

# Plattform des DTA: Texte kuratieren und nachnutzen

Alexander Geyken

Team: Matthias Boenig, Alexander Geyken, Susanne Haaf,  
Bryan Jurish, Christian Thomas, Frank Wiegand

CLARIN-Zentrum BBAW  
[clarin.bbaw.de](http://clarin.bbaw.de), [deutschestextarchiv.de](http://deutschestextarchiv.de)

2.2. Hamburg

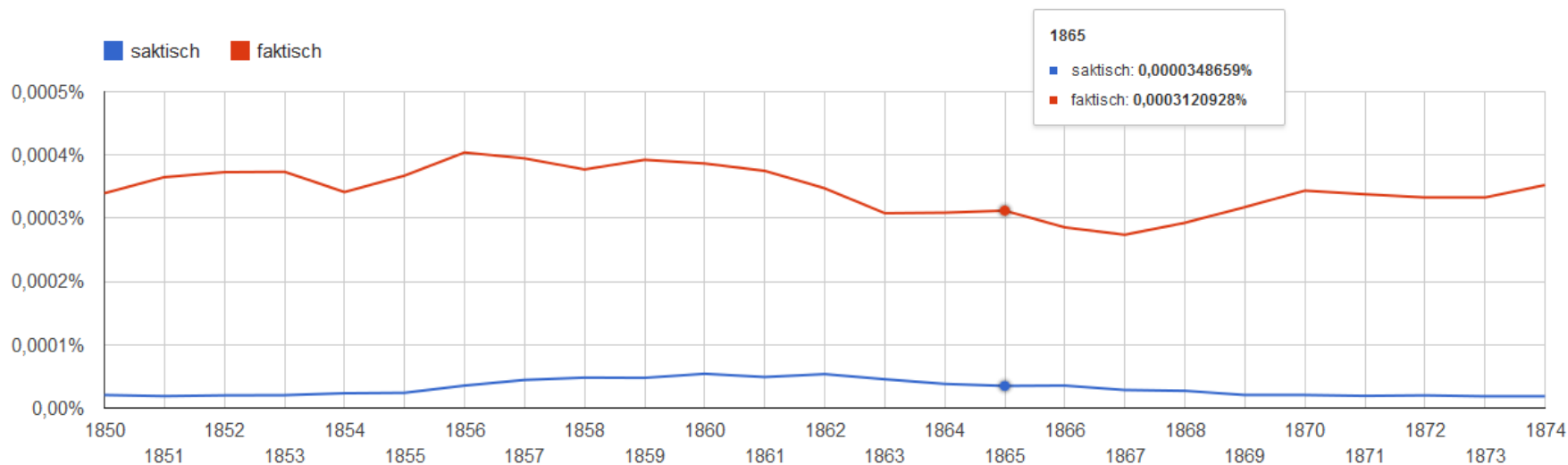
- Wie können wir traditioneller arbeitenden Geisteswissenschaftler\*innen die Nutzung digitaler Methoden nahebringen?
- Welche Möglichkeiten haben wir, die 'digitalen Bedarfe' in den geisteswissenschaftlichen Forschungsgemeinschaften zu ermitteln?
- Wie können Beratung und Dissemination konkret in Forschung und Lehre umgesetzt werden?
- Wie können wir die Nachhaltigkeit von Beratung und Dissemination im Bereich der Digital Humanities sichern?

1. Digitale Historische Quellen: Ausgangspunkt
2. Deutsches Textarchiv-Überblick
3. Text kuratieren und integrieren in das DTA
4. (Recherchieren im DTA)
5. Nachnutzung
  - Mehrwert Linguistische Annotation
  - Mehrwert Download-Pakete
6. Perspektiven des DTA für die digitalen Geisteswissenschaften

# 1. Ausgangspunkt: Heterogene Ressourcen



# 1. Ausgangspunkt: Fehlerhafte Texte



Pfeifer: EtymWB (via DWDS): [...] **faktisch** Adj. ‘tatsächlich, wirklich’ (**Ende 18. Jh.**), meist adverbial gebraucht [...]

frühester Beleg in GoogleBooks: 1729, <http://bit.ly/15pSELY>  
(transkribiert als "saktisch", +Metadatenfehler: Buch ist von 1918...)

