

dhmuc. Workshop

„Digitale Editionen und Auszeichnungssprachen“

Domänenspezifische Notationen

Eine (noch) unterschätzte Technologie für die Digital Humanities

Eckhart Arnold, IT/Digitale Geisteswissenschaften
arnold@badw.de



Bayerische
Akademie der Wissenschaften

Überblick

- I. Der Weg der Daten in DH-Projekten
- II. Maßgeschneiderte Notationen als Eingabetechnologie
- III. Beispiele
- IV. Zum technischen Hintergrund
- V. Schlussfolgerungen

Blatt 2



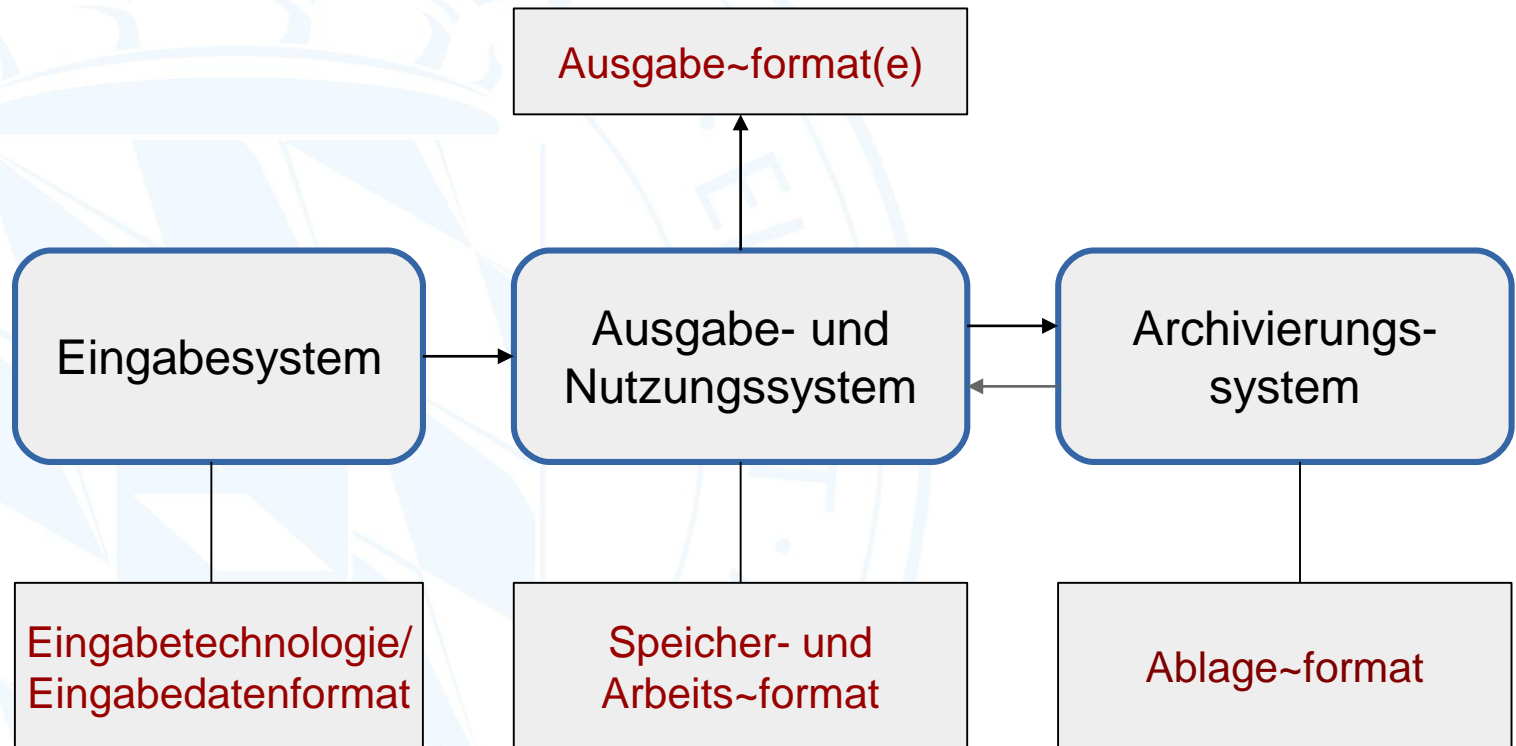
