Susanne Haaf

Das DTA-Basisformat zur TEI-XML-konformen Annotation historischer Textressourcen

Vortrag
Digital Humanities Kolloquium an der Berlin-Brandenburgischen Akademie der Wissenschaften ; (Berlin) : 2017.09.01

Persistent Identifier: urn:nbn:de:kobv:b4-opus-28079

Die vorliegende Datei wird Ihnen von der Berlin-Brandenburgischen Akademie der Wissenschaften unter einer Creative Commons Attribution- ShareAlike 3.0 Germany (cc by-sa 3.0) Licence zur Verfügung gestellt.

Das DTA-Basisformat zur TEI-XML-konformen Annotation historischer Textressourcen

Susanne Haaf

Vortrag beim Berliner DH-Kolloquium
1. September 2017

Gliederung des Vortrags

1. Einführung zum Thema
2. Einführung zum DTA-Basisformat-Tagset
3. Anwendung des DTA-Basisformats
4. "Hinter den Kulissen" – Technische Umsetzung
5. Verwendungsszenarien
6. Ausblick

Gliederung des Vortrags

1. Einführung zum Thema
2. Einführung zum DTA-Basisformat-Tagset
3. Anwendung des DTA-Basisformats
4. "Hinter den Kulissen" – Technische Umsetzung
5. Verwendungsszenarien
6. Ausblick

Zweyter Abschnitt.

Von den organisirten Körpern überhaupt.

§. 5.

Im allgemeinen werden die organisirten Körper (§. 2.) von ihres Gleichen *) erzeugt, dann durch eigene Kraft lebenslang ernährt, und dadurch ihre Selbsterhaltung und Wachsthum, und wenn sie zu ihrer Reife gelangt, auch ihre Fortpflanzungsfähigkeit bewirkt.

§. 6.

Zu diesen großen Verrichtungen werden sie eben durch die Organisation ihres Baues, und durch die mit derselben verbundenen Lebenskräfte geschickt gemacht. Denn durch diese letztern erhalten die Organe sowohl ihre Empfänglichkeit für reitzende Eindrücke (stimuli) als ihr Bewegungsvermögen, ohne welches beides, weder Ernährung noch Wachsthum, noch wechselseitige Einwirkung der Theile zur zweckmäßigen Erhaltung des Ganzen, und umgekehrt **), denkbar seyn könnte.

§. 7.

Sich die Entstehung der organisirten Körper zu erklären, hat man, zumahl neuerlich, die so genannte Evolutions-Hypothese bequem gefunden, und gemeint, es werde gar kein Mensch, und kein anderes Thier, und keine Pflanze erzeugt, — sondern sie lägen alle schon seit der ersten Schöpfung als völlig präformirte Keime ***) bei ihren Aeltern und

*) s. oben S. 2. Not. *)

**) Vergl. Kant's Critik der Urtheilskraft. S. 285 u. f.

***) »Denn« (so sagt Haller, das Haupt der neuern Evolutionisten —) »alle Eingeweide und die Knochen selbst waren schon im »unsichtbaren Keim vorhero gebaut gegenwärtig, obgleich in einem fast »flüssigen Zustande.«

Und das ist doch wenigstens bestimmte Sprache.

Welche Strukturen sind zu erkennen?

Zweyter Abschnitt.

Von den organisirten Körpern überhaupt.

§. 5.

Im allgemeinen werden die organisirten Körper (§. 2.) von ihres Gleichen *) erzeugt, dann durch eigene Kraft lebenslang ernährt, und dadurch ihre Selbsterhaltung und Wachsthum, und wenn sie zu ihrer Reife gelangt, auch ihre Fortpflanzungsfähigkeit bewirkt.

§. 6.

Zu diesen großen Verrichtungen werden sie eben durch die Organisation ihres Baues, und durch die mit derselben verbundenen Lebenskräfte geschickt gemacht. Denn durch diese letztern erhalten die Organe sowohl ihre Empfänglichkeit für reitzende Eindrücke (stimuli) als ihr Bewegungsvermögen, ohne welches beides, weder Ernährung noch Wachsthum, noch wechselseitige Einwirkung der Theile zur zweckmäßigen Erhaltung des Ganzen, und umgekehrt **), denkbar seyn könnte.

§. 7.

Sich die Entstehung der organisirten Körper zu erklären, hat man, zumahl neuerlich, die so genannte Evolutions-Hypothese bequem gefunden, und gemeint, es werde gar kein Mensch, und kein anderes Thier, und keine Pflanze erzeugt, — sondern sie lägen alle schon seit der ersten Schöpfung als völlig präformirte Keime ***) bei ihren Aeltern und

*) s. oben S. 2. Not. *)

**) Vergl. Kant's Critik der Urtheilskraft. S. 285 u. f.

***) »Denn« (so sagt Haller, das Haupt der neuern Evolutionisten —) »alle Eingeweide und die Knochen selbst waren schon im »unsichtbaren Keim vorhero gebaut gegenwärtig, obgleich in einem fast »flüssigen Zustande.«

Und das ist doch wenigstens bestimmte Sprache.

Welche Strukturen sind zu erkennen?

Wofür "strukturelle" Textauszeichnung?

- Layout-Informationen, die inhaltliche Informationen repräsentieren (sehr altes Konzept)

  → Informationsgehalt der Quelle bewahren
- Wissenschaftliche Editionen (editorische Anmerkungen, kritischer Apparat, angemessene Wiedergabe der Quelle)

- Unterbrechungen des Fließtextes dokumentieren für die linguistische Analyse
- Korpusanalysen basierend auf bestimmten Strukturen

- TEI
  - Text Encoding Initiative, www.tei-c.org
  - Gegründet 1987
  - Struktur: Board of Directors, Technical Council, Community
- TEI-Guidelines
  - Richtlinien für die XML-Auszeichnung von Texten
  - Annotation semantischer, typographischer, linguistischer, graphischer etc. Phänomene

```
<hi rendition="#bold">Platon</hi>
<persName>Platon</persName>
```

»The impact of the TEI on digital scholarship has been enormous. «

- »Encoding Schema der Wahl für:
  - Kritische Editionen
  - Wissenschaftliche Referenzwerke
  - Große linguistische Korpora
  - Detaillierte Metadaten zu elektronischen Texten und Kulturerbe-Sammlungen«

Cf. http://www.tei-c.org/About/history.xml

Projects, using the TEI

WE USE TEI

ECCO TCP Eighteenth Century Collections Online Text Creation Partnership

eebo tcp Early English Books Online Text Creation Partnership

DTA

TCP Transcribed by hand. Owned by libraries. Made for everyone.

DWDS

University of Oxford Text Archive

HENRIK IBSENS SKRIFTER

Blumenbach – online

TextGrid Repository

DIGITALISIERUNG DES POLYTECHNISCHEN JOURNALS

Sandrart.NET

etc. ...

Cf. www.tei-c.org/Activities/Projects/

TEI und Interoperabilität

- Wenn jeder nach den TEI/P5-Richtlinien auszeichnet: Erhalten wir interoperable Texte?

  Oder:

- Sind TEI-Texte per se interoperabel? (Unsworth 2011)

- Interoperabilität:
  - Möglichkeit, Texte mit denselben Methoden und Tools auszuwerten mit verlässlichen Resultaten

TEI und Interoperabilität

- TEI bietet unterschiedliche Markup-Optionen für ähnliche Phänomene

  z.B. Eigennamen

TEI

```
<rs type="propNounPersName">Sokrates</rs>,
<name type="person">Platon</name> und
<persName>Aristoteles</persName> waren Philosophen.
```

- Warum?
  - TEI als ein “set of building blocks for creating a schema suitable for a particular project”
  - Philosophie, Projekten Angebote, nicht Vorschriften für das Tagging zu machen

Cf. http://www.tei-c.org/Guidelines/Customization/odds.xml

DTA-Basisformat (DTABf)

- Format für die Auszeichnung von historischen Texten in Korpora
- Echte Teilmenge von TEI-P5
- Datengrundlage: DTA-Korpora
  - 4.247 Texte des 17.-19. Jh.
  - unterschiedliche Textsorten und Disziplinen

  → breite Datenbasis
- Leitgedanken:
  - Identisches Tagging für semantisch ähnliche Phänomene
  - Homogene Textstrukturierung des gesamten Korpus
  - Möglichst hohe Abdeckung von Phänomenen in geschriebenen historischen Texten
  - Vermeidung von Ambiguitäten (Interoperabilität gewährleisten)
  - Dokumentation von Entscheidungen

Gliederung des Vortrags

1. Einführung zum Thema
2. Einführung zum DTA-Basisformat-Tagset
3. Anwendung des DTA-Basisformats
4. "Hinter den Kulissen" – Technische Umsetzung
5. Verwendungsszenarien
6. Ausblick

DTA-Basisformat (DTABf)

TEI

```
<rs type="propNounPersName">Sokrates</rs>,
<name type="person">Platon</name> und
<persName>Aristoteles</persName> waren Philosophen.
```

DTABf

```
<persName>Sokrates</persName>,
<persName>Platon</persName> und
<persName>Aristoteles</persName> waren Philosophen.
```

Was wird ausgezeichnet

- Formale (Layout-) Textauszeichnung

  (Spaltensatz, Abbildungen und Trennlinien, Formeln, Tabellen, Listen, Seiten-/Zeilenumbrüche, ...)

- Inhaltliche (Logische, Konzeptuelle) Textauszeichnung

  (Titelblätter, Kapitel, Gedichte, Sprechakte im Drama, Erläuterungen/Metatext, Zitate, ...)

- Auszeichnung von Metadaten

  (Aufbewahrungsort der Quelle, Signatur, Autor, Erscheinungsjahr, Textsorte, Sprache, Schriftart, ...)

- Richtlinien zur Transkription

  (Nähe zur Quelle, Umgang mit Ligaturen, Sonderzeichen, Interpunktion, ...)

Über das Tagset

Tagset:	Echtes Subset der TEI
TEI:	**552** Elemente + Attribute
DTABf:	**129** Elemente + Attribute und Werte **83** im Textbereich (<text>) (davon 7 nur in Manuskripten, 76 in Drucken) **46** im Headerbereich (<teiHeader>)

Über das Tagset

- Festlegungen im DTABf nicht nur für Elemente, sondern auch für Attribute und deren Werte
  - *Element*: übergeordnete Semantik
  - *Attribute und Werte*: nähere Spezifikation der Elemente
- **Attribute**: Beispiel Anmerkungen
  - Element:

    `<note>`
  - Attribute:

    „lokale“ Attribute (= TEI P5): `@place, @type, @n`

    globale Attribute (= TEI P5): `@xml:id, @sameAs, @corresp, @prev, @next, @rendition`

    nicht gültig (↔ TEI P5): `@cert, @resp, @subtype, @evaluate, @facs, @ana, @targetLang, @target, ...`

Über das Tagset

- **Werte**: drei Möglichkeiten
  - Wert-Auswahl je Attribut festgelegt (je Element oder Klasse)

    Beispiel `<note>`:

    `@place="left|right|end|foot"`

    `@type="editorial"`

  - Attribut festgelegt, Datenstruktur des Wertes festgelegt

    Beispiel `<lg>`:

    `@n="[data.count]"`

  - keine Festlegung beim Wert (wenn unabdingbar)

    Beispiel `<note>` :

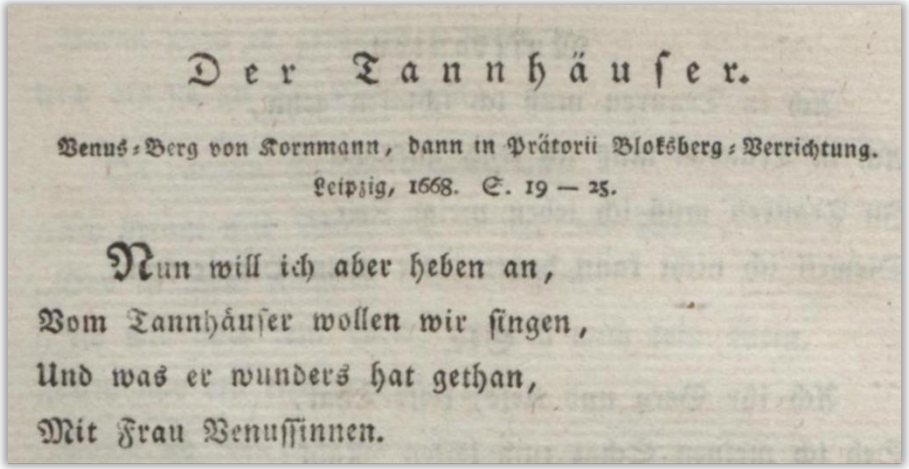
    @n (für das Fußnotenzeichen)

Tagging nach dem DTABf

```
… haben in dero
Welt-<lb/>
<note place="left">
   zu Mu͛nchen
</note>beru͛hmten
Residenz …
```

Sandrart, Joachim von: L'Academia Todesca. Bd. 1,1. Nürnberg, 1675.
http://www.deutschestextarchiv.de/sandrart_academie0101_1675

Tagging nach dem DTABf



Der Tannhäuser.