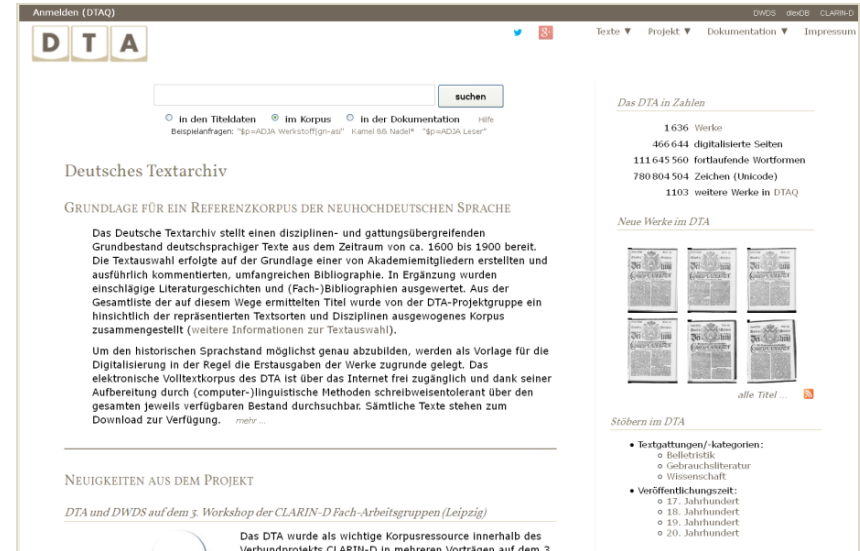
frühester Beleg "factifch" im DTA: 1790, <http://bit.ly/1dGxKbj>

Ziel: „...einen **disziplinenübergreifenden** Kernbestand deutschsprachiger Texte aus der Zeit von ca. 1650 bis 1900 nach den **Erstausgaben** zu digitalisieren und als **linguistisch annotiertes Volltextkorpus** im Internet bereitzustellen.“

(DFG-Projekt DTA: **2007-2016**,  
[www.deutschestextarchiv.de](http://www.deutschestextarchiv.de))

# 2. DTA: Aufbau des Kernkorpus

- Zeitraum: 1650-1900
- 1500 Werke, ca. 100 Millionen Textwörter
- Interoperabel (DTABf)
- durchsuchbar



Anmelden (DTAQ) DWDS BreCB CLARIN-D

DTA

Suchen

in den Titeldaten Im Korpus in der Dokumentation Hilfe

Beispielanfragen: "sp=ACDA Werkstoffhigh-art" "Kamel B&B Nadel" "sp=ACDA Leser"

### Deutsches Textarchiv

GRUNDLAGE FÜR EIN REFERENZKORPUS DER NEUHOCHDEUTSCHEN SPRACHE

Das Deutsche Textarchiv stellt einen disziplinen- und gattungsübergreifenden Grundbestand deutschsprachiger Texte aus dem Zeitraum von ca. 1600 bis 1900 bereit. Die Textauswahl erfolgte auf der Grundlage einer von Akademikern erstellten und ausführlich kommentierten, umfangreichen Bibliographie. In Ergänzung wurden einschlägige Literaturgeschichten und (Fach-)Bibliographien ausgewertet. Aus der Gesamtliste der auf diesem Wege ermittelten Titel wurde von der DTA-Projektgruppe ein hinsichtlich der repräsentierten Textsorten und Disziplinen ausgewogenes Korpus zusammengestellt (weitere Informationen zur Textauswahl).

Um den historischen Sprachstand möglichst genau abzubilden, werden als Vorlage für die Digitalisierung in der Regel die Erstausgaben der Werke zugrunde gelegt. Das elektronische Volltextkorpus des DTA ist über das Internet frei zugänglich und dank seiner Aufbereitung durch (computer-)linguistische Methoden schreibweiseintolerant über den gesamten jeweils verfügbaren Bestand durchsuchbar. Sämtliche Texte stehen zum Download zur Verfügung. [mehr ...](#)

NEUIGKEITEN AUS DEM PROJEKT

DTA und DWDS auf dem 3. Workshop der CLARIN-D Fach-Arbeitsgruppen (Leipzig)

Das DTA wurde als wichtige Korpusressource innerhalb des Verbundprojekts CLARIN-D in mehreren Vorträgen auf dem 3.

