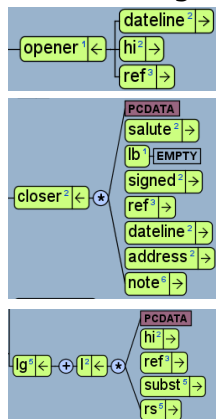
- bei Briefen: z.B. Absender, Anrede, Briefkorpus, Grußformel, Postscriptum ...
- bei Notizbüchern: Art der Notiz, Sub-Dokumente ...
- bei Gedichten: Strophen, Verse ...
- bei Szenischen Texten: Akte, Figuren, Sprechertext, Regieanweisungen ...

TEI:

- Briefe: opener mit dateline, address, closer mit salute
- Gedichte/Szen.Texte: lg/l, role, actor, cast, caesura ...

Repräsentation des Textes

Strukturdiagramm:



Strukturbeispiel:

```
div type="body" >
  pb n="Bl_1_r" <
  opener
    dateline rend="right" > 11.10.21. < dateline
  closer
    salute rend="center" > In vorzüglicher Hochachtung lb <
    Ihr < salute
    lb <
    signed rend="center" > p. P. rs key="PRhausmann type=persName" > R Hausmann < signed
    lb <
    + Details freibleibend!
  < closer
  lg type="stanza"
  l > Du ehrleibiger Gierlappe <
  l > Du gierst und tönst auf pappe <
  l > Dein Schmiergefing zum Ather stinkt. <
  l > Dafür das hi rend="underlined" > Dummen < hi geld im Kasten klingt. <
  l > Ein rothariges wutgeheule <
  l > Geht auf Deine rs key="PUsilbergäule type=Publikation" > Silbergäule < rs <
  l > Du blauer Lüftungreis (hast rote Knochenhaare) <
  l > Tran und rotäugiges Mümmelgreis. <
  < lg
  lg type="stanza"
```



Kennzeichnung von Metainformationen

Dazu gehören z.B.:

- Personennamen
- Datumsangaben
- Ortsangaben

Strukturanforderung:

- bei Personennamen: Vor- und Nachname, Namensteile
- bei Datumsangaben: Tag-Monat-Jahr, gesichert/ungesichert
- bei Ortsangaben: geografische bzw. politische Zuordnung, historischer Kontext, Klassifizierung z.B. als Stadt, Land ...

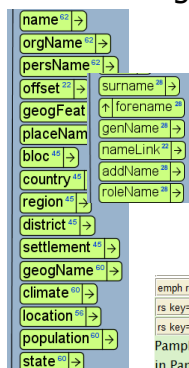
TEI (verschiedene Möglichkeiten):

- rs type=person vs. persName/forename+surname, role
- rs type=date vs. date/@when/@from/@to/@precision
- rs type=place vs. placeName, geogName, region, state



Kennzeichnung von Metainformationen

Strukturdiagramm:



Strukturbeispiel:

```

titleStm n=1 > Nummer des Briefs in der Publikation <
title > An persName forename Ludwig forename surname Ysenburg von Buri surname persName
placeName type=Absenderort > Frankfurt a. M. < placeName ,
-> Datum <
date when=1764-05-23 type=Briefdatum > 23. Mai 1764 < date . date type=Wochentag > Mittwoch < date
-> Zeitraum <
date from=1765-12-06 to=1765-12-10 precision=low type=Briefdatum > 6.-10? Dezember 1765 < date .
date precision=low type=Wochentag > Freitag-Dienstag? < date
-> < placeName precision=low type=Empfängerort > Neuhof < placeName > < title < titleStm!

emph rend=Lemma > rs key=PRBecker type=persName Abu Becker < rs < emph 1921 annoncierte der rs key=INsteegemann type=Institution
rs key=PRsteegemann type=persName Paul Steegemann < rs < Verlag < rs < die Herausgabe eines Buches von
rs key=PUbeckerStinnes type=Publikation Abu Becker: Gott Stinnes! Die Philosophie eines Emporkömmlings. Ein ethisches
Pamphlet < rs . Dieses erschien jedoch nicht. 1922 erschien stattdessen rs key=PUortnerStinnes type=Publikation Gott Stinnes.
in Pamphlet gegen den vollkommenen Menschen von rs key=PRortner type=persName Eugen Ortner < rs < rs im

```

Textkritik

Dazu gehören:

- Beschreibung der Textvariationen, die am Textträger erkennbar sind, z.B.
 - nachträgliche Korrekturen des handschriftlichen Texts
 - Anstreichungen oder Anmerkungen im gedrucktem Material, ggf. mit verschiedenen Stiften
 - einschließlich der Schreib-Topographie (oben, unten, drüber...)

Strukturanforderung:

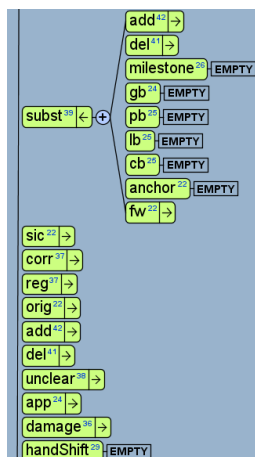
- Streichungen, Ergänzungen, Änderungen
- verschränkte Unterstreichungen
- Angabe der Position

TEI:

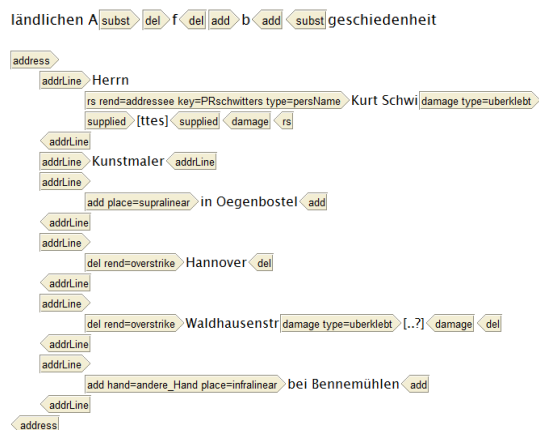
- subst, del, add, restore, unclear, damage ...
- app, rdg, ..., sic, supplied ...

Textkritik

Strukturdiagramm:



Strukturbeispiel:



Überlieferungskritik und Kommentierung

Dazu gehören:

- Kennzeichnung der Stelle im Primärtext
- Kommentartext mit Auszeichnungen, Metainformationen, Verweisen auf andere Textstellen ...