Venus-Berg von Kornmann, dann in Prätorii Bloksberg-Verrichtung.
Leipzig, 1668. S. 19 — 25.

Nun will ich aber heben an,
Vom Tannhäuser wollen wir singen,
Und was er wunders hat gethan,
Mit Frau Venussinnen.

Nun will ich aber heben an,
Vom Tannhäuser wollen wir singen,
Und was er wunders hat gethan,
Mit Frau Venussinnen.

```
<lg type="poem">
  <lg n="1">
    <l>Nun will ich aber heben an,</l><lb/>
    <l>Vom Tannhäuser wollen wir singen,</l><lb/>
    <l>Und was er wunders hat gethan,</l><lb/>
    <l>Mit Frau Venussinnen.</l>
  </lg><lb/>
<lg><lb/>
```

Arnim/Brentano: Des Knaben Wunderhorn. Bd. 1. Heidelberg, 1806, S. 86,
http://www.deutschestextarchiv.de/arnim_wunderhorn01_1806/95

Metadaten

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title type="main">Des Knaben Wunderhorn</title>
      <title type="sub">Alte deutsche Lieder</title>
      <title type="volume" n="1"></title>
      <author>
        <persName ref="http://d-nb.info/gnd/118504177">
          <surname>Arnim</surname>
          <forename>Achim von</forename>
        </persName>
      </author>
      <author>
        <persName ref="http://d-nb.info/gnd/118515055">
          <surname>Brentano</surname>
          <forename>Clemens</forename>
        </persName>
      </author>
      [...]
      <availability>
        <licence target="http://creativecommons.or
          <p>Distributed under the Creative Common
              Unported License.</p>
        </licence>
      </availability>
      [...]
        <publicationStmt>
            <publisher><name>Mohr u: Zimmer</name>
            <pubPlace>Heidelberg</pubPlace>
            <pubPlace>Frankfurt</pubPlace>
          <date type="publication">1806</date>
        </publicationStmt>
      [...]
```

Arnim, Achim von; Brentano, Clemens: Des Knaben Wunderhorn. Bd. 1. Heidelberg, 1806.

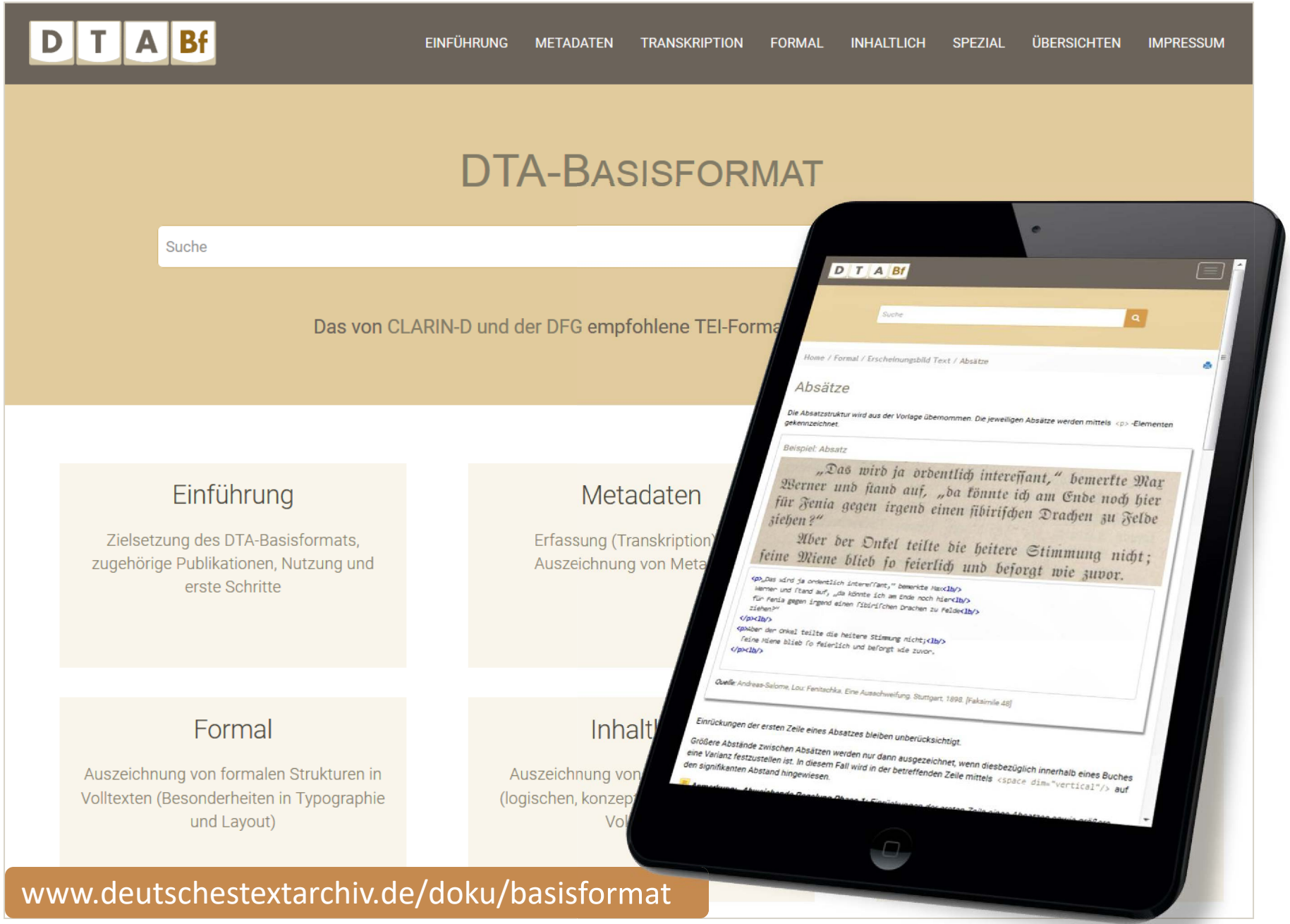
Bibliographische Angaben

URN:	urn:nbn:de:kobv:b4-20090519157
Titel:	Des Knaben Wunderhorn
Untertitel:	Alte deutsche Lieder
Autor/innen:	Achim von Arnim (GND, Wikipedia, ADB/NDB) Clemens Brentano (GND, Wikipedia, ADB/NDB)
Erscheinungsjahr:	1806
Verlag:	Mohr u: Zimmer
Ort:	Heidelberg; Frankfurt
Band:	1 von 3
Auflage:	1. Auflage
Bibliothek:	Staatsbibliothek zu Berlin – Preußischer Kulturbesitz
Signatur:	SBB-PK, 19 ZZ 2004-1 R

Gliederung des Vortrags

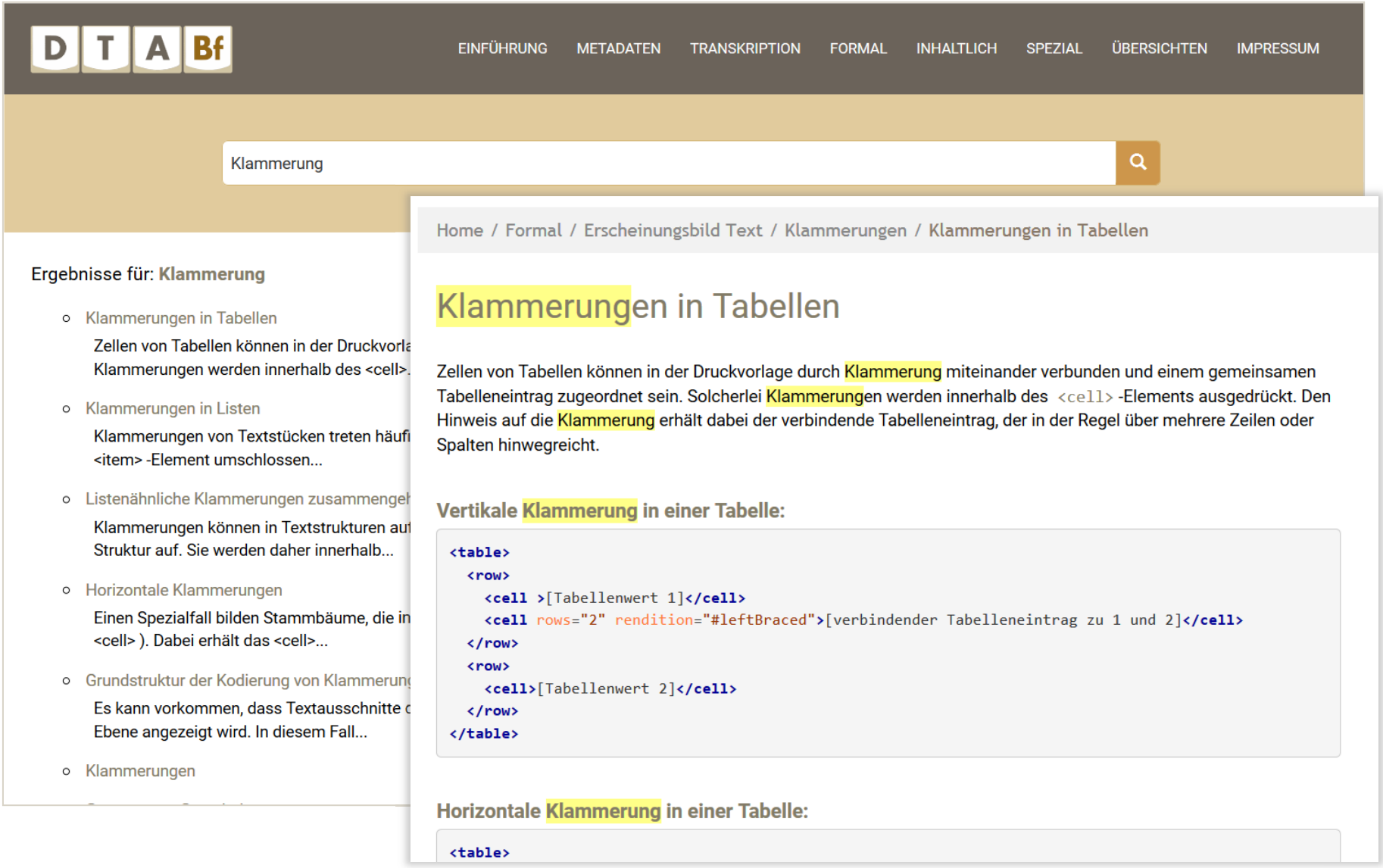
1. Einführung zum Thema
2. Einführung zum DTA-Basisformat-Tagset
3. Anwendung des DTA-Basisformats
4. "Hinter den Kulissen" – Technische Umsetzung
5. Verwendungsszenarien
6. Ausblick

Dokumentation des Formats

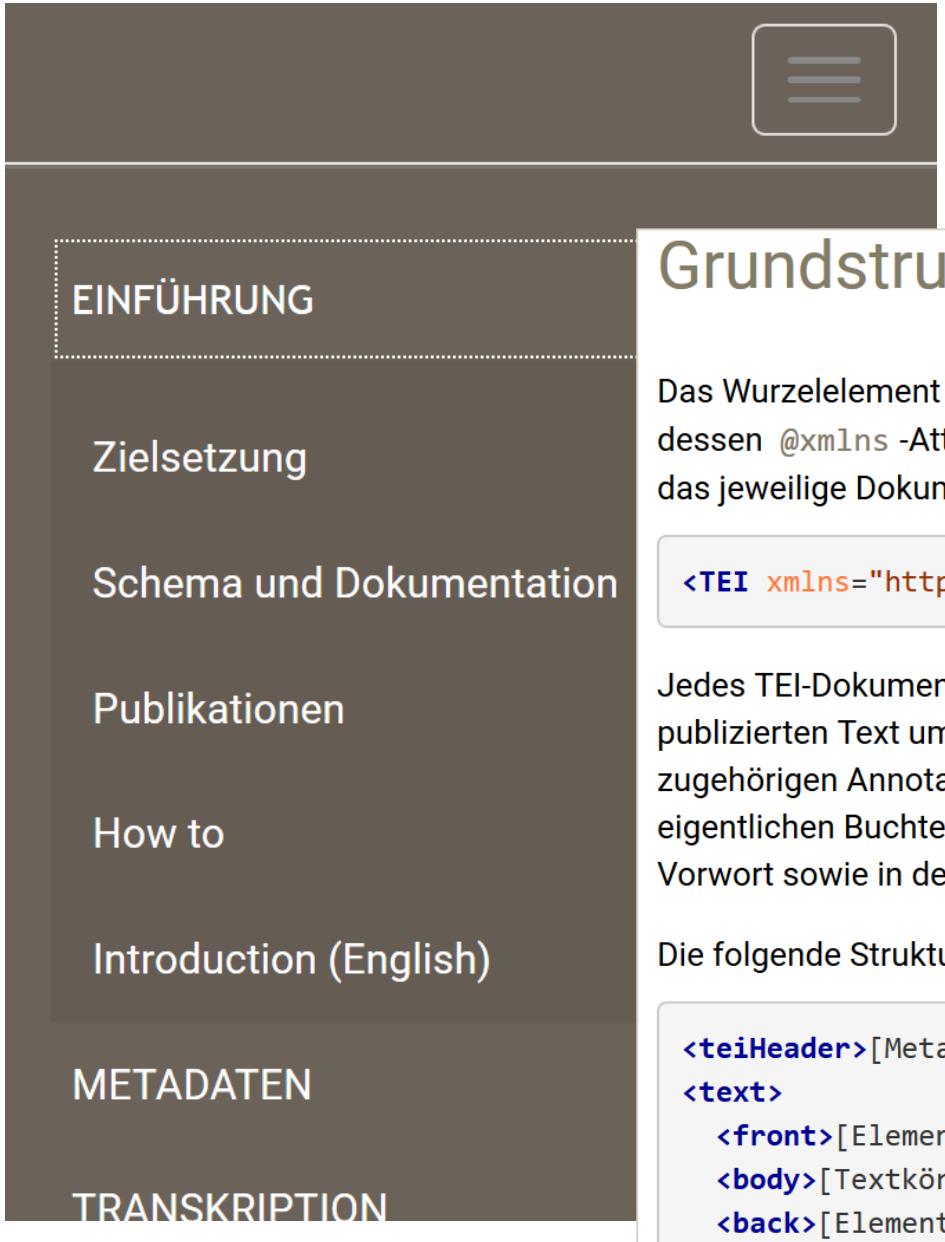


www.deutschestextarchiv.de/doku/basisformat

Dokumentation: Volltextsuche



Dokumentation: Systematischer Einstieg



Grundstruktur jedes TEI-Dokuments

Das Wurzelelement jedes TEI-Dokuments im DTA bildet das `<TEI>` -Element, dessen `@xmlns` -Attribut den TEI-Namensraum (namespace) spezifizieren, der für das jeweilige Dokument gilt:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">...</TEI>
```