I. Der Weg der Daten in DH-Projekten

Blatt 3



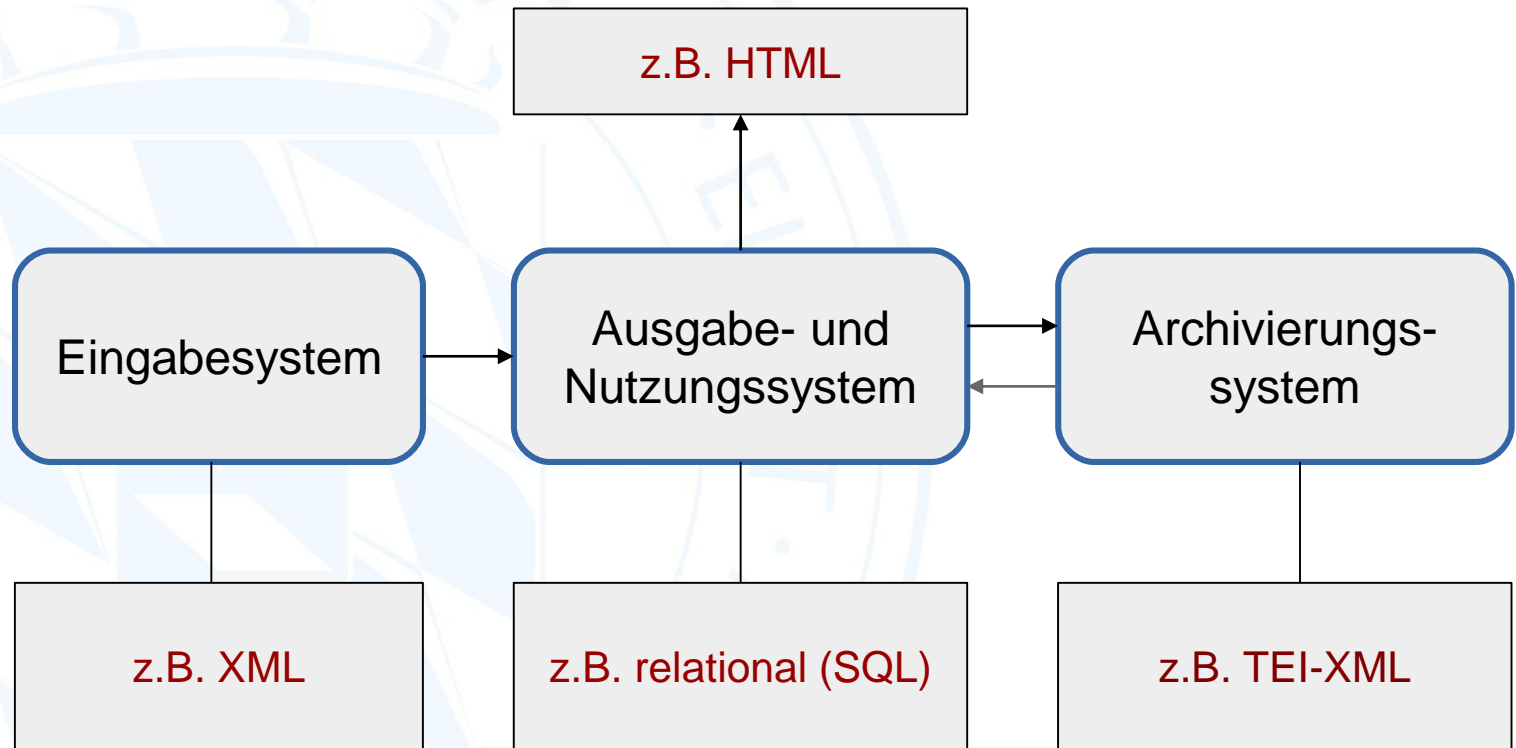
Bayerische
Akademie der Wissenschaften

Der Weg der Daten in DH-Projekten



Blatt 4

Der Weg der Daten in DH-Projekten



Blatt 5



II. Maßgeschneiderte Notationen als Eingabetechnologie

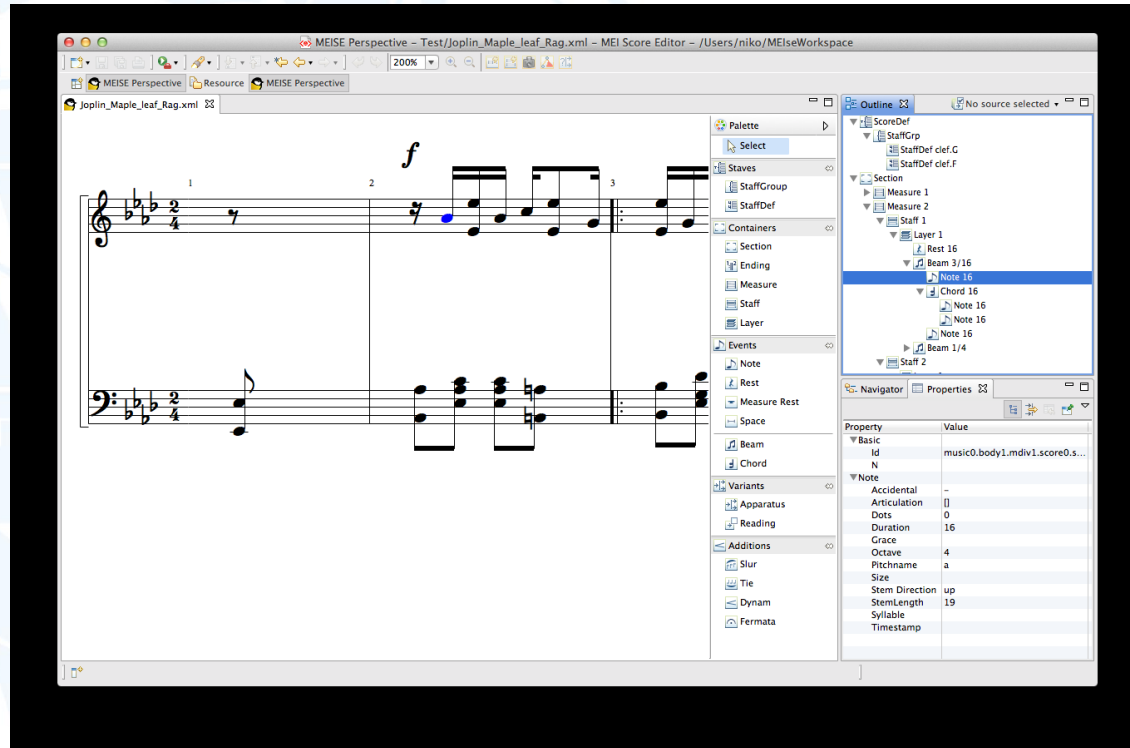
Blatt 6



Eingabetechnologien

1. **GUI-System** (grafische oder (formularbasierte (Web-)Anwendung)
2. **WYSIWIG/M bzw. „Rich-Text“-Eingabe** (Word, Oxygen-Editor, LyX)
3. **Textbasierte Notation** (z.B. TeX/LaTeX)