Das DTA in Zahlen

- 1636 Werke
- 466.644 digitalisierte Seiten
- 111.645.560 fortlaufende Wortformen
- 780.804.504 Zeichen (Unicode)
- 1103 weitere Werke in DTAQ

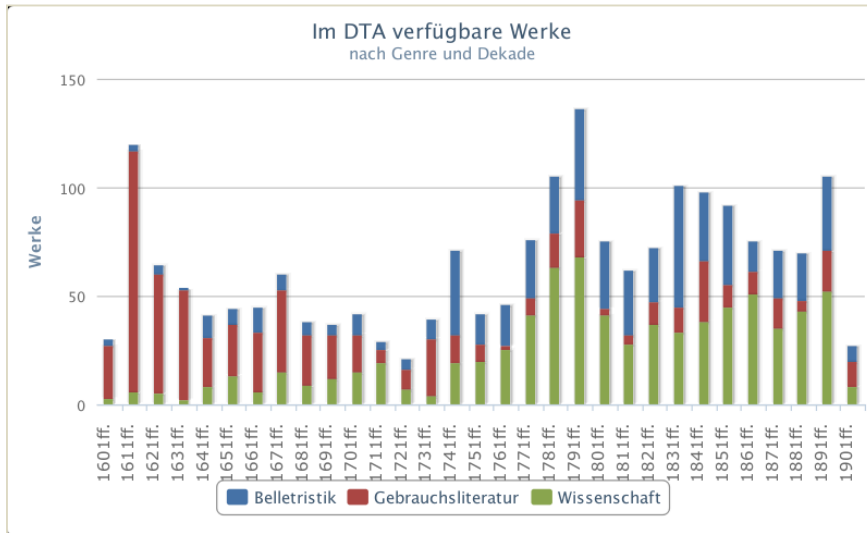
Neue Werke im DTA



alle Titel ...

Stöbern im DTA

- Textgattungen/-kategorien:
  - Belletristik
  - Gebrauchsliteratur
  - Wissenschaft
- Veröffentlichungszeit:
  - 17. Jahrhundert
  - 18. Jahrhundert
  - 19. Jahrhundert
  - 20. Jahrhundert




Deutsches Textarchiv - F. X

www.deutschestextarchiv.de/book/show/forster\_reise01\_1778

Forster, Georg: Johann Reinhold Forster's [...] Reise um die Welt. Bd. 1. Berlin, 1778.

BIBLIOGRAPHISCHE ANGABEN

URN: urn:nbn:de:koebv:b4-200905191395  
 Titel: Johann Reinhold Forster's [...] Reise um die Welt während den Jahren 1772 bis 1775  
 Untertitel: In dem von seiner itzregierenden Großbritannischen Majestät auf Entdeckungen ausgesandten und durch den Captain Cook geführten Schiffe the Resolution unternommen  
 Autor/in: Georg Forster (GND, Wikipedia, ADB/NDB)  
 Erscheinungsjahr: 1778  
 Verlag: Haude und Spener  
 Ort: Berlin  
 Band: Erster Band (Bd. 1 von 2)  
 Auflage: 1. Auflage  
 Bibliothek: Staatsbibliothek zu Berlin - Preußischer Kulturbesitz  
 Signatur: SBB-PK, 4 Pt 1275-1<b> R

ZUGEHÖRIGE WERKE

- Forster, Georg: Johann Reinhold Forster's [...] Reise um die Welt. Bd. 2. Berlin, 1780.

Suche im Werk

Hilfe

Ansichten für dieses Werk

- Text-Bild-Ansicht
- alle Faksimiles
- DTAQ (Qualitätssicherung)

Download

XML (TEI P5) - HTML - Text  
 TCF (text annotation layer)  
 TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)

Metadaten

TEI-Header · CMDI · Dublin Core

Statistiken

Scans: 514  
 Zeichen: ca. 1.178.245  
 Tokens: ca. 168.403  
 Oberflächentypen: ca. 19.738