Jedes TEI-Dokument im DTA besteht aus einem Header, der Metadaten zum publizierten Text umfasst, und einem Textbereich, der alle Transkriptionen mit den zugehörigen Annotationen enthält. Dieser Volltext umfasst dabei nicht allein den eigentlichen Buchtext, sondern alle Textbestandteile, so auch Titelseite und Vorwort sowie in der Regel die Register, Beigaben und Anhänge.

Die folgende Struktur wird in jedem TEI-Dokument eingehalten:

```
<teiHeader>[Metadaten]</teiHeader>
<text>
  <front>[Elemente vor Beginn des Buchtextes]</front>
  <body>[Textkörper]</body>
  <back>[Elemente nach Abschluss des Buchtextes]</back>
</text>
```

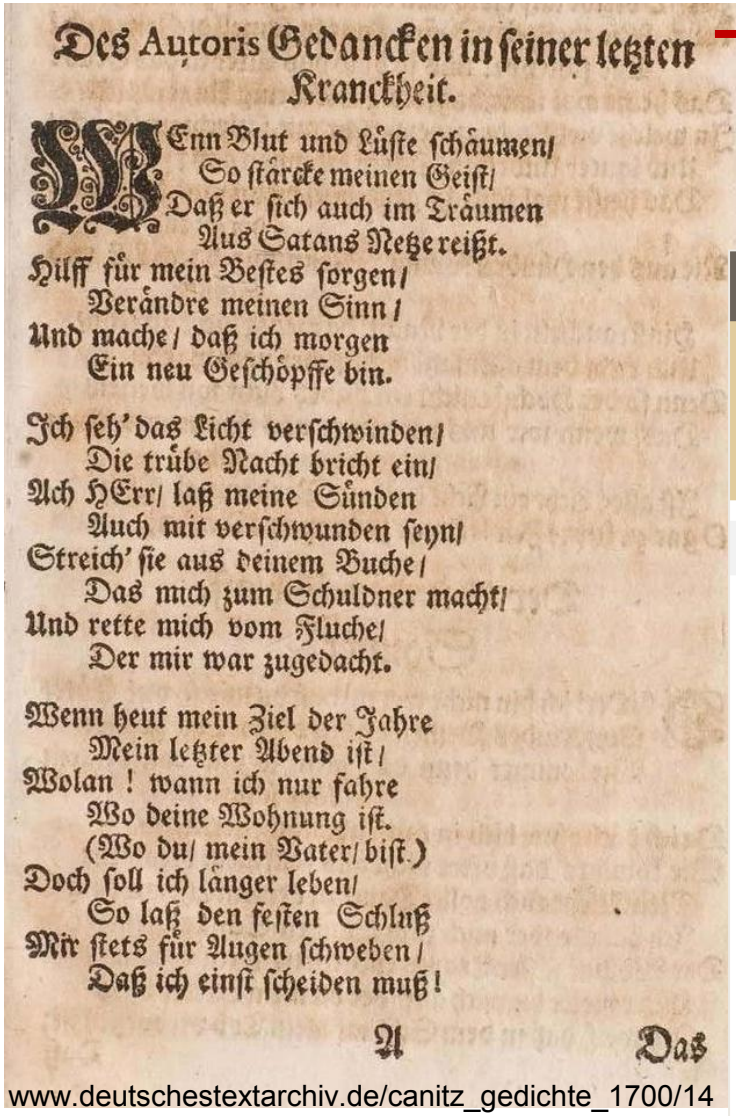
Element	Beschreibung	Attribute/Anmerkungen	Funktionale Kategorie	Level
`ab` Doku: TEI	Anonymer Block/Container	*stattdessen* `div` , `floatingText` *oder* `p` *benutzen*	textStructure	4 (unzulässig)
`abbr` Doku: TEI, DTABf	Abkürzung		phraseStructure, editorial	3 (fakultativ)
`actor` Doku: TEI, DTABf	Name eines Schauspielers innerhalb einer Darstellerliste		drama, phraseStructure	2 (empfohlen)
`corr` Doku: TEI, DTABf	(bei Korrekturen:) korrigierte Form	`@resp`	Manuskripte: Person, die verantwortlich ist für die Ergänzung [`"[pointer]"`]	editorial	3 (fakultativ)
		`@type`	Ursprung der Korrektur [`"addenda"` , `"corrigenda"` , `"editorial"`]		
`date` Doku: TEI, DTABf	Datumsangabe	`@when`	Datum in ISO-8601-Form	phraseStructure	3 (fakultativ)

Schema zur Validierung der XML-Instanzen

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="http://www.deutschestextarchiv.de/basisformat.rng"
type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<?xml-model href="http://www.deutschestextarchiv.de/basisformat.sch"
type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
<fileDesc>
<titleStmt>
<title type="main">Von wahrem Christenthumb</title>
```

```
mut/ Gedult/ Gottesfurcht. Sihe das<lb/>
heilige gebet des <hi rendition="#cap">Herr</hi>n das Vater<lb/>
vnſer an. Daſſelbe gehet wider dich vnd<lb/>
fu&#x0364;r dich. Sol Gottes Name allein<lb/>
```