Beispiel für GUI-Eingabe

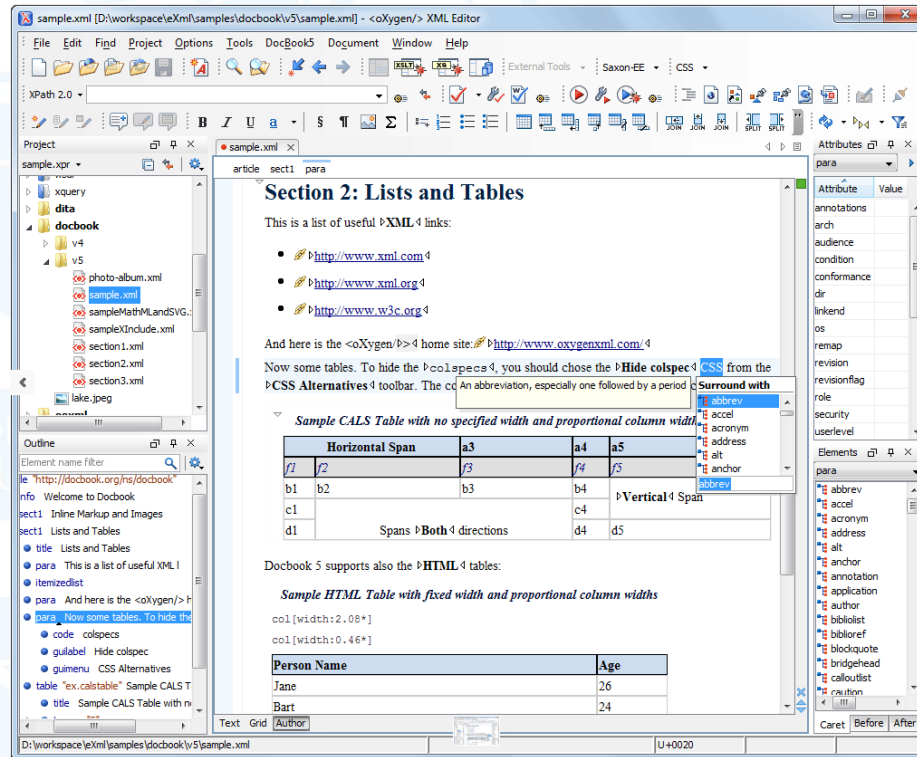


MEI Score Editor - <https://wiki.de.dariah.eu/display/TextGrid/MSMEISE>



Bayerische
Akademie der Wissenschaften

Beispiel für WYSIWYG (Rich-Text) Eingabe



Oxygen mit vorkonfigurierter Rich-Text-Eingabeoberfläche für DocBook

https://www.oxygenxml.com/xml_author/docbook_editor_structured_editing.html



Bayerische
Akademie der Wissenschaften

Beispiel: Textbasierte Notation

Einbettung von Lizenzinformationen

=====

Einsatz-Szenarien


- Einbettung von Metadaten bei der Erzeugung von Dokumenten (automatisiert / von Hand)
- Einbettung oder Veränderung von Metadaten bei der Weiterverarbeitung von Dokumenten (automatisiert / von Hand)

PDFLaTeX

```
.. code-block:: latex
```

```
\usepackage{xmpincl}  
\includemp{[METADATENDATEI]}
```

Sinnvollerweise sollte die Metadaten-Datei alle benötigten Metadaten enthalten. Das können neben den Lizenz-Metadaten z.B. auch bibliographische Metadaten sein.

 einbettung.rst 3.34 KB

Einbettung von Lizenzinformationen

Einsatz-Szenarien

- Einbettung von Metadaten bei der Erzeugung von Dokumenten (automatisiert / von Hand)
- Einbettung oder Veränderung von Metadaten bei der Weiterverarbeitung von Dokumenten (automatisiert / von Hand)

PDFLaTeX

```
\usepackage{xmpincl}  
\includemp{[METADATENDATEI]}
```

Sinnvollerweise sollte die Metadaten-Datei alle benötigten Metadaten enthalten. Das können neben den Lizenz-Metadaten z.B. auch bibliographische Metadaten sein.