Strukturanforderung:

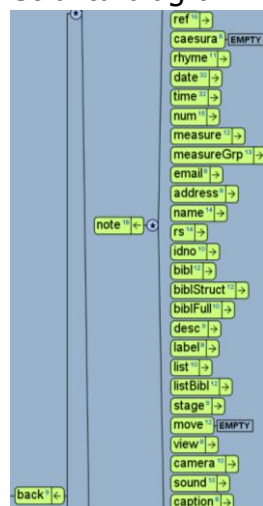
- Textanker:
eindeutige Benennung, eindeutiger Bezug
- Kommentartext:
vielfältige Textauszeichnung, da hier z.B. auch Gedichte, Zitate, Übersetzungen und Hinweise auf andere Kommentare, Primärtextstellen oder Sekundärliteratur ihren Ort haben

TEI (jeweils verschiedene Möglichkeiten):

- Textanker: ref vs. anchor vs. milestone
- Kommentartext: note vs. ref (als Verweisziel)

Überlieferungskritik und Kommentierung

Strukturdiagramm:

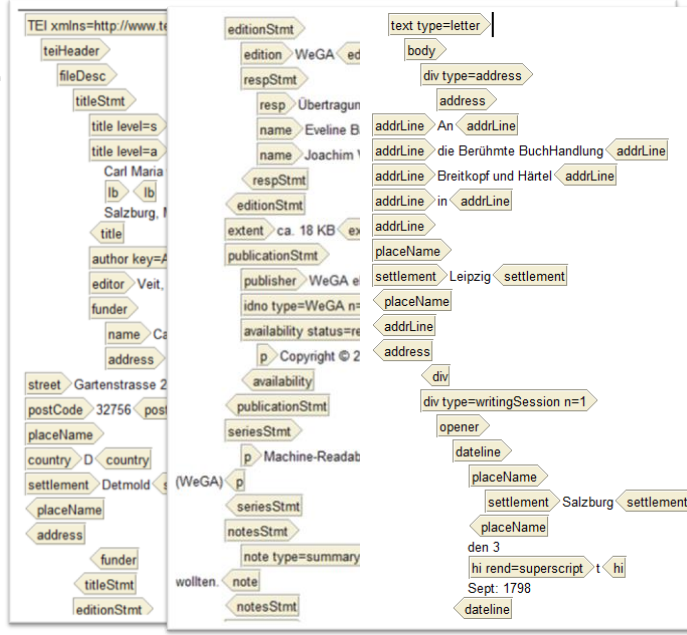


Strukturbeispiel:

```
div type="notes"
note xml:id="FN6" type="footnote">
  emph rend="Lemma" <emph>Geschäftlich!...nicht.</emph> Satz mit rotem Buntstift geschrieben.
  quote <quote>Geschäftlich</quote> auf die folgende nummerierte Aufzählung. <note>
note xml:id="FN7" type="footnote">
  emph rend="Lemma" <emph>Lieber Kurt!</emph> Anrede steht in einem in schwarz gezeichnete
  (angedeutet durch einen Pfeil) aus einem in rot gezeichneten Pistolenschaft (ohne Lau
note xml:id="GB_1_1a" resp="editor" ref target="k0002" type="Lemma" <ref> Wohlgebahrner ... Herr <ref> p
ehrerbietige Tonfall und die altertümelnde Wortwahl verweisen darauf, dass Goe
der Rang Buris und wohl im Allgemeinen auch die Satzungen der <emph rend="Anführung
Gesellschaft <emph> bekannt waren, in denen u. a. festgelegt war, <quote> wie in Sch
Mitglieder u. Vorgesetzte je nachdem sie Prinzen von hohem oder niederem Ade
aus angesehenen Familien usw sind, die Anreden u Briefanschreiben gehalten se
```



Beispiele Briefe: C.M. v. Weber



Beispiel Briefe: C.M. v. Weber

Carl Maria von Weber > Korrespondenz > A040080

Brieftext XML Faksimile

CARL MARIA VON WEBER AN BREITKOPF UND HÄRTEL IN LEIPZIG
SALZBURG, MONTAG, 3. SEPTEMBER 1798

An
die Berühmte BuchHandlung
Breitkopf und Härtel
in
Leipzig

Salzburg den 3^t Sept: 1798

S. D

Wollten Euer Hochedel 100 Stük Exemplare von meinen so eben die Preße verlaßenen Sechs Fugetten gegen gewöhnliche pro Cente in Dero Verlag nehmen, so bitte mich mit umgehender Post geneigetest zu benachrichtigen, auch zugleich mir die Weisung zu machen, auf welche beste Art ich diese 100 Stük, so jedes nur in einem Bogen besteht, am fügichsten übermachen könnte. In Erwartung habe die Ehre zu seyn

Dero
ergebenster
Karl Marie v. Weber

Von Hiesiger Mayerscher BuchHandlung habe dero Adresse erhalten.



Beispiel Briefe: C.M. v. Weber

Carl Maria von Weber > Korrespondenz > A040080

Brieftext XML Faksimile

```

<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:xsl="http://www.w3.org/XML/1998/namespace"
xml:id="A040080">
  <teiHeader>
    <fileDesc>
      <titleStm
      <title level="s">WeGA, Briefe, Digitale Edition</title> <title level="a">
        Carl Maria von Weber an Breitkopf und Härtel in Leipzig<lb/>Salzburg, Montag, 3. September
        1798
      </title> <author key="A002068">Weber, Carl Maria von</author> <editor>Weit, Joachim</editor>
      <funder>
        <name>Carl-Maria-von-Weber-Gesamtausgabe</name> <address>
          <street>Gartenstrasse 20</street> <postCode>32756</postCode> <placeName>
            <country>D</country> <settlement>Detmold</settlement>
          </placeName>
        </address>
      </funder>
    </titleStm> <editionStm
    <edition>WeGA</edition> <respStm
    <resp>Übertragung</resp> <name>Eveline Bartlitz</name> <name>Joachim Weit</name>
    </respStm
    </editionStm> <extent>ca. 18 KB</extent> <publicationStm
    <publisher>WeGA elektronisch</publisher> <idno type="WeGA" n="74"/> <availability
    status="restricted">
      <p>Copyright © 2009, Carl-Maria-von-Weber-Gesamtausgabe, Detmold</p>
    </availability
    </publicationStm> <seriesStm
    <p>Machine-Readable Transcriptions of Texts from the Carl Maria von Weber Complete Edition
    (WeGA)</p>
    </seriesStm> <notesStm
    <note type="summary">fragt an, ob B&H 100 Ex. seiner 6 Fughetten zum Verkauf übernehmen
    wollten.</note>
  