```
mut/ Gedult/ Gottesfurcht. Sihe das<lb/>
heilige gebet des <hi rendition="#k">Herr</hi>n das Vater<lb/>
vnſer an. Daſſelbe gehet wider dich vnd<lb/>
fu&#x0364;r dich. Sol Gottes Name allein<lb/>
```



Des Autoris Gedancken in seiner letzten Kranckheit.

WEnn Blut und Lüste schäumen/
So stärcke meinen Geist/
Daß er sich auch im Träumen
Aus Satans Netze reißt.
Hilff für mein Bestes sorgen/
Verändre meinen Sinn/
Und mache/ daß ich morgen
Ein neu Geschöpffe bin.

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

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

A Das

www.deutschestextarchiv.de/canitz_gedichte_1700/14

1. Textanalyse → Gedicht!

2. Recherche in der DTABf-Doku:

Suche

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

Gedichte in Gedichtbänden

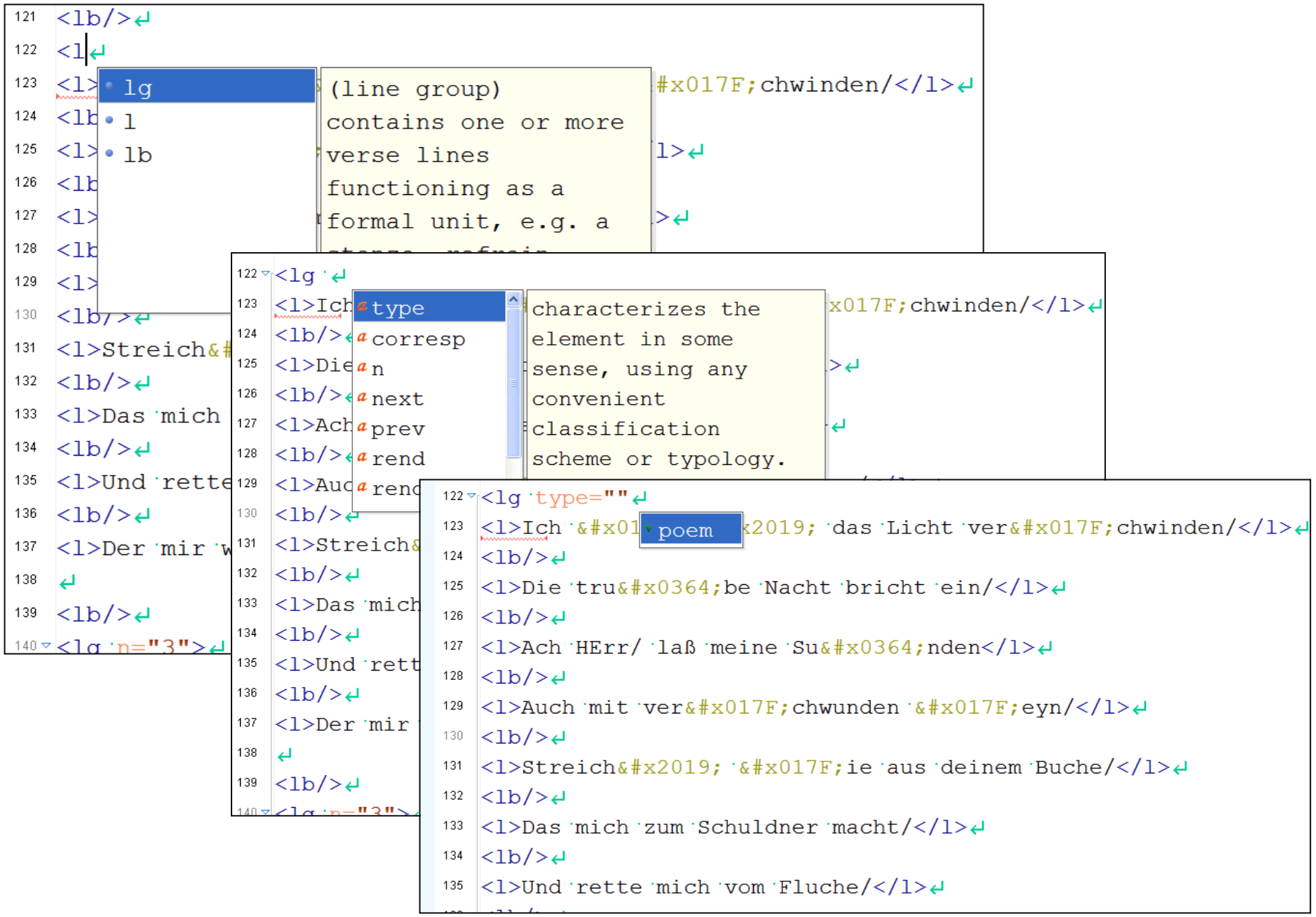
In Gedichtbänden können innerhalb von Gedichten neben dem Gedichttitel und den einzelnen Strophen zusätzliche Textbestandteile vorkommen, z.B. Datumszeilen, Sprecher oder zusätzliche Angaben zur Entstehungsgeschichte des Gedichts. In Gedichtbänden mit solcherlei Gedichten ebenso wie in literarischen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

Auszeichnung von Gedichten in Gedichtbänden (beispielhaft):

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

3. Annotation mithilfe des Schemas

Validierung der XML-Instanzen im oXygen



Tools zur Verwendung des DTABf

- DTA-oXygen-Framework
  - Framework für die Autoransicht im oXygen
  - DTABf-Elemente sind über Buttons zugänglich
  - WYSIWYG-Ansicht des Dokuments
- DTA-Metadatenformular
  - Webformular zur Eingabe von Metadaten
  - TEI-Header wird daraus automatisch erstellt
- **Ediarum** (S. Dumont/M. Fechner)
  - Anpassung der Editionsumgebung für das DTABf
  - Verwendung von ediarum → Erstellung DTABf-konformer Ressourcen
  - Bereits jetzt in verschiedenen BBAW-Vorhaben eingesetzt
  - Cf. Dumont/Haaf 2017

Gliederung des Vortrags

1. Einführung zum Thema
2. Einführung zum DTA-Basisformat-Tagset
3. Anwendung des DTA-Basisformats
4. "Hinter den Kulissen" – Technische Umsetzung
5. Verwendungsszenarien
6. Ausblick

DTABf-Komponenten

1. **Dokumentation:**
   http://www.deutschestextarchiv.de/doku/basisformat

2. **ODD**: http://www.deutschestextarchiv.de/basisformat.odd
3. **Schema (RNG):**
   http://www.deutschestextarchiv.de/basisformat.rng

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

4. **Schematron-Regeln:** ~/basisformat.sch

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

Dokumentation in DITA

- **DITA**: Darwin Information Typing Architecture
- XML-Dokumentationsformat
- Ein Thema (topic) = ein Dokument (z.B. fussnote.dita)
- DITA-Map legt fest, welche *topics* in der Dokumentation landen

- Vielerlei Ausgabeformate:

  Webhelp responsive, Webhelp, PDF, …
- Eigene Anpassungen am Design der Ausgabe möglich

- oXygen-XML-Editor: Support für das DITA-Open-Toolkit

DITA-Map

```
<part href="einfuehrung.dita">
  <topicref href="ziel.dita"/>
  <topicref href="schema.dita"/>
  <topicref href="publikationen.dita"/>
  <topicref href="howto.dita">
    <topicref href="benutzungDTABfSchema.dita"/>
    <topicref href="zurDokumentation.dita"/>
    <topicref href="hilfreicheTools.dita"/>
    <topicref href="grundstrukturDokument.dita"/>
  </topicref>
  <topicref href="introduction_en.dita"/>
</part>
```

Cf. https://github.com/deutschestextarchiv/dtabf/blob/master/documentation/dtabf.ditamap

DITA-Map

```
<part href="einfuehrung.dita">
  <topicref href="ziel.dita"/>
  <topicref href="schema.dita"/>
  <topicref href="publikationen.dita"/>
  <topicref href="howto.dita">
    <topicref href="benutzungDTABfSchema.dita"/>
    <topicref href="zurDokumentation.dita"/>
    <topicref href="hilfreicheTools.dita"/>
    <topicref href="grundstrukturDokument.dita"/>
  </topicref>
  <topicref href="introduction_en.dita"/>
</part>
```

- Individuelle Anpassungen an die Dokumentation möglich
- Nachnutzung der Dokumentation in anderen Projektkontexten möglich (CC-Lizenz)
- Erstellung eigener Derivate

Schema auf Basis von ODD

- ODD-Formalismus (One document does it all)
- Metasprache zur Anpassung der TEI-Guidelines
- TEI-Subset (Modul *tagsdocs*)
- Nutzer müssen nicht das TEI-Schema direkt ändern, sondern dokumentieren Änderungswünsche im ODD
- Transformationsskripte ermöglichen die automatische Erstellung eines TEI-Schemas auf Grundlage des ODD

ODD für DTABf:

```
<moduleRef key="drama" include="actor castList …"/>
<moduleRef key="core" include=" abbr bibl cb head …"/>
```

Cf. http://www.deutschestextarchiv.de/basisformat_all.odd

Schema auf Basis von ODD

```
<elementSpec ident="gap" module="core" mode="change">
  <attList>
    <attDef ident="n" mode="delete"/>
    <attDef ident="reason" mode="change">
      <valList type="closed" mode="replace">
        <valItem ident="insignificant"/>
        <valItem ident="illegible"/>
        <valItem ident="fm"/>
        <valItem ident="lost"/>
      </valList>
    </attDef>
    <attDef ident="quantity" mode="change" usage="opt">
      <datatype minOccurs="1" maxOccurs="unbounded">
        <rng:ref name="data.count"/>
      </datatype>
    </attDef>
    ...
  </attList>
</elementSpec>
```

DTABf – ODD: Spezifikation für das Element <gap> (Exzerpt)

Cf. http://www.deutschestextarchiv.de/basisformat_all.odd

DTA-Basisformat: Schematron

```
<pattern id="teiHeaderElements">
  <rule context="tei:addName | tei:address | tei:addrLine |
    tei:email | tei:biblFull | tei:country | tei:forename |
    tei:genName | tei:measure | tei:msDesc | tei:nameLink |
    tei:publicationStmt | tei:resp | tei:respStmt |
    tei:roleName | tei:surname | tei:titleStmt | tei:title">
    <report test="ancestor::tei:text" role="ERROR">
      [E0001] Element "<name/>" not allowed anywhere within
       element "text".
    </report>
  </rule>
</pattern>
```

Cf. http://www.deutschestextarchiv.de/basisformat.sch

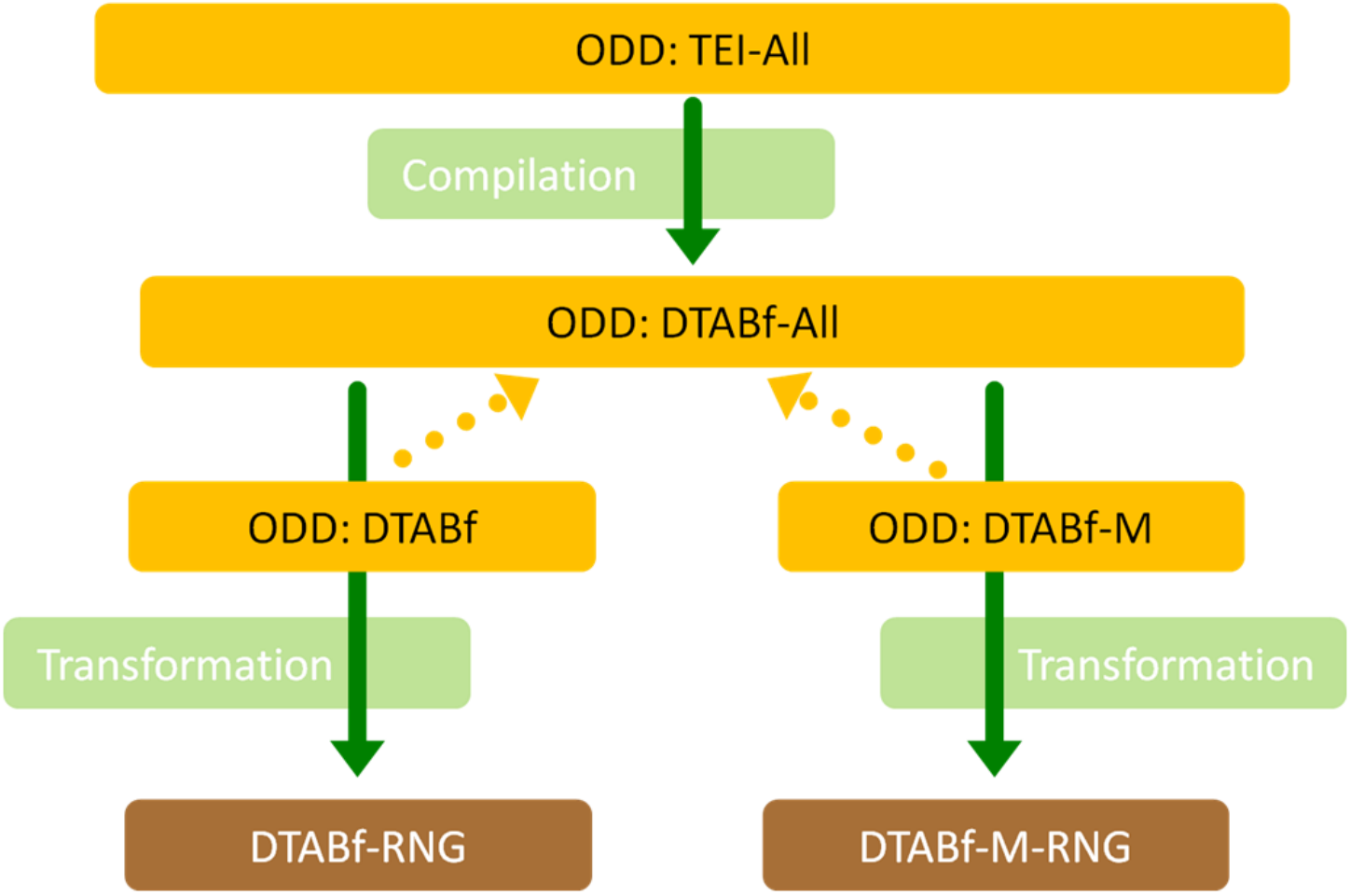
DTABf: Modularisierung

- DTA-Korpora wachsen stetig
  - ➢ speziell Textkuration (in CLARIN-D)
- Anpassungen für neue Texttypen
  - ➢ Manuskripte, Zeitungen, Funeralschriften
- Anpassungen an bislang unbekannte Phänomene
- Konsequenz für das DTABf: Modularisierung!

  a. des Schemas

  b. der Dokumentation

Modularisierung des Schemas durch ODD-Chaining



DTABf: Chaining ODDs -- Beispiel

ODD für DTABf-All:[1]