Restructured-Text-Quelle

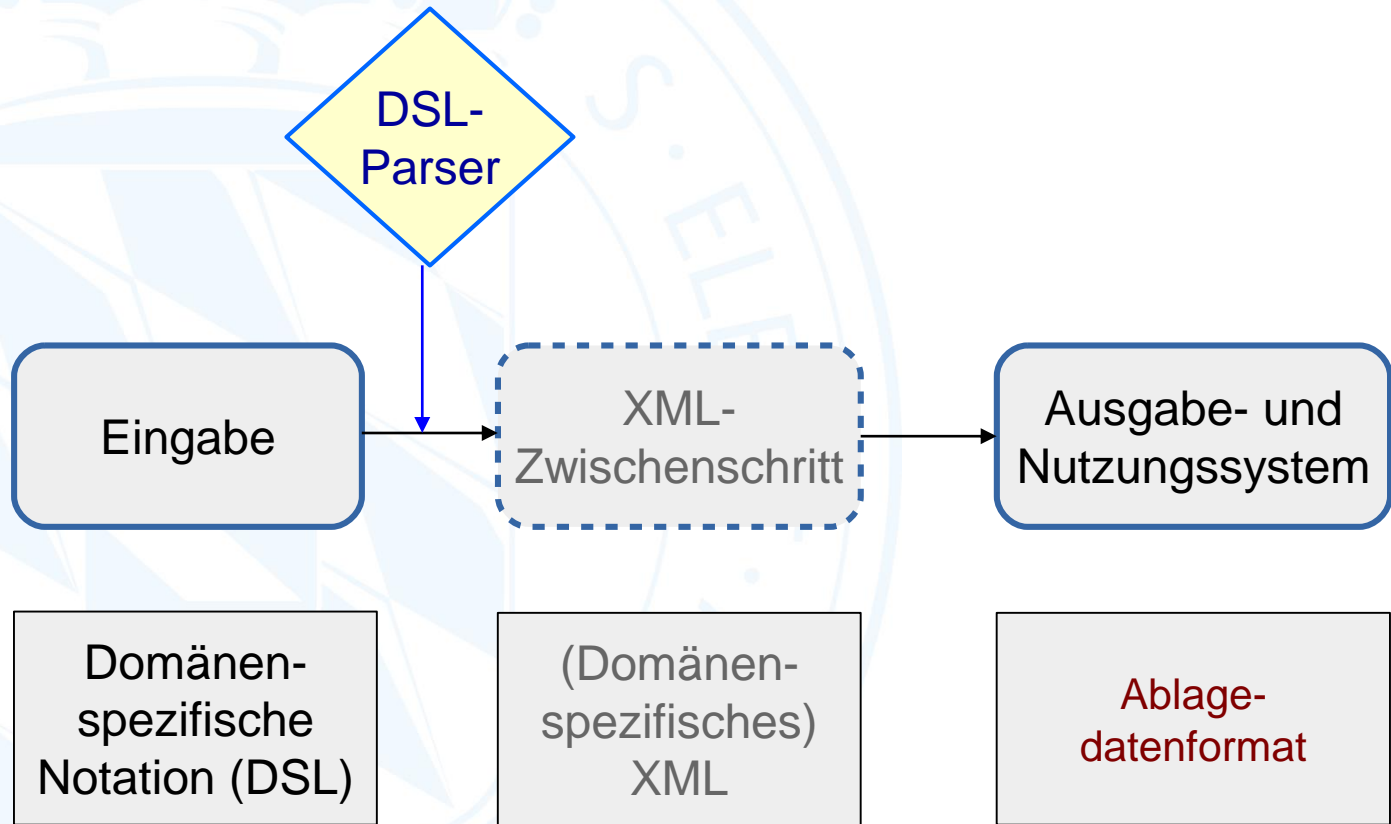
HTML-Ausgabe

Quelle: https://gitlab.lrz.de/badw-it/digitale_geisteswissenschaften_doks/blob/master/Offene_Lizenzen_2015/workshop/einbettung.rst

Bayerische
Akademie der Wissenschaften



Datenfluss bei textbasierten Notationen (Beispiel)



Blatt 11

III. Beispiele

Blatt 12



Bayerische
Akademie der Wissenschaften

Beispiel 1: Wörterbuchartikel

inconsolabilis, -e. *cui consolatio nulla est* – untröstlich, trostlos: **1 de morte**: Albert. M. Iob 17,14 subdit de fine -inconsolabili secundum corpus: 'eqs.' (*antea*: finis ... talis consolationem non habet). **2 de maerore**: Thangm. Bernw. 54 p. 781,10 ubique vociferatio lugubris exoritur, moeror -inconsolabilis ... auditur. Annal. Altah. a. 1056 p. 53,8 annus iste ... luctum -inconsolabilem intulit multis. Elis. Schon. viar. 13 p. 101,10 -inconsolabilis tristicia. Hist. de exp. Frid. imp. p. 92,23 -inconsolabili planctu in katedrali ecclesia ... corpus eius (*imperatoris*) reposuerunt. Albert. M. Iob 17,15 ex his concludit dolorem esse -inconsolabilem. *al.*

LEMMA inconsolabilis, -e

ERKLÄRUNG cui consolatio nulla est --- untröstlich

BEDEUTUNG de morte

* Albert. M. Iob 17,14 subdit de fine -inconsolabili secundum corpus: 'eqs.' (*antea*: finis ... talis consolationem non habet).