```



Beispiele Briefe: C.M. v. Weber

Vorausgehend keine Angaben gefunden	; den 3 ^{er} Sept: 1798 is <u>Fughetten</u> gegen ost geneigtest zu Stük, so jedes nur yn Dero ergebenster Carl Marie v. Weber	Editorial (einblenden)
Folgend keine Angaben gefunden		Kontext
Personen Weber, Carl Maria von		Absolute Chronologie
Werke 6 Fughetten		Vorausgehend 1798-01-01: An Johann Peter Heuschkel
Orte Leipzig Salzburg		Folgend 1799-01-01: An unbekannt 1799-07-10: Von W. Frein von Marschall genannt Greiff
Rollen keine Angaben gefunden		Korrespondenzstelle
		Vorausgehend keine Angaben gefunden
		Folgend keine Angaben gefunden



Beispiele Briefe: C.M. v. Weber

Carl Maria von Weber > Tagebücher

Korrespondenz | Schriften | Tagebücher | Werke

Chronologie

- ☑ 1810 bis 1813 (675)
- ☑ 1814 bis 1815 (508)
- ☑ 1816 bis 1817 (731)
- ☑ 1818 bis 1819 (709)
- ☑ 1820 bis 1821 (731)
- ☑ 1822 bis 1823 (730)
- ☑ 1826 (154)

Personen

- Weber, Caroline von (1518)
- Benedict, Julius (460)
- Liebich, Johann Karl (380)
- Jungh, Philipp (174)
- Gänsbacher, Johann (170)
- Schlesinger, Adolph Martin (170)
- Brunetti d. Ä., Therese (168) mehr...

Werke

- Der Freischütz, romantische Oper in drei Aufzügen (224)

Montag, 26. Februar 1810
d: 26t Februar von Stuttgart mit OberP:K: Götz abgereist. im Vermögen gehabt 40 f. in Heilbronn übernachtet.
A064400

Dienstag, 27. Februar 1810
d: 27t bis Fürfeld mit Götz gereist, von ihm Abschied genommen, und noch 25 f erhalten. da kein anderes Fuhrwerk zu bekommen war muste ich Extrapost nehmen. PostGeld bis Sinzheim 5 f 15. Chaussée und ...
A064401

Mittwoch, 28. Februar 1810
d: 28t --- ihm zum KapellMstr Ritter, zu Hoffmann, zu Franks gegangen, welche eine große Freude hatten, mich zu sehen, und sogleich für ein Monats Quartier sorgten. Nachmittags noch zu Babo, Weber, ...
A064402

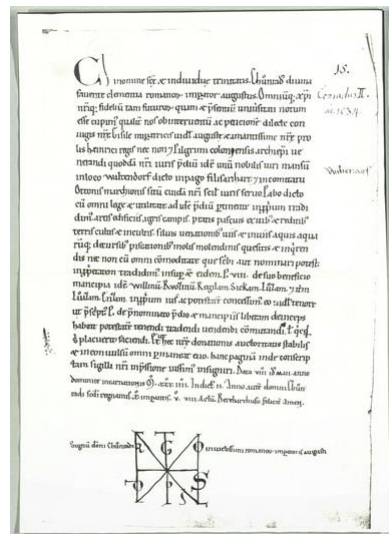
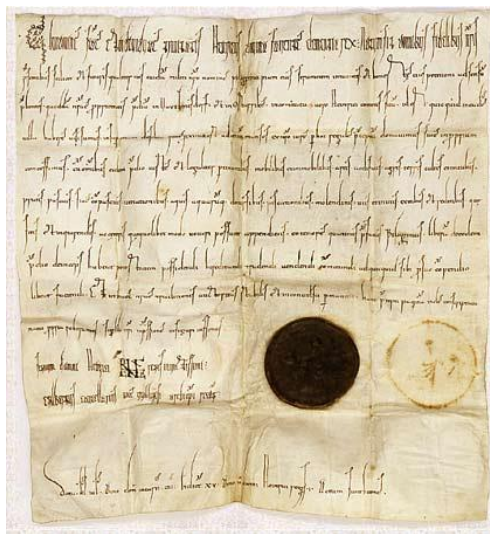
Donnerstag, 1. März 1810
d: 29t11 März mit Hoffman und Berger nach Heidelberg gefahren. da zu H. KapellMst: Hoffmann, H. von Tusch, Heinrich Voß und Michaelis gegangen, auch zufällig Seligmann angetroffen. für Mittagessen 3 f ...
A064403

Freitag, 2. März 1810
d: 2t Kaibel und Maier aufgesucht, und zum erstenmale im Schwarzen Bären zu Mittag geessen 36. im großen Concert gespielt die Var. aus C. Abends 27.
A064404

Samstag, 3. März 1810
d: 3t Mittag und Abendeßen 1. -
A064405



Urkunden-Editionen





Lexicon musicum latinum

- Eintrag: entry
- Bedeutungen: sense

Wörterbuch Vorwort Quellenverzeichnis Lit

F bis fa

F, f, ff, FF, F
diatonicum - F
chromaticum
fa
faburdon
falsa
falseta, falsetum
famulus
falsus
farsa
fauces
faulxbourdon,
faburdon
fenestra
ferenix

F, f, ff
Mensuren
Ausführung
Romanus
Schlüssel
die sechste
designates
II letter th
for perform
Romanian
identifies I
I Marke
that design
measurum
I im Sys
II

body
div type=LML xml:id=M
div type=section xml:id=MF
head F head
entry xml:id=MF00001
form form type=lemma F form, form type=variant xml:id=MF00002 f form,
form type=variant xml:id=MF00003 ff form, form type=variant xml:id=MF00004 FF form form
sense
sense level=1 xml:lang=ger hi rend=italics I. Markierungspunkt in Mensuren II. Tonbuchstabe III. Au-
ührungshinweis (littera significativa, ‚Romanusbuchstabe‘) IV. Schlüsselbuchstabe
V. Kennbuchstabe für die sechste Kirchentonal (< hi sense —
sense level=1 xml:lang=eng -> \im< hi rend=italics I. letter that designates a specific point in
measurements II. letter that designates a note III. direction for performance (littera significativa‚ &
Romanian letter) IV. clef sign V. letter that identifies the sixth mode (< hi sense
sense level=2 n=1   num hi rend=bold I< hi num
sense xml:lang=ger hi rend=italics Markierungspunkt in Mensuren (< hi sense —
sense xml:lang=eng -> \im< hi rend=italics letter that designates a specific point in measure
< hi sense
sense level=4 n=1   num hi rend=bold 1< hi num
sense xml:lang=ger hi rend=italics im System A - LL (< hi sense —
sense xml:lang=eng hi rend=italics in the system A - LL (< hi sense
sense level=5 n=a   num hi rend=bold a< hi num
sense xml:lang=ger hi rend=italics F ür die lichanos hypaton chromaticos und FF ür die
hyperboleon diatonos (< hi sense —