```
<moduleRef key="transcr" include="fw metamark subst..."/>
<moduleRef key="core" include="add del list p sp ..."/>
```

ODD für DTABf:[2]

```
<moduleRef key="transcr" except="metamark subst"/>
<moduleRef key="core" except="add del ..."/>
```

ODD für DTABf-M:[3]

```
<moduleRef key="transcr"/>
<moduleRef key="core" except="sp ..."/>
```

[1] http://www.deutschestextarchiv.de/basisformat_all.odd
[2] http://www.deutschestextarchiv.de/basisformat.odd
[3] http://www.deutschestextarchiv.de/basisformat_ms.odd

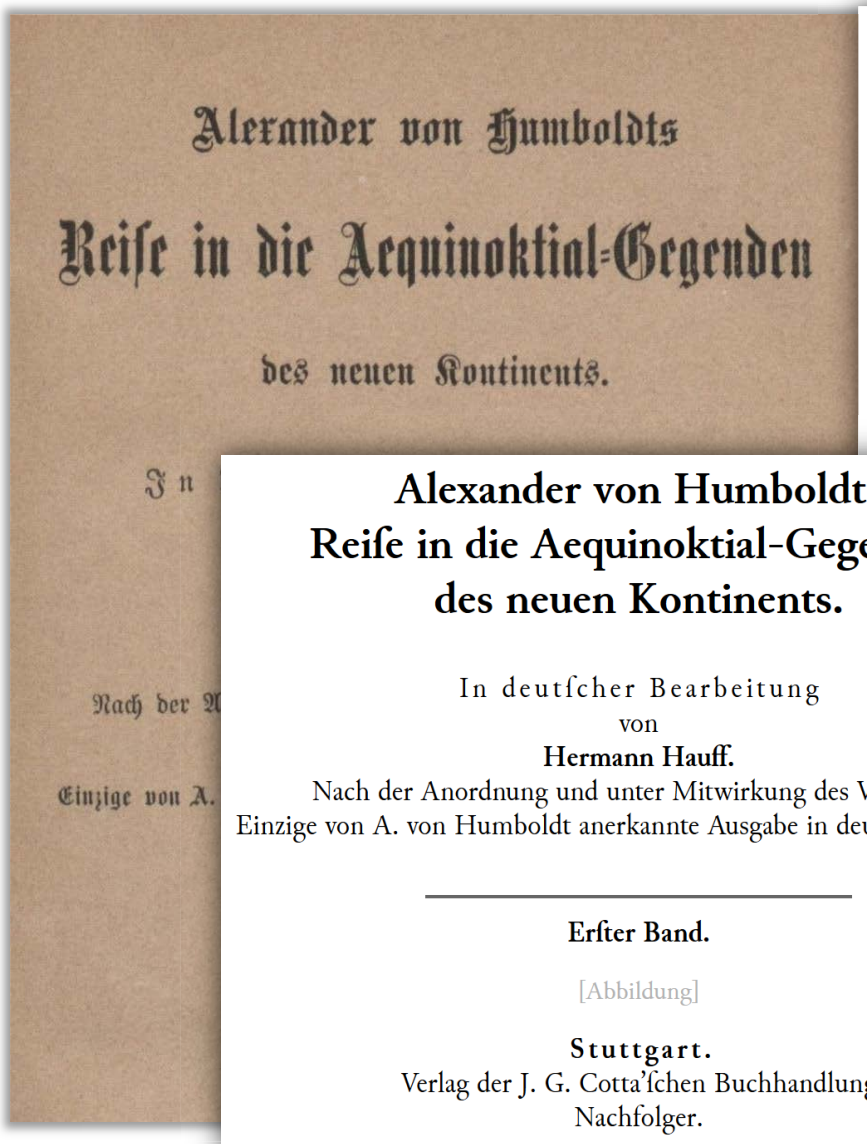
Gliederung des Vortrags

1. Einführung zum Thema
2. Einführung zum DTA-Basisformat-Tagset
3. Anwendung des DTA-Basisformats
4. "Hinter den Kulissen" – Technische Umsetzung
5. Verwendungsszenarien
6. Ausblick

DTABf-Nutzung

- Kooperationsprojekte des DTA
  - Digitalisierung nach DTA-Richtlinien
  - Konvertierung und Integration digitaler Daten
  - z.B. Hamburgischer Correspondent, Blumenbach Edition
- Projekte außerhalb des DTA
  - z.B. Hamburger Schlüsseldokumente, Thüringer Editionsportal
- Projekte innerhalb des Hauses
  - Anpassung von ediarum für das DTABf
  - z.B. Alexander von Humboldt auf Reisen, Marx-Engels-Gesamtausgabe

DTABf-Nutzung, Bsp. 1: Rendering



Alexander von Humboldts
Reiſe in die Aequinoktial-Gegenden
des neuen Kontinents.

In deutſcher Bearbeitung
von
Hermann Hauff.
Nach der Anordnung und unter Mitwirkung des Verfaſſers.
Einzige von A. von Humboldt anerkannte Ausgabe in deutſcher Sprache.

Erſter Band.

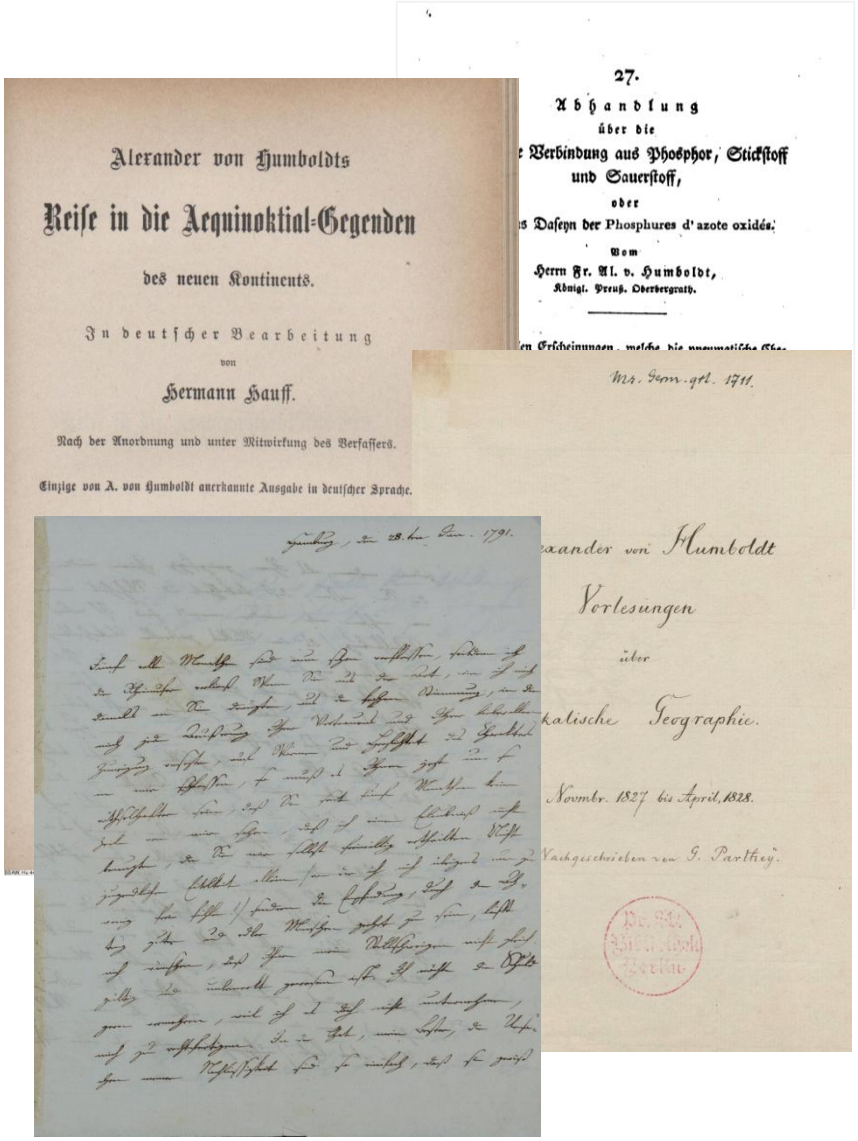
[Abbildung]

Stuttgart.
Verlag der J. G. Cotta'ſchen Buchhandlung
Nachfolger.