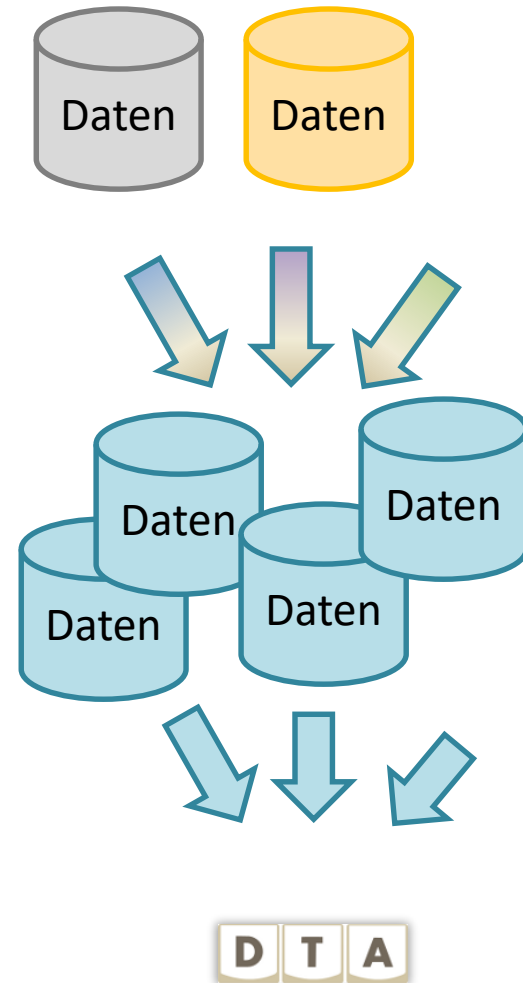
Wortwölkchen

Projekte/Forschende/Studierende,  
die ihre Daten ...

- entsprechend den DTA-Richtlinien erstellen (damit interoperabel sind)
- Im DTA korrigieren
- im DTA veröffentlichen

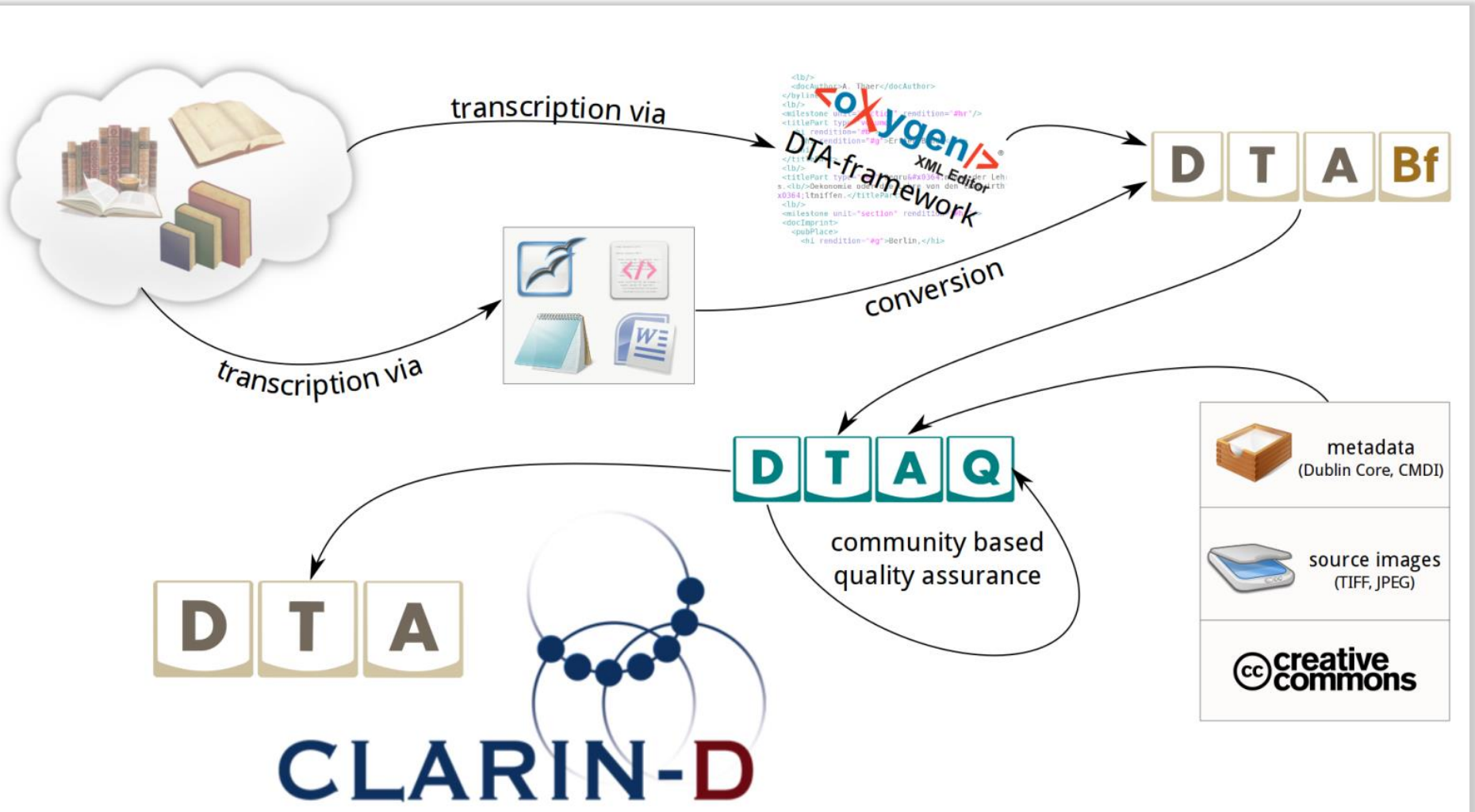
Voraussetzung für die Texte:

- Hohe Datenqualität
- Bilder verfügbar
- Rechte ausreichend





# 3. Integration externer Texte in das DTA



## DTA-BASISFORMAT

Suche

- ❖ TEI-Format für die Auszeichnung von historischen Texten
- ❖ Schema & Dokumentation
- ❖ Ziel: Interoperabilität gewährleisten (Ambiguität vermeiden)
- ❖ DFG-Empfehlung
- ❖ Literatur:  
<https://sprache.hypotheses.org/147>



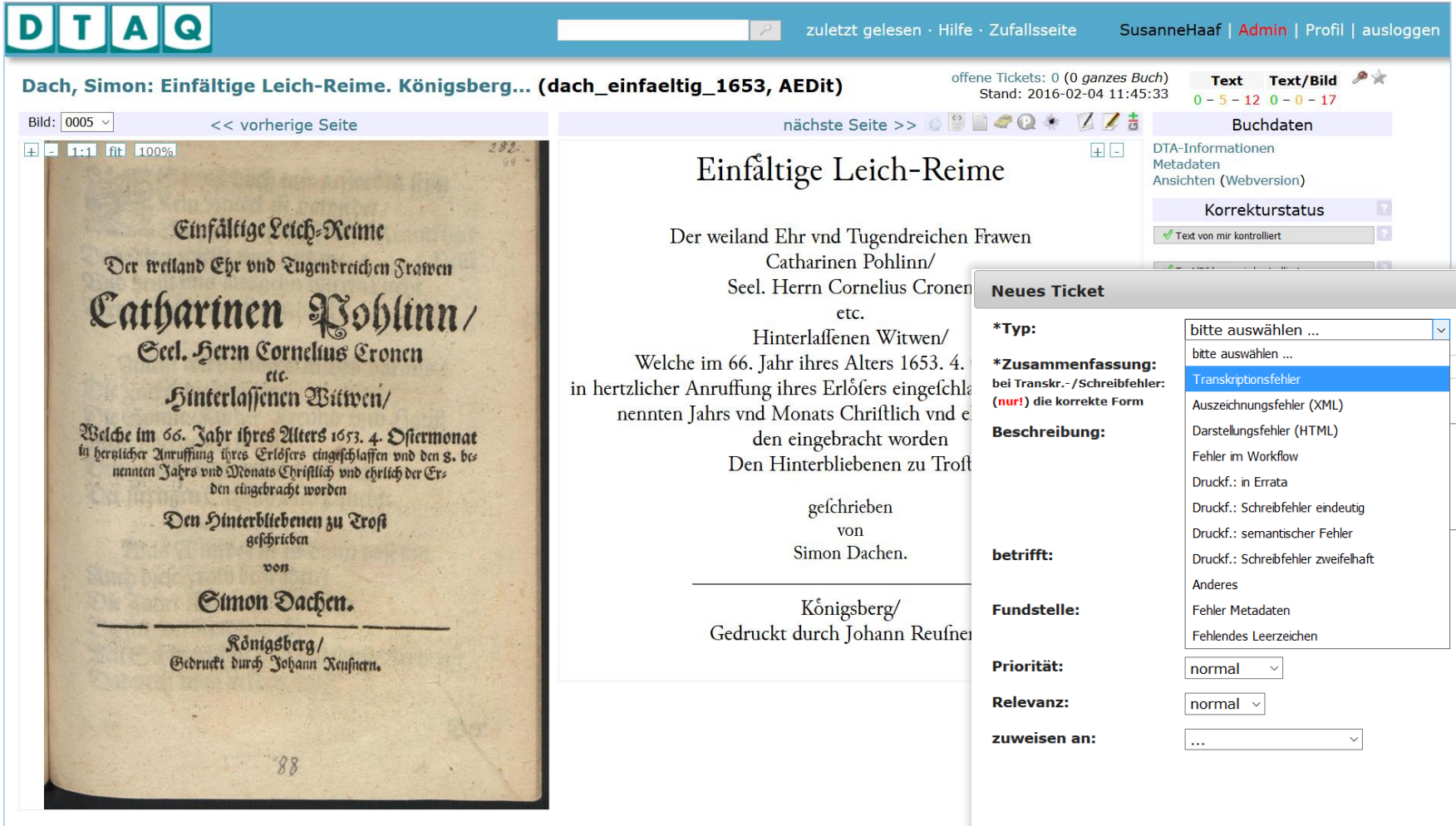
**Level 1: notwendig:** <cb/>, <list>, <lg>, <note>