BEDEUTUNG de maerore

* Thangm. Bernw. 54 p. 781,10 ubique vociferatio lugubris exoritur, moeror -inconsolabilis ... auditur.

* Annal. Altah. a. 1056 p. 53,8 annus iste ... luctum -inconsolabilem intulit multis.

* Elis. Schon. viar. 13 p. 101,10 -inconsolabilis tristicia.

* Hist. de exp. Frid. imp. p. 92,23 -inconsolabili planctu in katedrali ecclesia ... corpus eius {*imperatoris*} reposuerunt.

* Albert. M. Iob 17,15 ex his concludit dolorem esse -inconsolabilem.

Blatt 13



DSL und XML Repräsentation

LEMMA inconsolabilis, -e

ERKLÄRUNG cui consolatio nulla est ---
untröstlich

BEDEUTUNG de morte

* Albert. M. Iob 17,14 subdit de fine -
inconsolabili secundum corpus : 'eqs.'
(antea : finis ... talis consolationem non
habet).

BEDEUTUNG de maerore

* Thangm. Bernw. 54 p. 781,10 ubique
vociferatio lugubris exoritur, moeror -
inconsolabilis ... auditur.
* Annal. Altah. a. 1056 p. 53,8 annus iste
... luctum -inconsolabilem intulit multis.
* Elis. Schon. viar. 13 p. 101,10 -
inconsolabilis tristitia.
* Hist. de exp. Frid. imp. p. 92,23 -
inconsolabili planctu in kathedrali
ecclesia ... corpus eius {imperatoris}
reposuerunt.
* Albert. M. Iob 17,15 ex his concludit
dolorem esse -inconsolabilem.