```
<TEI>
  <text>
    <front>
      <pb facs="#f0007" n="[III]"/>
      <titlePage type="main">
        <docTitle>
          <titlePart type="main"> <hi
rendition="#b">Alexander von Humboldts<lb/>
Rei&#x017F;e in die Aequinoktial-Gegenden<lb/>
des neuen Kontinents.</hi> </titlePart>
        </docTitle><lb/>
        <byline><hi rendition="#g">In
17F;cher Bearbeitung</hi><lb/>
<docAuthor><hi rendition="#b">Hermann
hi></docAuthor><lb/>
 Anordnung und unter Mitwirkung des
017F;&#x017F;ers.<lb/>
von A. von Humboldt anerkannte Ausgabe
#x017F;cher Sprache.</byline><lb/>
<milestone rendition="#hr"
ction"/>
<titlePart type="volume"> <hi
n="#b">Er&#x017F;ter Band.</hi>
art><lb/>
<figure/><lb/>
<docImprint><pubPlace><hi
n="#b"><hi rendition="#a">Stuttgart</hi>.
```

Bsp. 2: Vernetzung von Daten – AvH-R im DTA

1. Werke AvHs im DTA-Kernkorpus
2. Projekt der HU Berlin: Nachschriften zu Humboldts Kosmos-Vorlesungen
3. DTAE-Projekt: Aufbereitung der unselbständigen Schriften AvHs
4. BBAW-Vorhaben: Alexander von Humboldt auf Reisen

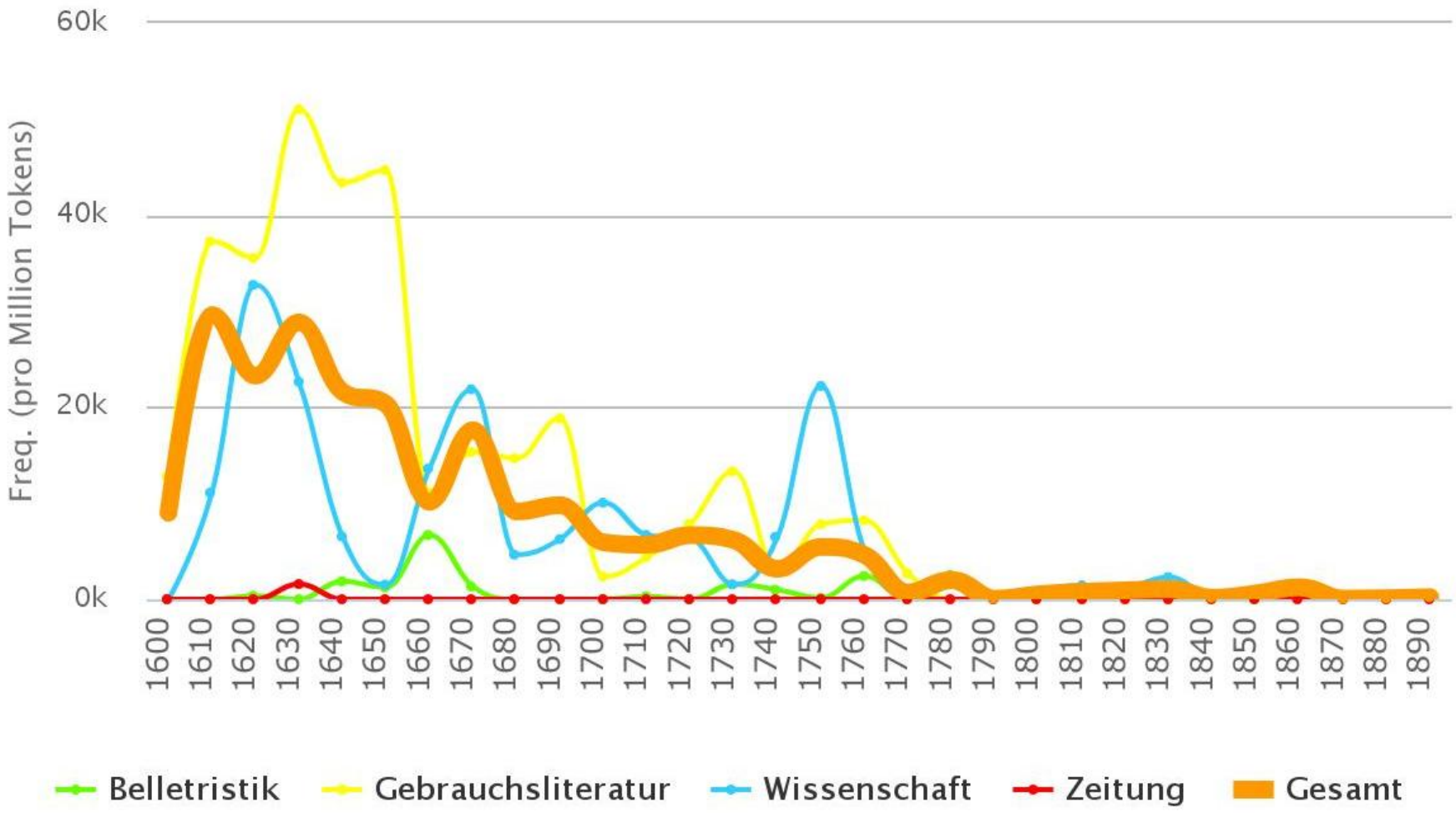


Bsp. 2: Vernetzung von Daten – AvH-R im DTA

8:	[dta:humboldt_aequinoktial04_1859:163]	Man erzählt rührende Fälle, wo	afrikaniſche	Sklaven ihr Leben aufs Spiel ſetzten, um...
9:	[dta:humboldt_aequinoktial04_1859:191]	... Demerary Feuergewehre, und als der Handel mit	amerikaniſchen	Sklaven (Poitos) in Blüte ſtand,...
10:	[dta:humboldt_aequinoktial04_1859:253]	... frei zur Welt kommen, nicht einer des	andern	Sklaven werden.“
11:	[dta:humboldt_aequinoktial04_1859:296]	... auf dem Feſtlande von Spaniſch-Amerika die Zahl der	afrikaniſchen	Sklaven ſo unbedeutend, daß ſie ſich zur...
12:	[dta:humboldt_aequinoktial01_1859:203]	... Sklavenhandel abgeſchafft hat, und dennoch waren die	erſten	Sklaven, die wir aufgeſtellt ſahen, auf...
13:	[dta:humboldt_aequinoktial01_1859:203]	Die zum Verkauf	ausgeſetzten	Sklaven waren junge Leute von fünfzehn bis zwanzig...
14:	[dta:humboldt_aequinoktial01_1859:204]	Der Handel mit	afrikaniſchen	Sklaven, den die ſpaniſchen Geſetze niemals begünſtigt...
15:	[dta:humboldt_aequinoktial01_1859:308]	... behelfen ſich bei ihrer mühſamen Arbeit mit einem	einzigen	Sklaven; ſie brechen mit eigener Hand den...
16:	[dta:humboldt_grenzen_1851:133]	...; allein er würde mir immer ein Haufe	ernährter	Sklaven, nicht eine Vereinigung freier, nur...
17:	[dta:humboldt_aequinoktial02_1859:75]	... ſein Wohlſtand gehoben hatte, kaufte er zwei	ſchwarze	Sklaven, und dies ward ſein Verderben:
18:	[dta:humboldt_aequinoktial02_1859:186]	... wir drei entſprungene Neger einbringen, vor kurzem	gekaufte	Sklaven.
19:	[dta:humboldt_aequinoktial02_1859:207]	... und Indigo erzeugen kann, und daß die	unglücklichen	Sklaven Bauern, Pächter und Grundbeſitzer werden können...
20:	[dta:humboldt_aequinoktial02_1859:229]	In dieſes Bad werfen ſich die	unglücklichen	Sklaven, wenn ſie gegen Sonnenuntergang, mit...
21:	[dta:humboldt_aequinoktial02_1859:266]	... mit der Förderung und dem Schmelzen des Erzes	beſchäftigten	Sklaven beträgt nur 60 bis 70.
22:	[dta:humboldt_reizmittel_1797:19]	... man, einem alten Herkommen gemäß, die	todten	Sklaven, ehe man ſie ins Meer ſenkt...
23:	[dta:humboldt_verhaeltnis_1826:2]	... dem Feſtlande vom ſpaniſchen Amerika die Zahl der	afrikaniſchen	Sklaven verhältnißmäßig zur Sklavenbevölkerung von Braſilien oder vom...
24:	[dta:humboldt_verhaeltnis_1826:2]	... pflanzt und die faſt keine dem afrikaniſchen Gebiet	entriſſene	Sklaven hat.

DDC: "$p=ADJA Sklave" #has[author, /Humboldt/]

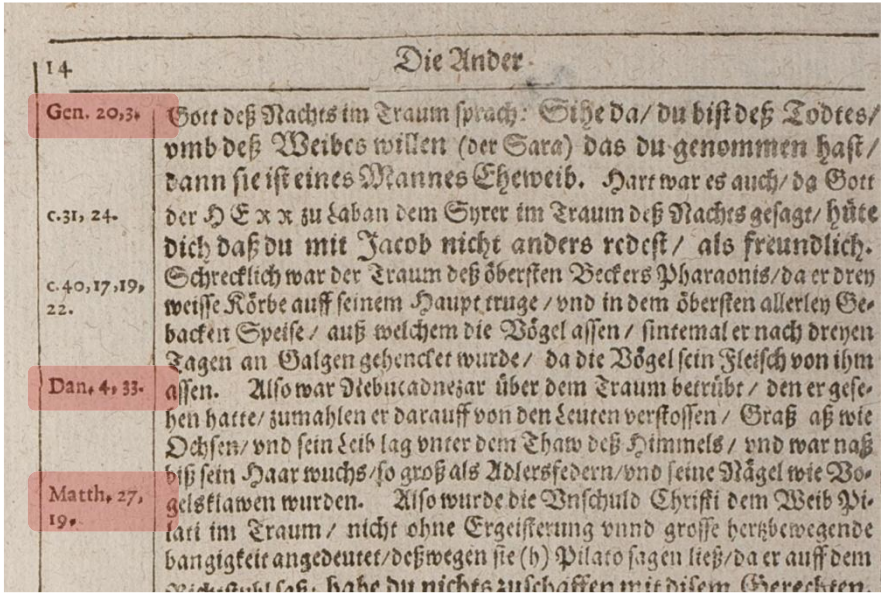
Bsp. 3: Korpusanalyse – Entwicklung von Marginalien



TEI: <note place="left | right">

DDC: $con=/left/ || $con=/right/

Bsp. 3: Korpusanalyse – Entwicklung von Marginalien



14 Die Ander

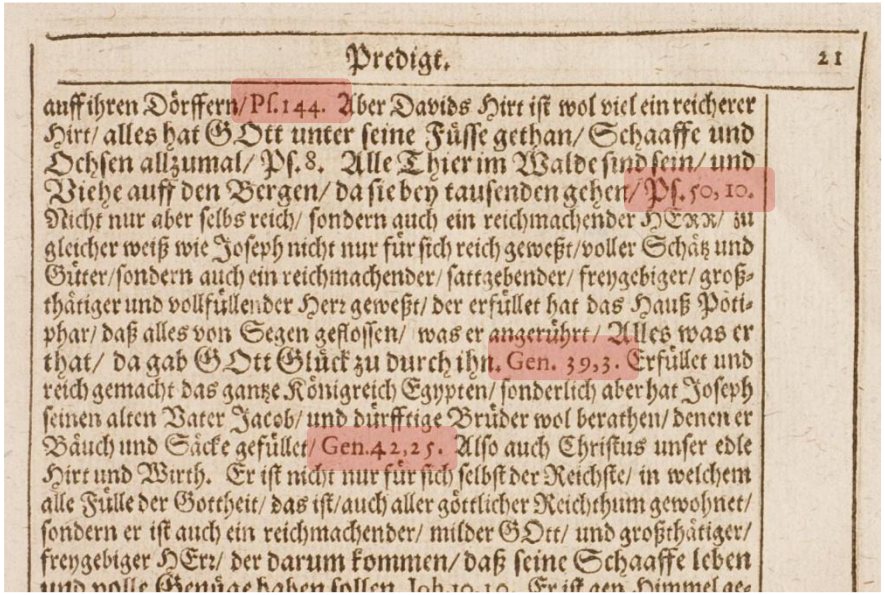
Gen. 20,3. Gott deß Nachts im Traum sprach: Sihe da/ du bist deß Todes/ vmb deß Weibes willen (der Sara) das du genommen hast/ dann sie ist eines Mannes Eheweib. Hart war es auch/ da Gott

c.31, 24. der HErr zu Laban dem Syrer im Traum deß Nachts gesagt/ hüte dich daß du mit Jacob nicht anders redest/ als freundlich.