**Level 2: empfohlen:** <choice>, <fw>, <lb/>

**Level 3: fakultativ:** <foreign>, <persName>

*Level 4: unzulässig:* <ab>, <div1>, <g>

- unterschiedliche editorische Erschließungstiefen
  - DTA-Kernkorpus: bis Level 2
- Alternativen für unzulässige Elemente



The screenshot shows the DTAQ web interface. At the top, there is a navigation bar with the DTAQ logo, a search bar, and user information: "zuletzt gelesen · Hilfe · Zufallsseite" and "SusanneHaaf | Admin | Profil | ausloggen". Below this, the document title is "Dach, Simon: Einfältige Leich-Reime. Königsberg... (dach\_einfaeltig\_1653, AEDit)". The document content is displayed in a large font, with a left sidebar showing a thumbnail of the document page and a right sidebar with "Buchdaten" and "DTA-Informationen". A "Neues Ticket" dialog box is open in the foreground, showing a list of error types with "Transkriptionsfehler" selected. The dialog includes fields for "Typ:", "Zusammenfassung:", "Beschreibung:", "betrifft:", "Fundstelle:", "Priorität:", "Relevanz:", and "zuweisen an:". Buttons for "Ticket erstellen" and "Abbrechen" are at the bottom.

❖ Fehler melden (Ticketing System)

Einfältige Leich-Reime  
Der weiland Ehr vnd Tugendreichen Frawen  
**Catharinen Pohlinn/**  
Seel. Herrn Cornelius Cronen

etc.  
Hinterlassenen Witwen

Welche im 66. Jahr ihres Alters 1653. 4. Ostermonat  
in hertzlicher Anruffung ihres Erlösers eingeschlaffen  
nennten Jahrs vnd Monats Christlich vnd ehrlich der Erden  
den eingebracht worden

Den Hinterbliebenen zu Trost  
geschrieben

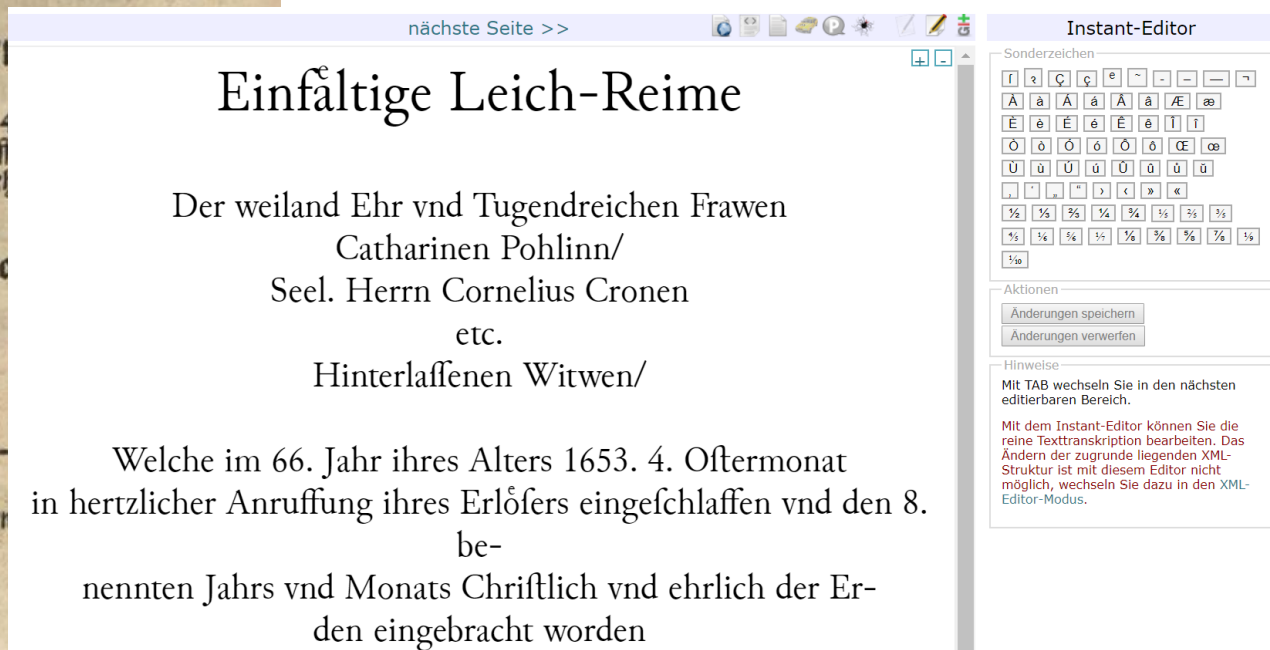
von

**Simon Dachen.**

Königsberg/  
Gedruckt durch Johann Kufner

88

- ❖ Korrektur/Annotation im Text-Modus „Instant-Editor (seitenweise) – xml wird „gekapselt“