Bayerisches Wörterbuch

(work in progress)

Ache, Ohe

F., fließendes Gewässer, °OB, °NB **verrein.** °Ocha „alle Gewässer, die aus einem See kommen“
Hemhof RO; °gehn wir in d'Ach zum Schwimmen? „eine Ache ist größer als ein Bach“ Bayerisoiien
SOG; °Ohe „jeder Bach“ SchönbG GRA: Und wie d'Ach graben worden ist, sind die Grabenmacher
dran hinkommen Berg SOB (BÖCK Sagen Neuburg-Schrobenhsn 20); So daz Eliases piuoet in erda
kitrifit, so inprimant die perga... oha artrubnest 9 Jh. (SKD 68.50-52 (Muspilli)); das zye
Beclagte an gemeltem Subentage [Sonwendtag] R
hinunter gelaffen 1749 (StA Mch Hofmark Amera
Ohe, häufig in Komposita, z.B. Salzach, Kolbach; se
(M), mit falscher Restitution od. verdunkelt z.B. Aiz

Etm.: Ahd. *ahha*, mhd. *ache*, *stf.*, urverwandt mit lat. *aca*

Ltg. Appellativum $a^{h_2}o$ (SOG), $a^{h_2}o$ (WOR), $a^{h_2}o$ (M) (GRA); im Namenskomp. abgeschwächt zu -*a*.

SCHMELLER I,21f.; WESTENRIEDER Gloss. 4.
WBO I,115f.; Schwab. Wb. 1.88; Schw. Id. 1.63.
°DWB I,1349f.; Frühhd. Wb. 1.535; LEXER HWb.

W-129a/2.

nur eine Bedeutung

artikel
lemma
lemma-position
lemma **Ache** lemma -> ->
lemma-variante °Ohe lemma-variante -> CR
grammatik genus=F. wortart=N
lemma-position ->, (nur wenn keine Bedeutungs-Nr. folgt)
Bedeutung
bedeutung-position
bedeutung fließendes Gewässer bedeutung
>, (stets nach Bedeutung, auch wenn mehrere)
Verbreitung
verbreitung-position
verbreitung-angabe belegtmach1958 häufigkeit=verrein. bezirk=OB
>, (immer zwischen den Angaben)
verbreitung-angabe belegtmach1958 häufigkeit=verrein. bezirk=NB
verbreitung-position
Belege
beleg-position
beleg-angabe
beleg-kontext belegtmach1958 Oche belegtext
beleg-kontext anfuhrungsja -> (immer) -> alle Gewässer, die aus einem See
kommen -> (immer) -> beleg-kontext
beleg-region ort Hemhof ort landkreis RO landkreis beleg-region
beleg-angabe -> (immer am Ende der Belegangabe)
beleg-angabe beleg-text belegtmach1958 gehn wir in d'Ach zum Schwimmen? beleg-text
beleg-kontext anfuhrungsja -> -> eine Ache ist größer als ein Bach ->
beleg-kontext
beleg-region ort Bayerisoiien ort landkreis SOG landkreis beleg-region beleg-angabe
>, (immer am Ende)
beleg-angabe beleg-text belegtmach1958 Ohe belegtext

lemma-position
bedeutung-position
bedeutung
verbreitung-position
beleg-position
erläuterung-position
etymologie-position
lautung-position
literatur-position
komposita-position
lemma-kompositum
bedeutung-position
lautung-position
literatur-position
verfasser



Sprachwörterbuch

Inhaltsgenerierung

```

lemma flex=-o-a compagno lemma
art il art lemma compagno lemma art la art lemma compagna lemma lemma_compj
phon kom'paɲno phon phon_comp phon [CDATA] kom'paɲno [CDATA] phon phon [CDATA] kom'paɲna
gram gen=m pos=n [CDATA] phon phon_comp

```

Inhalteverdichtung

```

lem_gp lemma flex=-o-a compagno lemma
phon kom'paɲno phon
gram gen=m pos=n
lem_gp
trans_gp trans flex=-in Partner trans
gram gen=m pos=n
trans_gp

lem_gp lemma flex=-o-a compagno lemma
phon kom'paɲno phon
gram gen=m pos=n
lem_gp
trans_gp trans flex=-e-in Lebensgefährte trans
gram gen=m pos=n
trans_gp

trans_gp trans_comp art der art trans Partner trans art die art trans Partnerin trans
art der art trans Lebensgefährte trans art die art trans Lebensgefährtin trans trans_comp