c.40,17,19, 22. Schrecklich war der Traum deß öbersten Beckers Pharaonis/ da er drey weisse Körbe auff seinem Haupt truge/ vnd in dem öbersten allerley Gebacken Speise/ auß welchem die Vögel assen/ sintemal er nach dreyen Tagen an Galgen gehencket wurde/ da die Vögel sein Fleisch von ihm

Dan. 4, 33. assen. Also war Nebucadnezar über dem Traum betrübt/ den er gesehen hatte/ zumahlen er darauff von den Leuten verstossen/ Graß aß wie Ochsen/ vnd sein Leib lag vnter dem Thaw deß Himmels/ vnd war naß biß sein Haar wuchs/ so groß als Adlersfedern/ vnd seine Nägel wie Vogelsklawen wurden. Also wurde die Vnschuld Christi dem Weib Pi-

Matth. 27, 19. lati im Traum/ nicht ohne Ergeisterung vnnd grosse hertzbewegende bangigkeit angedeutet/ deßwegen sie (h) Pilato sagen ließ/ da er auff dem



Predigt. 21

auffihren Dörffern/ Pf. 144. Aber Davids Hirt ist wol viel ein reicherer Hirt/ alles hat GOtt unter seine Füsse gethan/ Schaaffe und Ochsen allzumal/ Pf. 8. Alle Thier im Walde sind sein/ und Viehe auff den Bergen/ da sie bey tausenden gehen/ Pf. 50, 10. Nicht nur aber selbs reich/ sondern auch ein reichmachender HErr/ zu gleicher weiß wie Joseph nicht nur für sich reich geweßt/ voller Schätz und Güter/ sondern auch ein reichmachender/ sattgebender/ freygebiger/ großthätiger und vollfüllender Herr geweßt/ der erfüllet hat das Hauß Potiphar/ daß alles von Segen geflossen/ was er angerührt/ Alles was er that/ da gab GOtt Glück zu durch ihn. Gen. 39, 3. Erfüllet und reich gemacht das gantze Königreich Egypten/ sonderlich aber hat Joseph seinen alten Vater Jacob/ und dürfftige Brüder wol berathen/ denen er Bäuch und Säcke gefüllet/ Gen. 42, 25. Also auch Christus unser edle Hirt und Wirth. Er ist nicht nur für sich selbst der Reichste/ in welchem alle Fülle der Gottheit/ das ist/ auch aller göttlicher Reichthum gewohnet/ sondern er ist auch ein reichmachender/ milder GOtt/ und großthätiger/ freygebiger HErr/ der darum kommen/ daß seine Schaaffe leben und volle Genüge haben sollen. Joh. 10, 10. Er ist gen Himmel ge-

Dannhauer, Johann Conrad: Catechismus Milch. Bd. 4. Straßburg, 1653[1]

Dannhauer, Johann Conrad: Catechismus Milch. Bd. 9. Straßburg 1672[2]

[1] http://www.deutschestextarchiv.de/dannhauer_catechismus06_1657
[2] http://www.deutschestextarchiv.de/dannhauer_catechismus09_1672

Bsp. 4: Wortuntersuchungen – „frohlocken”

- Duden (1999): 3 Bedeutungen
  1. *„Schadenfreude empfinden”*
  2. *„vor Freude jubeln; jauchzen“*
  3. *„lobsingen (veraltet)”*
- Korpusanalyse zur Verteilung von "frohlocken"
  - DDC: frohlocken with $xpath=/\b**p**\b/ (in Prosa-Paragraphen)
  - DDC: frohlocken with $xpath=/\b**l**\b/ (in Lyrik-Versen)
- Resultat:
  - übliche Verwendung: frohlocken (+ PP, z.B. "über")
  - *nur in der Lyrik*: frohlocken + Dativobj. ("jem.m frohlocken")
  - entspricht Bedeutung 3

Bsp. 4: Wortuntersuchungen – „frohlocken”

„frohlocken” mit Dativ in der Lyrik:

Frolocket eurem Held/ nach Streit hat er jetzt Ruh.

(Mühlpfort 1686: 48)

Mir hůpfen die Berg’, und die Hůgel! Mir frohlocket der Wald!

(Klopstock 1773: 147)

Frohlockt dir eine halbe Welt.

(Zachariae 1764: 254)

Wiehert ihm Rosse! frohlockt ihm, Adler!

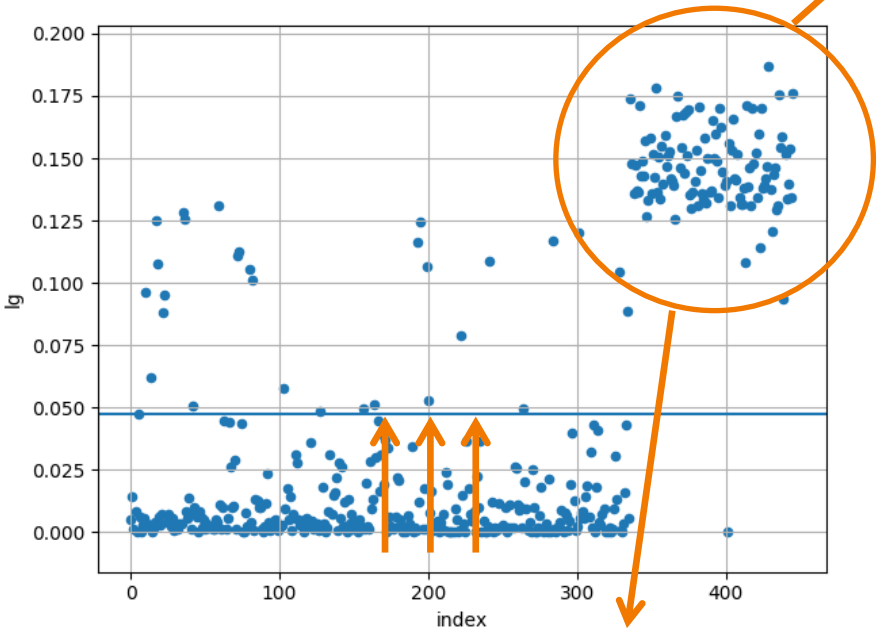
(Kosegarten 1798: 41)

Frohlockt ihm, alle Frommen!

(Candidus 1854: 21)

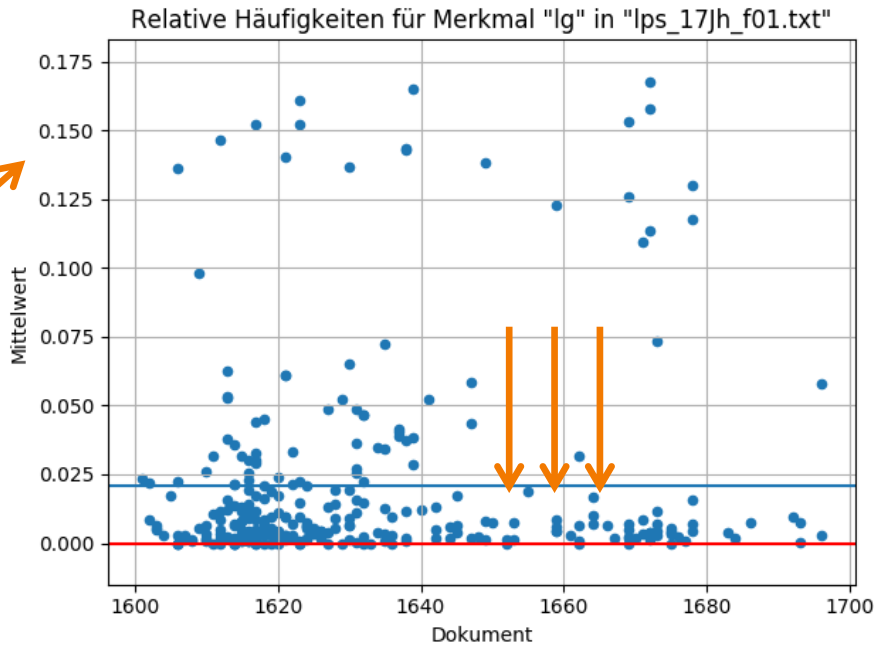
Cf. *Haaf 2016*

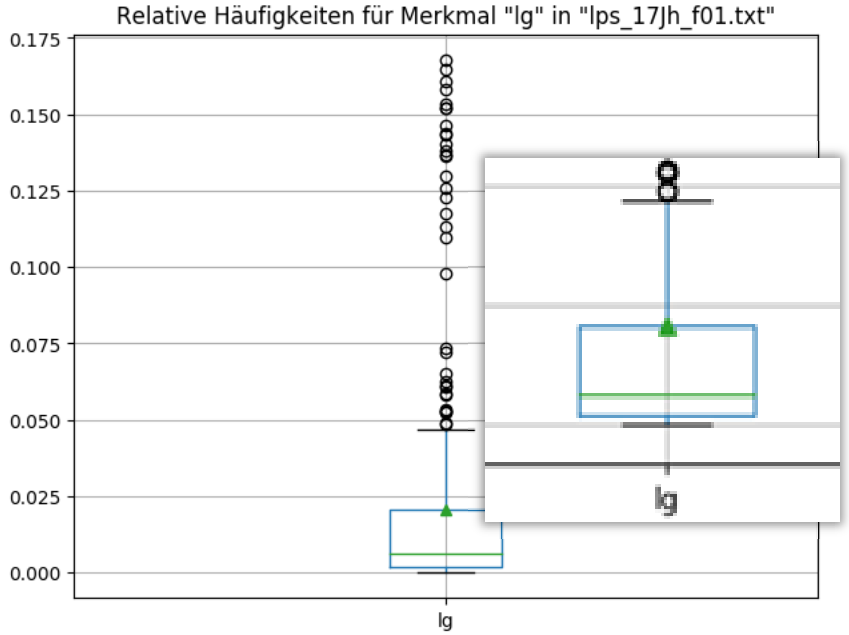
Bsp. 5: Überprüfung des Korpus – Textsorte: "Leichenpredigt"



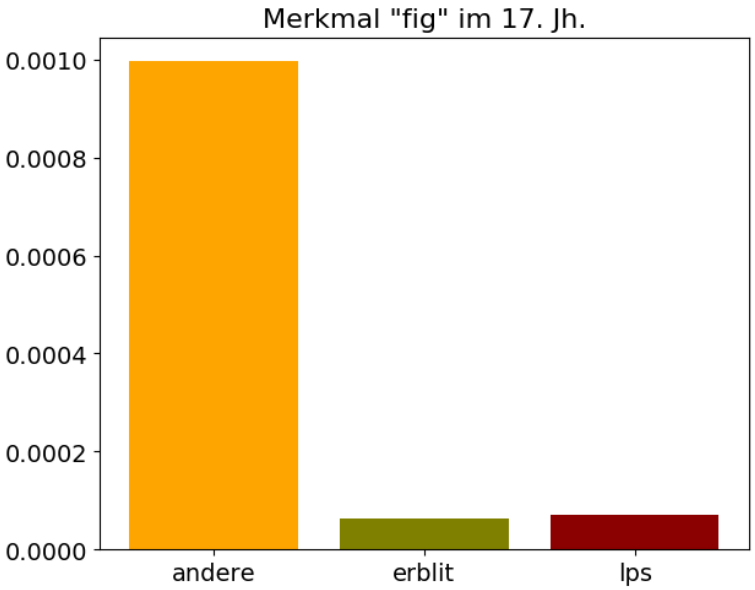
Epicedien Simon Dach

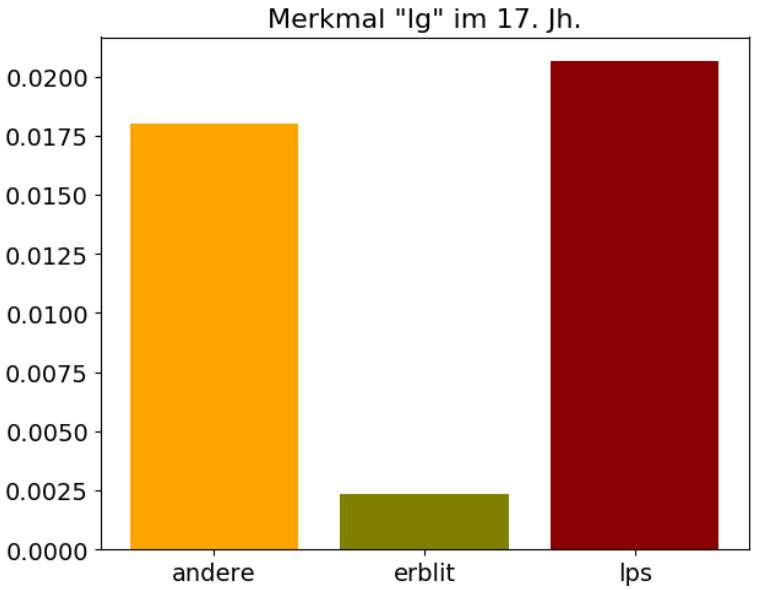
LPs ohne Simon Dach

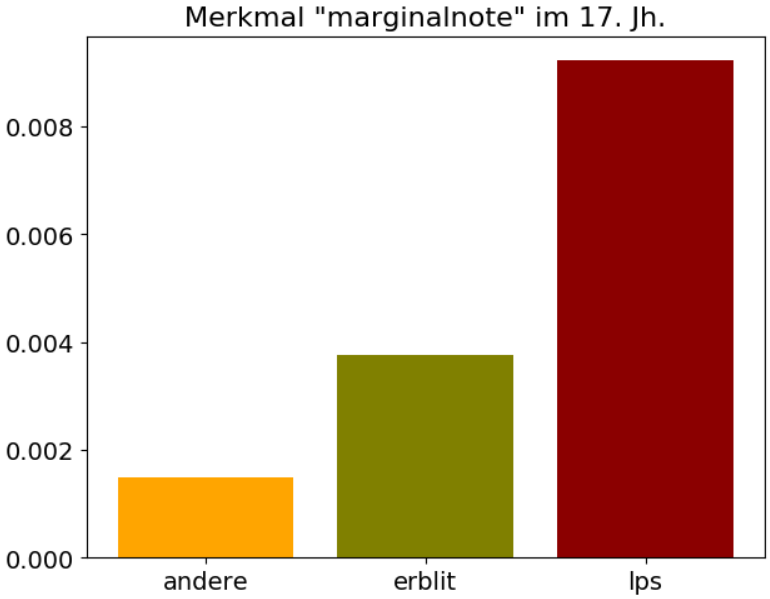


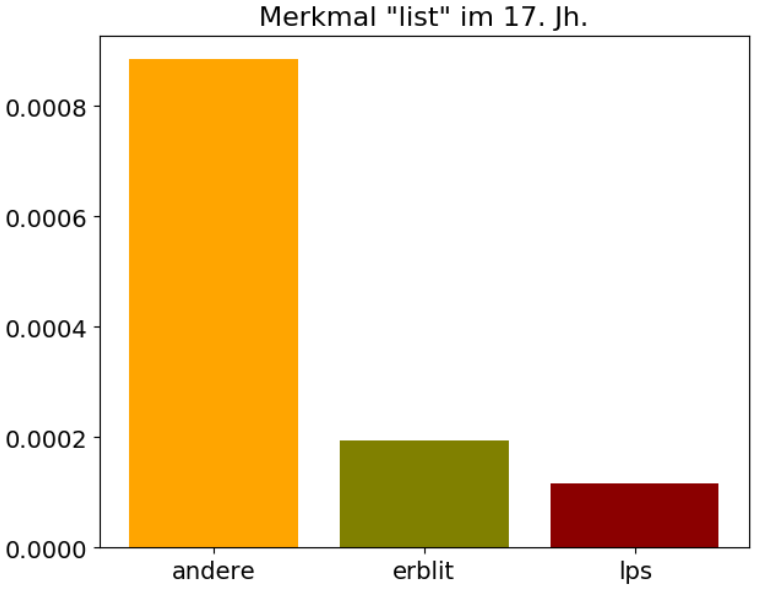


Bsp. 6: Strukturmerkmale im Vergleich



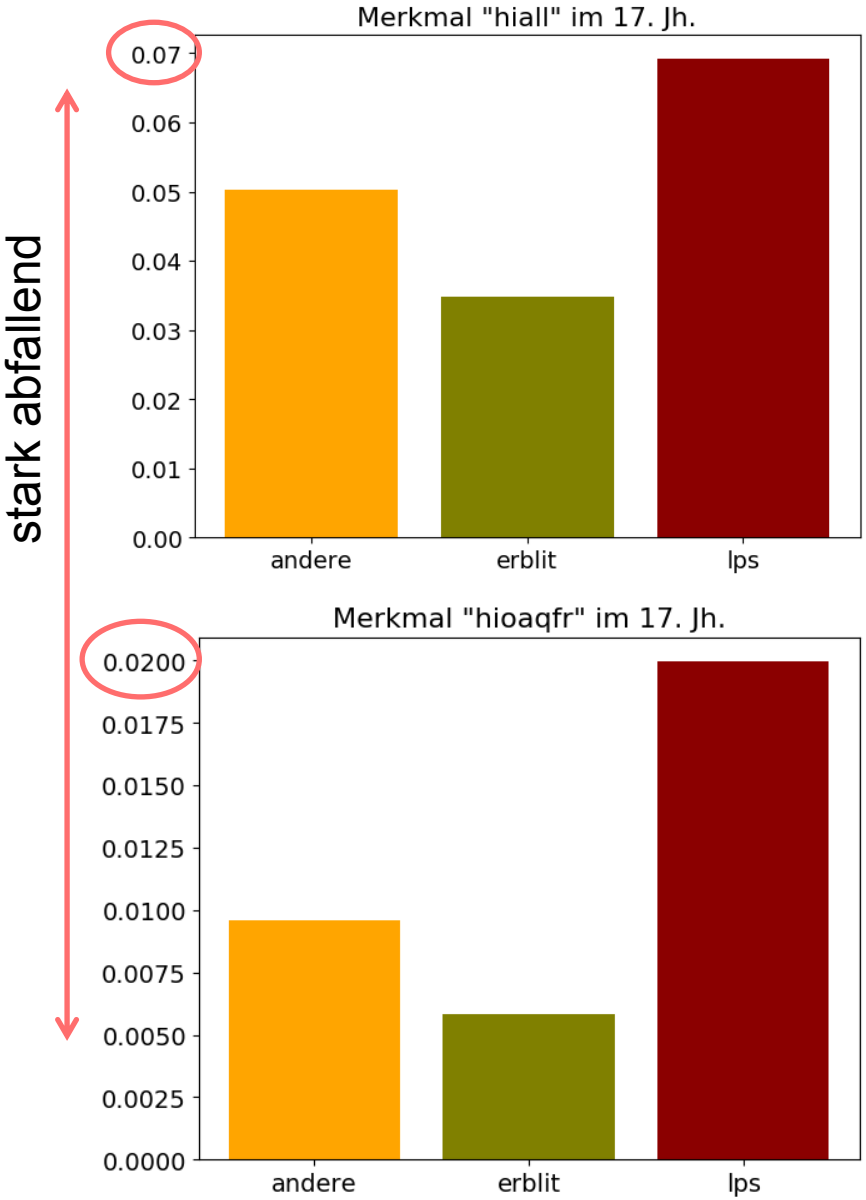






→ relative Häufigkeiten des Merkmals im Verhältnis zur Token-Anzahl

Bsp. 6: Strukturmerkmale im Vergleich – Hervorhebungen



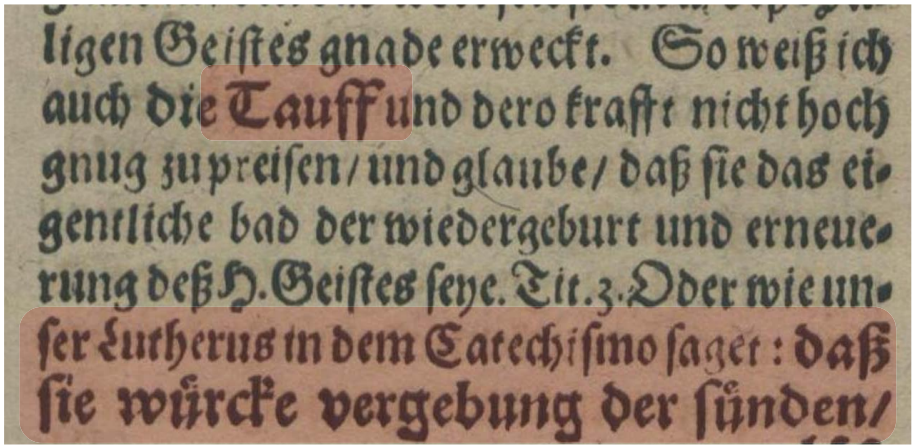
Hervorhebungen allgemein

Hervorhebungen außer Wechsel zur Antiqua und Frakturwechsel, i.e. *Zentrierung, Einrückung, Sperrdruck, Kursivdruck*, …

Bsp. 6: Strukturmerkmale im Vergleich – Hervorhebungen

- Zahlenmäßig häufig in der prosaischen Erbauungsliteratur: Mittel der Intertextualität (vgl. Pfefferkorn 2005) und der Orientierung