nächste Seite >>

## Einfältige Leich-Reime

Der weiland Ehr vnd Tugendreichen Frawen  
Catharinen Pohlinn/  
Seel. Herrn Cornelius Cronen  
etc.  
Hinterlassenen Witwen/

Welche im 66. Jahr ihres Alters 1653. 4. Ostermonat  
in hertzlicher Anruffung ihres Erlösers eingeschlaffen vnd den 8.  
be-  
nennten Jahrs vnd Monats Christlich vnd ehrlich der Er-  
den eingebracht worden

**Instant-Editor**

Sonderzeichen

ſ	z	ç	ç	ë	~	-	-	-	-
À	à	Á	á	Â	â	Æ	æ		
É	é	Ê	ê	Ë	ë	Ï	ï		
Ô	ô	Ó	ó	Ö	ö	Œ	œ		
Û	û	Ü	ü	Û	ü	Û	ü		
¼	½	¾	¼	¾	½	¾	¾		
⅓	⅔	⅓	⅔	⅓	⅔	⅓	⅔		
⅙									

Aktionen

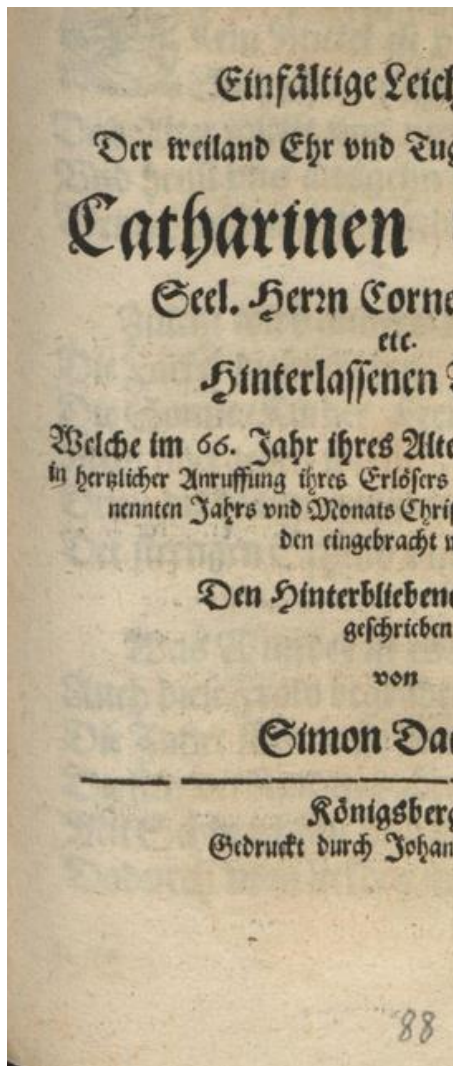
Anderungen speichern

Anderungen verwerfen

Hinweise

Mit TAB wechseln Sie in den nächsten editierbaren Bereich.

Mit dem Instant-Editor können Sie die reine Texttranskription bearbeiten. Das Ändern der zugrunde liegenden XML-Struktur ist mit diesem Editor nicht möglich, wechseln Sie dazu in den XML-Editor-Modus.



## ❖ Korrektur/Annotation im XML-Modus (seitenweise)

```

1 <TEI xmlns="http://www.tei-c.org/ns/1.0"><text>
  <front><pb facs="#f0005"/>
2 <titlePage type="main"><docTitle><titlePart
  type="main">Einfältige Leich-Reime</titlePart><lb/>
3 <titlePart type="desc">Der weiland Ehr vnd
  Tugendreichen Frawen<lb/>
4 Catharinen Pohlinn/<lb/>
5 Seel. Herrn Cornelius Cronen<lb/>
6 etc.<lb/>
7 Hinterlassenen Witwen/<lb/>
8 <lb/>
9 Welche im 66. Jahr ihres Alters 1653. 4.
  Oftermonat<lb/>
10 in hertzlicher Anruffung ihres Erlofers
  eingefchlaffen vnd den 8. be-<lb/>
11 nennten Jahrs vnd Monats Christlich vnd ehrlich der
  Er-<lb/>
12 den eingebracht worden<lb/>
13 Den Hinterbliebenen zu Trost</titlePart>
  </docTitle><lb/>
14 <byline>gefchrieben<lb/>
15 von<lb/>
16 <lb/>
17 <docAuthor>Simon Dachen</docAuthor>.</byline><lb/>
18 <milestone rendition="#hr" unit="section"/><lb/>
19 <docImprint><pubPlace>Königsberg</pubPlace>/<lb/>
20 Gedruckt durch <publisher>Johann Reufnern</publisher>.
  </docImprint></titlePage></front><body><lb/>
21 </body></text></TEI>
  