<lemma>

<wort>inconsolabilis</wort>

<flexion>-e</flexion>

<erklärung>

<interpretament>

<lat>cui consolatio nulla est</lat>

<deu>untröstlich</deu>

</interpretament>

<bedeutung>

<kategorie>de morte</kategorie>

<beleg>Albert. M. Iob 17,14 subdit de fine -inconsolabili
secundum corpus : 'eqs.' (antea : finis ... talis
consolationem non habet).</beleg>

</bedeutung>

<bedeutung>

<kategorie>de maerore</kategorie>

<beleg>Thangm. Bernw. 54 p. 781,10 ubique vociferatio
lugubris exoritur, moeror -inconsolabilis ...
auditur.</beleg>

<beleg>Annal. Altah. a. 1056 p. 53,8 annus iste ... luctum
-inconsolabilem intulit multis.</beleg>

<beleg>Elis. Schon. viar. 13 p. 101,10 -inconsolabilis
tristitia.</beleg>

<beleg>Hist. de exp. Frid. imp. p. 92,23 -inconsolabili
planctu in kathedrali ecclesia ... corpus eius

<erläuterung>imperatoris</erläuterung> reposuerunt.</beleg>

<beleg> Albert. M. Iob 17,15 ex his concludit dolorem esse
-inconsolabilem.</beleg>

</bedeutung>

</erklärung>

</lemma>



Beispiel 2: Mathematische Formeln

Der Bayes'sche Lehrsatz

$$P(p_i|q) = \frac{P(q|p_i)P(p_i)}{\sum_{i=1}^n P(q|p_i)P(p_i)}$$

$$P(p_i|q) = \frac{P(q|p_i)P(p_i)}{\sum_{i=1}^n P(q|p_i)P(p_i)}$$

LaTeX-Notation

```

<math
xmlns="http://www.w3.org/1998/Math/MathM
L">
<mrow>
<mi>P</mi>
<mrow>
<mo form="prefix">(</mo>
<msub>
<mi>p</mi></mi>
</msub>
<mo>|</mo>
<mi>q</mi>
</mrow>
<mo>=</mo>
<mfrac linethickness="1">
<mrow>
<mi>P</mi>
<mrow>
<mo form="prefix">(</mo>
<mi>q</mi>
<mo>|</mo>
<mi>p</mi>
</mrow>
<msub>
<mi>i</mi>
</msub>
</mrow>
<mo>)</mo>
<mi>P</mi>
</mrow>
</mfrac>
</mrow>

```

```

<mi>P</mi>
<mrow>
<mo form="prefix">(</mo>
<msub>
<mi>p</mi>
<mi>i</mi>
</msub>
</mrow>
<mo>=</mo>
<mfrac
displaystyle="true">
<munderover>
<mo>&#x02211;</mo>
<mrow>
<mi>i</mi>
<mo>=</mo>
<mn>1</mn>
</mrow>
</munderover>
</mstyle>
<mi>P</mi>

```

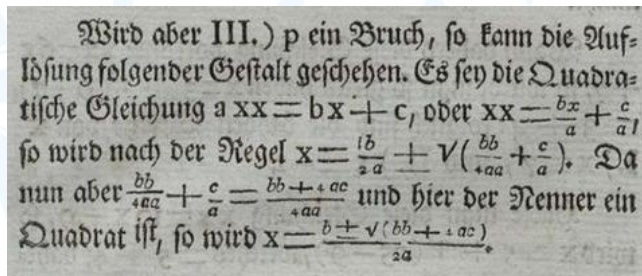
```

<mrow>
<mo form="prefix">(</mo>
<mi>q</mi>
<mo>|</mo>
<mi>p</mi>
<mi>i</mi>
</msub>
<mo>=</mo>
<mfrac
displaystyle="true">
<mo>&#x02211;</mo>
<mrow>
<mi>p</mi>
<mi>i</mi>
</mrow>
</mfrac>
</mrow>

```



Auch TEI-XML erlaubt Fremdnotationen



TEI-XML (Ausschnitt):

nun aber

```
<formula notation="TeX">\frac{bb}{4 aa}</formula> +  
<formula notation="TeX">\frac{c}{a}</formula> =  
<formula notation="TeX">\frac{bb + 4 ac}{4 aa}</formula>  
und hier der Nenner ein  
<lb/>
```

```
Quadrat i&#x017F;t, &#x017F;o wird<hi rendition="#aq">x</hi> =  
<formula notation="TeX">\frac{b \pm \sqrt{(bb + 4 ac)}}{2 a}</formula>.
```

Wird aber III.) p ein Bruch, so kann die Auflösung folgender Gestalt geschehen. Es sey die Quadratische Gleichung $axx = bx + c$, oder $xx = \frac{bx}{a} + \frac{c}{a}$,

so wird nach der Regel $x = \frac{1b}{2a} \pm \sqrt{\left(\frac{bb}{4aa} + \frac{c}{a}\right)}$. Da

nun aber $\frac{bb}{4aa} + \frac{c}{a} = \frac{bb + 4ac}{4aa}$ und hier der Nenner ein

Quadrat ist, so wird $x = \frac{b \pm \sqrt{(bb + 4ac)}}{2a}$.