  → Unterstützen die Vermittlung der Lehrinhalte strukturell

- **Marginalien:** Hervorhebung von Kernaussagen, bibl. Angaben, Orientierungshilfe
- **Wechsel der Fraktur:** Hervorhebung von Zitaten und Kernaussagen, Orientierungshilfe
- **Wechsel zur Antiqua:** Hervorhebung von lat. Zitaten und bibl. Angaben


ligen Geistes gnade erweckt. So weiß ich
auch die Tauff und dero krafft nicht hoch
gnug zu preisen / und glaube / daß sie das ei-
gentliche bad der wiedergeburt und erneue-
rung deß H. Geistes seye. Tit. 3. Oder wie un-
ser Lutherus in dem Catechismo saget: daß
sie würcke vergebung der sünden /

sie in ihren Zusamenkunfften gesprochen worden: Der HErr segne
vnd behüte dich: der HErr lasse sein Angesicht leuchten über
dir vnd sey dir gnädig: der HErr erhebe sein Angesicht über
dich / vnd gebe dir Friede. So muste ein Versöhnopffer / ein Fluch-
opffer vorgehen / sonderlich auff das hohe Versöhnungsfest / da der versöh-
Num. 6, 24 25. 26.
Lev. 18, 25.

Gliederung des Vortrags

1. Einführung zum Thema
2. Einführung zum DTA-Basisformat-Tagset
3. Anwendung des DTA-Basisformats
4. "Hinter den Kulissen" – Technische Umsetzung
5. Verwendungsszenarien
6. Ausblick

Status

- DTABf zugrunde gelegt für *TEI Simple*
- Empfohlenes Format (unter anderen):
  - für die Auszeichnung historischer Korpora (CLARIN 2012)
  - als Archivformat für literaturwissenschaftliche Editionen (DFG 2016)
  - als Annotationsformat für linguistische Korpora (DFG 2014)

- Präsenz auf github
  - Einladung zur Nachnutzung des Formats/der Doku/der Skripte
  - Teilhabe an Diskussionen
  - https://github.com/deutschestextarchiv/dtabf

Anstehende Aufgaben

- Pflege und Weiterentwicklung im Rahmen von CLARIN-D
  - Besonders jetzt Fokus auf Manuskripten (DTABf-M)
- Neue Entwicklungen in Parthenos
  - Dokumentation in Englisch
  - Überarbeitung DTAoX
  - etc.
- Schematron Quick Fixes?

PARTHENOS

Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

- Einrichtung einer Steuerungsgruppe mit externen Beteiligten

Anhang: Bibliographische Hinweise (1)

DTABf: Dokumentation und weitere Literatur

- http://www.deutschestextarchiv.de/doku/basisformat
- http://www.deutschestextarchiv.de/doku/basisformat/publikationen.html

DFG- und CLARIN-Empfehlungen des DTABf:

- bibliographisch zusammengestellt unter:
  http://www.deutschestextarchiv.de/doku/basisformat/ziel.html

Erwähnte Projekte:

- Alexander von Humboldt auf Reisen: http://edition-humboldt.de
- DTAE-Projekte: http://www.deutschestextarchiv.de/doku/textquellen...
  - Hamburgischer Korrespondent: ...#correspondent
  - Hidden Kosmos: ...#avhkv
  - Alexander von Humboldts unselbstständige Schriften: ...#avh

Anhang: Bibliographische Hinweise (2)

Dumont, Stefan; Fechner, Martin: Bridging the Gap: Greater Usability for TEI encoding. In: Journal of the Text Encoding Initiative [Online], Issue 8 | 2014-2015. URL: http://jtei.revues.org/1242

Dumont, Stefan/Susanne Haaf: *Die vernetzte Edition*. Vortrag im Rahmen des Workshops "Editionsportale" an der Universität Jena, Historisches Institut, 3./4. August 2017. Folien: http://www.deutschestextarchiv.de/files/DumontHaaf_VernetzteEdition_Jena2017.pdf

Unsworth, John: Computational Work with Very Large Text Collections. Interoperability, Sustainability, and the TEI. In: Journal of the Text Encoding Initiative 1 (2011), http://jtei.revues.org/215.

Pfefferkorn, Oliver: Übung der Gottseligkeit. Die Textsorten Predigt, Andacht und Gebet im deutschen Protestantismus des späten 16. und 17. Jahrhunderts. Frankfurt/Main 2005 (= Deutsche Sprachgeschichte 1)

Anhang: Verwendete Korpora, Bsp. 5 & 6 (Folie 52–54)

- Grundlage: Kernkorpus des Deutschen Textarchivs
  - 1600 Werke, 17.—19. Jh., verschiedene Disziplinen
  - Annotation nach DTA-Basisformat (TEI-P5)
  - http://www.deutschestextarchiv.de
- Erbauungsliteratur Prosa
  - 17. Jh.: 25 Bände (10 Autoren, 10.501 Seiten)
- Funeralschriften
  - 17. Jh.: 334 Schriften (14.316 Seiten)
- Referenzkorpus
  - 17. Jh.: 187 Bände (60.798 Seiten)

Vielen Dank!

Dokumentation:
http://www.deutschestextarchiv.de/doku/basisformat

Kontakt:
haaf@bbaw.de