```



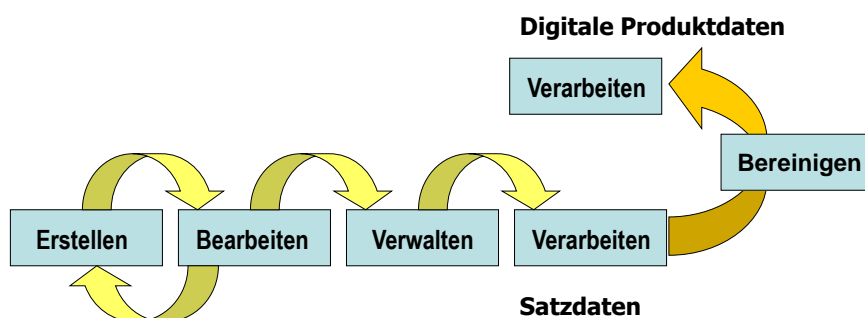
Teil 3: Arbeiten mit XML

XML-Workflow

1. Wann brauche ich XML überhaupt?
 - Wenn das Ergebnis eine hybride Publikation sein soll:
Print und Online.

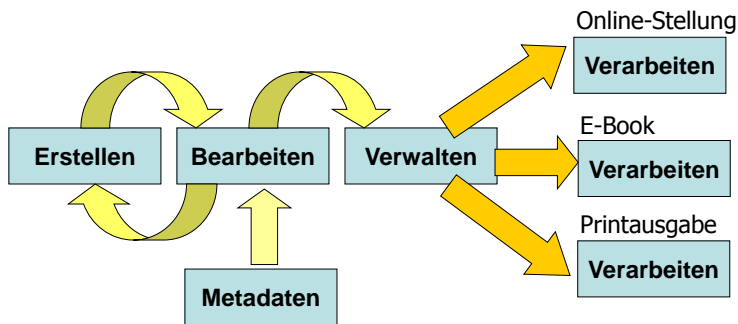
Crossmediales Publizieren

Konventionelle Produktion einer digitalen Ausgabe



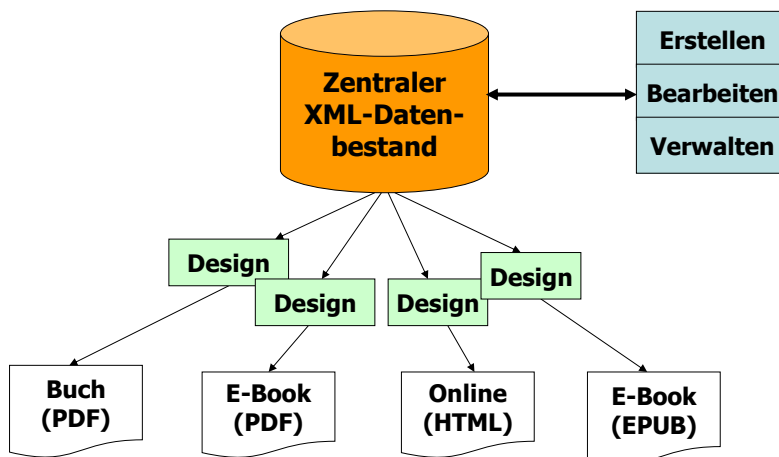
Crossmediales Publizieren

... muss ein crossmedialer Workflow werden!



Crossmediales Publizieren

Single Source Publishing





XML-Workflow

1. Wann brauche ich XML überhaupt?
 - Wenn das Ergebnis eine hybride Publikation sein soll: Print und Online.
- 1.1 Mit oder ohne Verlag für die Publikation?
 - Wenn mit Verlag, dann:
- 1.2 Was liefern wir dem Verlag? (Print-PDF oder XML?)
 - Wenn XML, dann sollten die (Print-)Korrekturen in XML ausgeführt werden!



XML-Workflow

2. An welcher Stelle im Workflow soll XML ins Spiel kommen?
 - Bei sehr kleinteilig auszuzeichnenden Substanzen (Briefeditionen, Wörterbücher etc.) ist XML von Beginn an empfehlenswert!

XML-Workflow

In der Redaktion



XML	XML	---	---	XML
Word	Word	X	XML	XML
Open Office	Open Office	X	XML	XML
TeX	TeX	X	XML	XML
CTE	CTE	X	XML	XML

XML-Workflow

Ausgewählte Inkunabeln der Dombibliothek Freising und der Bibliothek des Metropolitankapitels München

Die Sammlungen der Dombibliothek Freising und der Bibliothek des Metropolitankapitels München hüten in ihren Beständen insgesamt 285 Inkunabeln. Bei der Neukatalogisierung bzw. Einbringung der Inkunabelbestände in den Bibliotheksverbund Bayern wurde festgestellt, dass 18 Ausgaben bisher noch nicht in den Beständen der Bayerischen Staatsbibliothek nachgewiesen sind. Von einer Inkunabel (J 115) wird in Freising das einzige bekannte Exemplar aufbewahrt.

Dieser besondere Teilbestand wird hier in Form von Volidigitalisaten bereitgestellt und durch kurze Texte beschrieben.

Zur Geschichte der Dombibliothek Freising

Schon seit der Gründung des Bistums Freising 739 bestand auf dem Freisinger Domberg eine Bibliothek. Ein planmäßiger Aufbau von Bibliothek und Schreibschule begann jedoch erst unter Bischof Arbeo (764-784). Den Höhepunkt ihrer Produktivität erreichte die Schreibschule unter den Bischöfen Hitto (811-835) und Erchanbert (836-854). Insgesamt sind aus der Dombibliothek über 650 Codices mit antiken und mittelalterlichen Texten überliefert. Besonders bedeutend sind die "Freisinger Denkmäler", die ältesten Texte in altslawischer Sprache, enthalten in einem Handbuch für die Slawenmission, das unter Bischof Abraham (957-994) in Freising



Aus J 107: Exlibris Fürstbischof Johann Franz Eckher von Kapfings (1696)



XML-Workflow

2. An welcher Stelle im Workflow soll XML ins Spiel kommen?

- Bei sehr kleinteilig auszuzeichnenden Substanzen (Briefeditionen, Wörterbücher etc.) ist XML von Beginn an empfehlenswert!

2.1 Wie kann ich die Redaktion ins Boot holen?

- Nur durch überzeugen!

2.2 Welche Maßnahmen stehen zur Verfügung?

- Wenn XML, dann Ausführen der (Print-)Korrekturen in XML!



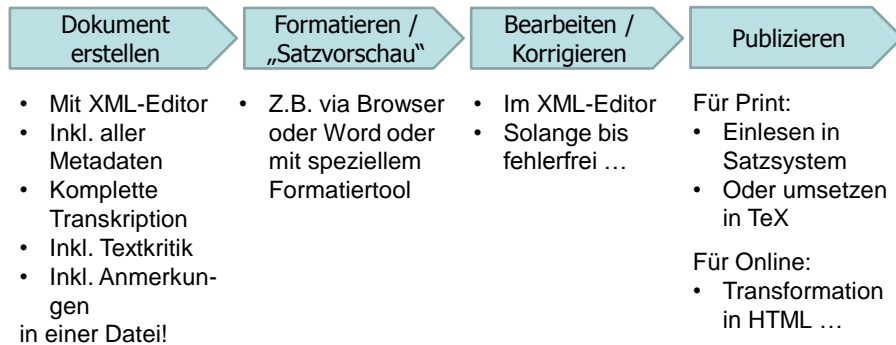
XML-Workflow

3. Mögliche Vorgehensweise:

- Eine inhaltsorientierte Struktur, in der sich der Sprachgebrauch der Redaktion wiederfindet – bis hin zur gemeinsamen Festlegung (und Benennung) der Strukturelemente.
- Eine Arbeitsumgebung, die ähnlich bequem zu benutzen ist wie die gewohnte Arbeitsumgebung.
- Immer wieder gemeinsam Probleme mit der Strukturnutzung oder auch mit der Arbeitsumgebung besprechen und lösen.
- ...