Quelle: Leonard Euler, Vollständige Anleitung zur Algebra, Band 2, St. Petersburg 1770, URL: http://www.deutschestextarchiv.de/book/view/euler_algebra02_1770?p=76

Beispiel 3: Markdown

Markdown

Markdown als HTML-Aufsatz

Markdown ist eine "sparsame" Auszeichnungssprache, die nur eine kleine Menge von Textelementen definiert. Reichen die vordefinierten Elemente nicht aus, so kann man innerhalb eines Markdown-Dokuments auch HTML-Elemente einbetten, z.B. eine Tabelle:

Markdown	HTML
[Website]	
(https://commonmark.org)	Website

Markdown
=====

Markdown als HTML-Aufsatz

Markdown ist eine "sparsame" Auszeichnungssprache, die nur eine kleine Menge von Textelementen definiert. Reichen die vordefinierten Elemente nicht aus, so kann man innerhalb eines Markdown-Dokuments auch HTML-Elemente einbetten, z.B. eine Tabelle:

```
<table>
<tr><td>Markdown</td><td>&nbsp;</td><td>HTML</td></tr>
<tr>
  <td>[Website] (https://commonmark.org)</td>
  <td>&nbsp;</td>
  <td>&lt;a href="https://commonmark.org"&gt;
    Website&lt;/a&gt;</td>
</tr>
</table>
```

```
<h1>Markdown</h1>
```

```
<h2>Markdown als HTML-Aufsatz</h2>
```

```
<p>Markdown ist eine &quot;sparsame&quot;Auszeichnungssprache,
die nur eine kleine Menge von Textelementen definiert.
Reichen die vordefinierten Elemente nicht aus, so kann
man innerhalb eines Markdown-Dokuments auch HTML-Elemente
einbetten, z.B. eine Tabelle:</p>
```

```
<table>
<tr><td>Markdown</td><td>&nbsp;</td><td>HTML</td></tr>
<tr>
  <td>[Website] (https://commonmark.org)</td>
  <td>&nbsp;</td>
  <td>&lt;a href="https://commonmark.org"&gt;
    Website&lt;/a&gt;</td>
</tr>
</table>
```

Markdown-Syntax

HTML-Syntax

IV. Zum technischen Hintergrund

Blatt 18



Bayerische
Akademie der Wissenschaften

Formale Spezifikation von DSLs

LEMMA inconsolabilis, -e

ERKLÄRUNG cui consolatio nulla est ---
untröstlich

BEDEUTUNG de morte

* Albert. M. Iob 17,14 subdit de fine -
inconsolabili secundum corpus : 'eqs.'
(antea : finis ... talis consolationem non
habet).

BEDEUTUNG de maerore

* Thangm. Bernw. 54 p. 781,10 ubique
vociferatio lugubris exoritur, moeror -
inconsolabilis ... auditur.
* Annal. Altah. a. 1056 p. 53,8 annus iste
... luctum -inconsolabilem intulit multis.
* Elis. Schon. viar. 13 p. 101,10 -
inconsolabilis tristitia.
* Hist. de exp. Frid. imp. p. 92,23 -
inconsolabili planctu in kathedrali
ecclesia ... corpus eius {imperatoris}
reposerunt.
* Albert. M. Iob 17,15 ex his concludit
dolorem esse -inconsolabilem.

lemma = "LEMMA" wort ["," flexion] erklaerung
Flexion = TEXT
erklaerung = "ERKLÄRUNG"
[interpretament] { bedeutung }
bedeutung = "BEDEUTUNG"
(interpretament | kategorie) { beleg }
interpretament = lateinisch "----" deutsch
lateinisch = WORT{!"----"(TEXTBLOCK|SONDERZEICHEN)}
Deutsch = WORT TEXT
beleg = "*" WORT { TEXT | erlaeuterung }
Erlaeuterung = "{" { TEXT | SONDERZEICHEN } }"
TEXT = { TEXTBLOCK | SONDERZEICHEN }
TEXTBLOCK = { WORT | LEERRAUM | ZAHL }
WORT = /[A-ZÄÖÜ]?[a-zäöüß]*/

EBNF-Grammatik für Wörterbuchartikel (unvollst. Auszug)



Transformation DSL => Datenstruktur

LEMMA inconsolabilis, -e

ERKLÄRUNG cui consolatio nulla est ---
untröstlich

BEDEUTUNG de morte

* Albert. M. Iob 17,14 subdit de fine -
inconsolabili secundum corpus : 'eqs.' (antea :
finis ... talis consolationem non habet).