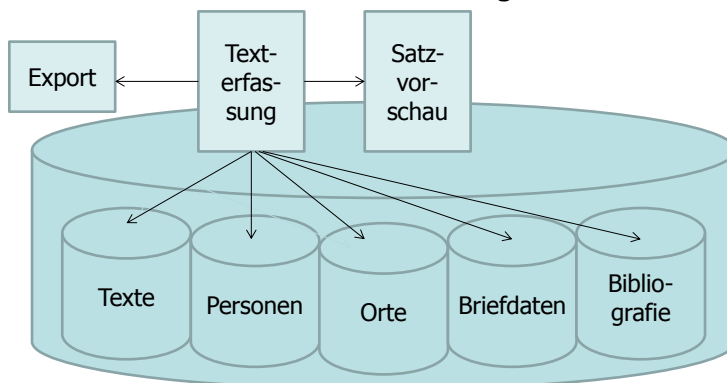
Typen von XML-Workflow

➤ Dokumenten-orientiertes Vorgehen (wie mit MS Word)



Typen von XML-Workflow

➤ Datenbank-orientiertes Vorgehen



XML-Editoren

- XMetaL von Just Systems
 - sehr guter XML-Editor für die redaktionelle Bearbeitung von XML-Dokumenten
 - kostenpflichtig
- Oxygen XML von SynchroSoft
 - sehr guter XML-Editor, der für Entwickler gedacht ist
 - mit einer eigenen Textbearbeitungssicht, die sehr Nutzerfreundlich zu konfigurieren ist
 - kostenpflichtig
- XML-Spy von Altova
 - Qualitätvoller XML- und HTML-Editor mit etlichen Zusatzprogrammen, der vor allem für Entwickler gedacht ist
 - Nicht anpassbar an Nutzerbedürfnisse
 - Kostenpflichtig

XML-Editor: Oxygen

Bayerische Akademie der Wissenschaften 89

XML-Editor: Oxygen

The screenshot shows the Oxygen XML Editor interface. The main window displays an XML document with a postcard layout. A search dialog box titled 'Werk einfügen' is open, showing a list of search results for the term 'Elektra'. The results include 'w0142 Chor aus Elektra von Sophokles' and 'w0612 Christas Herreise'. The background document shows a postcard with the following text:

(Adressangabe:)
 (gedr.: Postkarte |)
 An (Aktion: 1 |) (Titel: |) Herr Otto Singer |
 (gedr.: in (Aktion: 1 |) (Titel: |) München |)
 (gedr.: Wohnung |) (Straße und Hausnummer: |)
 (Aktion: 1 |) (Titel: |) Lieber Freund!
 Was haben Sie jetzt Alles von **Elektra** u. wie wo
 der Partitur meine Ziffern? Bitte um Angabe
 da ich die Ziffern jetzt nachtragen muß. Geh
 Ich werde von Anfang an doppelte Stühen druck
 kleines Orchester!
 Besten Gruß! Ihr
 (Datum: 3. April 08)

Bayerische Akademie der Wissenschaften 90

XML-Editor: Oxygen

The screenshot shows the Oxygen XML Editor interface. The main window displays an XML document with a postcard layout. A metadata editor dialog box titled 'Datum bearbeiten' is open, showing fields for date, author, and location. The background document shows a postcard with the following text:

Beilagen:
 Reproduktionen: (Bilddatei: |) (Quelle: |)
 (Bilddatei: |) (Bilddatei: |)
 Literatur:
 Editorische Besonderheiten dieser Übertragung:
 Klassifizierung:
 Postkarte
 Entstehungsdaten:
 Aktion Nr. 1 (+Verfasser: |) (+Verfasser-Org.: |)
 (Verfasser: |) (Adressat: |) (Datierung: |) (Ort: |) (Zielort: |)
 (Interne) Aufgaben:



XML-Workflow

In der Produktion

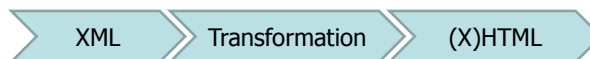


XML	InDesign	Ja	schlecht
XML	3B2/App	Ja	Ja
XML	TeX	Ja	---
XML	XSL-FO	Nein	---



XML-Workflow

Für die Online-Stellung





Qualitätssicherung und Weiterverarbeitung

Teilweise sehr umfangreiche Inhaltsmodelle, z.B.:

- Unterstrukturen in Überschriften: ca. 54 Unterelemente neben Text
- Unterstrukturen in Absätzen: ca. 54 Unterelemente neben Text



Probleme bei der Datenerstellung

editionsübergreifend unterschiedliche Auszeichnungen

ländchen A<subst> f <add> b <add> <subst> geschiedenheit

<app> <lem> eimge </lem>

<rdg> <del type=durchgestrichen> m ein <add type=ueberzeile> ig <add> e </rdg> </app>

<app> <lem> Ungeziefer </lem>

<rdg> Ungezi <add type=inzeile> e <add> fer </rdg> </app>

<note xml:id=FN6 type=footnote>
<emph rend=Lemma> Geschäftlich </emph> nicht.) <emph> Satz mit rotem Buntstift geschrieben, </emph>
<quote> Geschäftlich </quote> auf die folgende nummerierte Aufzählung. </note>

<note xml:id=FN7 type=footnote>
<emph rend=Lemma> Lieber Kurt! </emph> Anrede steht in einem in schwarz gezeichneten (angedeutet durch einen Pfeil) aus einem in rot gezeichneten Pistolenschaft (ohne Lauf </emph>
<note xml:id=GB_1_1a esp=editor> <ref target=k0002 type=Lemma> Wohlgebohrne ... Herr </ref> <ref> p </ref> </emph>
<emph> ehrerbietige Tonfall und die altertümliche Wortwahl verweisen darauf, dass Goet der Rang Buris und wohl im Allgemeinen auch die Satzungen der <emph rend=Anführung> Gesellschaft </emph> bekannt waren, in denen u. a. festgelegt war, <quote> wie in Schr Mitglieder u. Vorgesetzte je nachdem sie Prinzen von hohem oder niederem Adel aus angesehenen Familien usw sind, die Anreden u Briefanschriften gehalten sei



Probleme bei der Datenerstellung

trotz redaktioneller Vorgabeliste keine einheitlichen Attributwerte

```
<!ELEMENT rs (#PCDATA | add | damage | emph | hi | ref | rs | unclear | subst | del | sic)*>
<!ATTLIST rs
  key CDATA #IMPLIED
  rend (ubers | sender | addressee | Pseudonym | Peudonym) #IMPLIED
  type (persname | persName | perName | Zeitung | Zeitschrift | Werk_Lit_KS | Werk_Lit |
  Werk_KS_BK | Werk_BK_KS | Werk_BK | Publikation | Literatur | Lebensdaten |
  Institution | Institutionen | Institution | Institution | Event | Ausstellung) #REQUIRED
```

Vermischung von semantischer und typografischer Auszeichnung – auch englischer und deutscher Benennungen ...

```
<!ATTLIST p
  rend (vertical_left | right | no_lb | left_margin | left | framed | center |
  Tinte_schwarz | Kopierstift | Fraktur) #IMPLIED
```



Qualitätssicherung in der Redaktion

```
div type=body
  pb n=Bl_3_r
  p Ob ich Ihre Vorlesung schweigend g
  geschwiegen noch geduldet! Ich habe gelac
  subst del k del add K add subst unst im Cr
  tue ich nur, wenn ich eine Tageszeitung les
  signed rend=center
  milestone n=1 rend=KS_rot unit=Unterstreich
  signed
  lb
  dateline rend=left Hannover 11.12.21 dateline
  div
  div type=notes
  note xml:id=EN011 type=endnote
  emph rend=Lemma rs key=PRBecker type=persName
  rs key=PRSteegemann type=persName Paul Steeger
  rs key=PUBeckerStinnes type=Publikation Abu Becke
  Pamphlet rs Dies erschien jedoch nicht.
  in Pamphlet gegen den vollkommenen Men
  rs key=INsteegemann type=Institution rs key=PRSteege
  werden, dass rs key=PRBecker type=persName Ab
  nicht klar... (1890 Glaishammer b.
  In rs key=PRSchwitters type=persName KS rs Adre
  rs key=PROrtner type=persName Eugen Ortner rs
  rs key=PROrtner type=persName Ortner rs ohne
  332) und einer an rs key=PRBecker type=persName
```

Primärtext:

```
<Seitenwechsel [Bl_3_r]>
```

Ob ich Ihre Vorlesung schweigend geduldet^(ENxix) habe? Wahrhaftig nicht! Ich habe weder geschwiegen noch geduldet! Ich habe gelacht. Jenes befreiende Lachen gelacht, um des Willen alle [k] K[???] unst im Grund doch nur da – und überhaupt sinnvoll ist. Schweigend dulden tue ich nur, wenn ich eine Tageszeitung lese!

```
[s: Unterstreich] Abu Becker [e1] <Zeilenwechsel >
Hannover 11.12.21
```

Anmerkungen:

(ENxviii) *Abu Becker* 1921 annoncierte der *Paul Steegemann Verlag* die Herausgabe eines Buches von «*Abu Becker: Gott Stinnes! Die Philosophie eines Emporkömmlings. Ein ethisches Pamphlet*». Dieses erschien jedoch nicht. 1922 erschien stattdessen «*Gott Stinnes. Ein Pamphlet gegen den vollkommenen Menschen*» im *Paul Steegemann Verlag*. Es kann also vermutet werden, dass *Abu Becker* ein Pseudonym für *Eugen Ortner*

(Kommentar: wenn es tatsächlich ein Pseudonym wäre, müssten beide den gleichen key bekommen, ist aber nicht klar...)

(1890 Glaishammer b. Nürnberg – 1947 Traunstein), Dramatiker, Bühnenautor, ist in *KS Adressheft Merzgebiet 1a* steht auf S. 30 eine Adresse unter dem Namen *Eugen Ortner* in Leipzig und im *SN VI* ist unter Nr. 943, S. 325 ein Brief an *Ortner* (ohne Vorname), einer an *Abu Becker* (Nr. 1331, S. 332) und einer an *Abbu Becker* (Nr. 1362, S. 333) verzeichnet. Diese Briefe stehen jedoch nur im Register und sind nicht im Entwurf

Qualitätssicherung im XML-Workflow

- Struktur:
 - Strukturführung bei Content-Erstellung
 - Validieren gegen DTD
- Inhalte:
 - Validieren gegen XML-Schema
 - Validieren gegen Schematron Content Rules
 - spezifische Formatierung für typische Fehler
 - Formatierung im Produkt

```
Umschlag:
[[DIV: TTB]]
Cover Hannover: hs. beschr. Briefumschlag eines Briefs von [PRHannemann] Raoul
Hannemann an [PRSchwitters] KS, zugeschnitten und untereinander eingeklebt, 3.1.1920
Maße: 10,5 x 14,5 cm; #P01; #26a

~Seitenwechsel [Cover_Innenseite]~
Empfängeradresse:
Herrn Kunstdrucker [act]
Herrn [PRSchwitters] Kurt [act] #P01; #26b
Hannover
[unleserl. Hand, right: ] [Waldhausener 5] #P01; #26a
de Post wart seine Wohnung wold finden
er ist blind und dof [act]
Absenderadresse:
Falls nicht zu ermitteln,
bitte zurück an [PRHannemann] R. Hannemann
Berlin-Friedrich
Bismarckstr. 16
[Unklar: ?] #P01; #26a
Anmerkungen:
(PN1) Der Briefumschlag wurde in zwei Hälften geschnitten und dann aufgeklebt. Die
Schnittstellen, die meist quer durch die Wörter gehen, sind mit Senkrechten markiert.
(PN3) Kurt; eigentlich: Kurt
(PN4) Waldhausener 5] von anderer Hand
(EN) Der Umschlag gehört vermutlich zu dem Brief von Hannemann Bl. 42 r
(ENa) Waldhausener 5] [PRSchwitters] KS Wohnung in der Waldhausenerstraße 5 in
Hannover wurde offensichtlich von der Post ermittelte.
(ENb) b. [PRHoch] H. [PRHannemann] H. wohnt 1921 bei seiner damaligen Le-
bensgefährtin [PRHoch] Hannah Hoch, eigentlich [PRHoch] Johanna Hoch (Gotha 1889
– 1978 Berlin). Collagekünstlerin, Dadaistin, Mitglied der ' [Novembertagungs] No-
vembergruppe. Sie lebte [PRSchwitters] KS vermutlich 1919 zusammen mit [PRHannemann]
Hannemann in Berlin kennen. (vgl. [Hille, 2000] Hille 2000, S. 175). Seit 1921
pflegte [PRSchwitters] KS und [PRHoch] H. Hoch eine persönliche Freundschaft. Ange-
sichts soll [PRSchwitters] KS für die zweite h in ihrem Vornamen verantwortlich sein,
damit sie wie Anna Blasse von hanten wie von vorne HANNAH sei (vgl. [Rosen, 1999]
Rosen 1999, S. 113f).
```

XML-Syntax

Bestandteile von XML-Dokumenten

- Prolog
- Dokumentinstanz

(Minimal-)XML-Dokument:

```
<?xml version="1.0"?>
```

```
<dokument>text text text</dokument>
```

XML-Syntax: Zustände

- wohlgeformt (well-formed)
(= Instanz ohne DTD)
Die Instanz entspricht den generellen Syntaxregeln.
- valide (valid)
(= Instanz mit DTD)
 - Die Instanz ist wohlgeformt und
 - stimmt mit einer zugehörigen DTD überein
„conforms to a given Document Type Definition“

XML-Syntax

Dokumenttyp-Deklaration

Interne DTD

```
<!DOCTYPE sammlung  
[  
  <!ELEMENT sammlung ...>  
  ...  