BEDEUTUNG de maerore

* Thangm. Bernw. 54 p. 781,10 ubique vociferatio
lugubris exoritur, moeror -inconsolabilis ...
auditur.
* Annal. Altah. a. 1056 p. 53,8 annus iste ...
luctum -inconsolabilem intulit multis.
* Elis. Schon. viar. 13 p. 101,10 -
inconsolabilis tristicia.
* Hist. de exp. Frid. imp. p. 92,23 -
inconsolabili planctu in kathedrali ecclesia ...
corpus eius {imperatoris} reposuerunt.
* Albert. M. Iob 17,15 ex his concludit dolorem
esse -inconsolabilem.

<lemma>
<wort>inconsolabilis</wort>
<flexion>e</flexion>

<erklärung>
<interpretament>
<lat>cui consolatio nulla est</lat>
<deu>untröstlich</deu>
</interpretament>

<bedeutung>
<kategorie>de morte</kategorie>
<beleg>Albert. M. Iob 17,14 subdit de fine -inconsolabili
secundum corpus : 'eqs.' (antea : finis ... talis
consolationem non habet).</beleg>
</bedeutung>

<bedeutung>
<kategorie>de maerore</kategorie>
<beleg>Thangm. Bernw. 54 p. 781,10 ubique vociferatio
lugubris exoritur, moeror -inconsolabilis ...
auditur.</beleg>
<beleg>Annal. Altah. a. 1056 p. 53,8 annus iste ... luctum
-inconsolabilem intulit multis.</beleg>
<beleg>Elis. Schon. viar. 13 p. 101,10 -inconsolabilis
tristicia.</beleg>
<beleg>Hist. de exp. Frid. imp. p. 92,23 -inconsolabili
planctu in kathedrali ecclesia ... corpus eius
<erläuterung>imperatoris</erläuterung> reposuerunt.</beleg>
<beleg> Albert. M. Iob 17,15 ex his concludit dolorem esse
-inconsolabilem.</beleg>
</bedeutung>
</erklärung>
</lemma>

DSL-Syntax
(EBNF)

DSL-
Parser

Blatt 20



Resümé zu DLS

- Wissenschaftlich wohlverstandener theoretischer Hintergrund:
- (kontextfreie Sprachen, EBNF-Grammatiken, syntaxgetriebenes Übersetzen)

- Breite Verfügbarkeit von Werkzeugen,
- die aber ggf. zugeschnitten werden müssen

- Leichtgewichtige Technologie („Werkzeugkastenphilosophie“):
- (wenig eigener Code erforderlich, geringe Folgelasten)

- Flexible Technologie – Datenstrukturen werden nicht präjudiziert

- Breites Anwendungsfeld (u.a. für die Retrodigitalisierung)

- Unkompliziert in der Anwendung

- Aber: Technische Betreuung erfordert ein gewisses Know-How



Vergleich DSL <=> XML

DSL

- leichtgewichtig
- Gute Lesbarkeit und Handhabbarkeit
- breite Verfügbarkeit von Werkzeugen
- ermöglicht strukturierte Datenerfassung
- lose Kopplung an ein bestimmtes Datenmodell
- theoretisch fundiert und praxiserprobt
- höheres Technik-Know-How erfordert
- in den DH noch wenig verbreitet

XML

- leichtgewichtig
- mäßige Handhabbarkeit und Lesbarkeit für die Anwender
- breite Verfügbarkeit von Werkzeugen
- ermöglicht strukturierte Datenerfassung
- enge Kopplung an ein bestimmtes Datenmodell (Baumstruktur)
- theoretisch fundiert und praxiserprobt
- Technik relativ leicht zu erlernen
- in den DH wohletabliert



V. Fazit

TENDIT AD
ÆQUUM

59

Blatt 23



Bayerische
Akademie der Wissenschaften

Fazit

Für DSLs gibt es in den Digital Humanities ein breites Einsatzgebiet:

- als Alternative zu WYSIWYG/M - und Rich-Text-Editoren (Word, Oxygen, Lyx)
- als leichtgewichtige Technik für Cross-Media-Publishing (z.B. <http://scholarlymarkdown.com/>)
- als Alternative zu XML
- als alternative Parsing-Technologie zu regulären Ausdrücken

Blatt 24