```

Externe DTD

```
<!DOCTYPE sammlung SYSTEM "Gedichtsammlung.dtd">
```



XML-Syntax: Deklarationen Dokumenttyp-Definition (intern)

```

Altova XML Spy - [Rilke_Gedichte_mit-interner-DTD.xml]
Datei Bearbeiten Projekt XML DTD/Schema Schema-Design XSL/Query Authentiz Konvertieren Ansicht Browser WSDL SOAP Extras Fenster Hilfe
1 <?xml version="1.0" encoding="UTF-8" standalone="no"?>
2 <!DOCTYPE sammlung
3 [
4 <ELEMENT sammlung (gedicht)+ >
5 <ELEMENT gedicht (autor, titel, undertitel?, (vers+ | strophe+)) >
6 <ELEMENT autor (#PCDATA) >
7 <ELEMENT titel (#PCDATA) >
8 <ELEMENT undertitel (#PCDATA) >
9 <ELEMENT vers (#PCDATA) >
10 <ELEMENT strophe (vers+) >
11 ] >
12
13 <sammlung>
14 <gedicht>
15 <autor>Rainer Maria Rilke</autor>
16 <titel>Der Panther</titel>
17 <undertitel>Im Jardin des Plantes, Paris</undertitel>
18 <strophe>
19 <vers>Sein Blick ist vom Vorübergehn der Stäbe</vers>
20 <vers>so müd geworden, daß er nichts mehr hält.</vers>
21 <vers>Ihm ist, als ob es tausend Stäbe gäbe</vers>
22 <vers>und hinter tausend Stäben keine Welt.</vers>
23 </strophe>
24 <strophe>
25 <vers>Der weiche Gang geschmeidig starker Schritte,</vers>
26 <vers>der sich im allerkleinsten Kreise dreht,</vers>
27 <vers>ist wie ein Tanz von Kraft um eine Mitte,</vers>

```



XML-Syntax: Deklarationen Dokumenttyp-Definition (extern)

```

Altova XML Spy - [Rilke_Gedichte_ohne_alles.xml]
Datei Bearbeiten Projekt XML DTD/Schema Schema-Design XSL/Query Authentiz Konvertieren Ansicht Browser WSDL SOAP Extras Fenster Hilfe
1 <?xml version="1.0" encoding="UTF-8" standalone="no"?>
2 <!DOCTYPE sammlung SYSTEM "Gedichtsammlung.dtd">
3 <sammlung>
4 <gedicht>
5 <autor>Rainer Maria Rilke</autor>
6 <titel>Der Panther</titel>
7 <undertitel>Im Jardin des Plantes, Paris</undertitel>
8 <strophe>
9 <vers>Sein Blick ist vom Vorübergehn der Stäbe</vers>
10 <vers>so müd geworden, daß er nichts mehr hält.</vers>
11 <vers>Ihm ist, als ob es tausend Stäbe gäbe</vers>
12 <vers>und hinter tausend Stäben keine Welt.</vers>
13 </strophe>
14 <strophe>
15 <vers>Der weiche Gang geschmeidig starker Schritte,</vers>
16 <vers>der sich im allerkleinsten Kreise dreht,</vers>
17 <vers>ist wie ein Tanz von Kraft um eine Mitte,</vers>
18 <vers>in der betäubt ein großer Wille steht</vers>
19 </strophe>
20 <strophe>
21 <vers>Nur manchmal schiebt der Vorhang der Pupille</vers>
22 <vers>sich lautlos auf –. Dann geht ein Bild hinein,</vers>
23 <vers>geht durch der Glieder angespannte Stille –</vers>
24 <vers>und hört im Herzen auf zu sein.</vers>
25 </strophe>
26 </gedicht>
27

```

XML-Syntax

Generelle Regeln:

- ein Wurzelement
- alle Elemente paarig
- nur korrekt geschachtelte Elemente
- Beachtung von Groß-/Kleinschreibung
- Regeln für Namen:
 - an erster Stelle Buchstabe oder _
 - gefolgt von Buchstaben, Ziffern oder den Zeichen: _ - . :
 - nirgendwo "xml"

XML-Syntax

White Space (Leerraumzeichen)

- Leerschritt, Tabulator, Zeilenschaltung, Zeilenvorschub
- prinzipiell überall (außerhalb eines Elements) zur besseren Darstellung erlaubt
- werden bei der Verarbeitung ignoriert (in der Regel)



XML-Syntax

Sonderzeichenreferenzen:

für ü: **ü**; oder **ü**; oder **ü**

- bestehen aus & + Wert + ; nur notwendig für Zeichen, die nicht im verwendeten Zeichensatz sind
- können formuliert werden als:
 - a) benannte Entitäten
 - b) Dezimalwert (= #)
 - c) Hexadezimalwert (= #x)



XML-Syntax

XML-Entities (Defaultwerte)

< = <

> = >

& = &

" = "

' = '

XML-Syntax

Kommentare:

`<!-- uw: hier noch ein Bild einfügen -->`

- stehen in Spitzklammern
- sind kenntlich am Rufzeichen und je zwei Bindestriche vorn und hinten
- dazwischen beliebiger Kommentartext
- kann an jeder beliebigen Stelle stehen

Übung

Editieren eines XML-Dokuments mit epcEdit!

- Sich orientieren im Dokument:
Strukturbaum, Suche in Elementen, Sichten
- Prüfen der XML-Eigenschaften:
XML-Deklaration, DOCTYPE-Deklaration, Instanz-Syntax
- Editieren:
XML-Elemente und -Attribute einfügen, löschen, ändern, teilen ...
- Korrigieren von Zeichen und Auszeichnungen



**Vielen Dank für Ihre
Aufmerksamkeit!**

welsch@badw.de
ursula.welsch@welschmedien.de