```

Sonderzeichen

ſ	ꝛ	ç	ç	e	~	-	-
—	—						
À	à	Á	á	Â	â	Æ	æ
È	è	É	é	Ê	ê	Ï	ï
Ò	ò	Ó	ó	Ô	ô	Œ	
œ							
Ù	ù	Ú	ú	Û	û	Ü	ü
,	'	"	"	>	<	»	«
½	⅓	⅔	¼	¾	⅕	⅙	
⅔	⅘	⅙	⅚	⅞	⅛	⅜	
⅞	⅞	⅞	⅞				

Markup

Linie	<lb>	<cb>	LR hor.
LR vert.			
Organisation	Person	Ort	
Druckfehler	Normalisierung		
Abkürzung	<gap>	<supplied>	
#b	#g	#i	#K
#sub	#sup		
#aq	#fr	#n	#blue
#red			
-et	-c-	-right	

[http://www.deutschestextarchiv.de/dtaq/book/show/dach\\_einfaeltig\\_1653](http://www.deutschestextarchiv.de/dtaq/book/show/dach_einfaeltig_1653)

## Zwei kleine Beispiele ...

Beispiel 1: Einzeltexte beitragen

Beispiel 2: Briefe vernetzen

- im Rahmen einer Kolumne "Braugeschichte digital" habe ich im Newsletter der Gesellschaft für die Geschichte des Brauwesens e.V. ([www.ggb-berlin.de](http://www.ggb-berlin.de)) das Projekt DTA vorgestellt  
(Marcus Schmitz, 03/2017)
- Idee: Zeitschriftenbeitrag über das Dortmunder Adambier aus dem Jahre 1869) -> DTABf



## Adambier

### Dortmunder Adambier.

(Aus des Herausgebers Arbeit über deutsche Lokalbiere.)

Als wir im Dezember 1862 unsere „Schule d. Br.“ in die Welt sandten, gaben wir im Vorwort Kunde von unserem Vorhaben, die deutschen Lokalbiere aus eigener Anschauung kennen zu lernen und Alles, was wir darüber erfahren konnten, in diesen Kreisen mitzutheilen. Heute sehen wir ein, daß wir dieses Unternehmen etwas zu leicht genommen haben. Freilich rechneten wir dabei ganz wesentlich auf eine freundliche Mitwirkung der Brauer. Die ist uns auch namentlich Anfangs vielfach versprochen worden; aber nur von sehr vereinzelt Seiten her haben wir bis heute die gewünschte und wiederholt erbetene Theilnahme erhalten.

Versuchen wir heute einmal einen andern Weg als Reizmittel zur Mitwirkung. Sezen wir eine Biographie her. Der Dortmundener Adam bietet interessante Einzelheiten genug.

## DTABf-Text

- `<teiHeader> ...<title type="main">Dortmunder Adambier.</title><title type="sub">(aus des Herausgebers Arbeit über deutsche Lokalbiere.)</title>`
- `<editorialDecl> <p>Die Transkription erfolgte nach den unter<ref target="http://www.deutschestextarchiv.de/doku/basisformat/mdTranskription.html">http://www.deutschestextarchiv.de/doku/basisformat/mdTranskription.html</ref> formulierten Richtlinien.</p> <p>Verfahren der Texterfassung: manuell (einfach erfasst).</p>`

## Adambier

### Dortmunder Adambier.

(Aus des Herausgebers Arbeit über deutsche Lokalbiere.)

Als wir im Dezember 1862 unsere „Schule d. Br.“ in die Welt sandten, gaben wir im Vorwort Kunde von unserem Vorhaben, die deutschen Lokalbiere aus eigener Anschauung kennen zu lernen und Alles, was wir darüber erfahren konnten, in diesen Kreisen mitzutheilen. Heute sehen wir ein, daß wir dieses Unternehmen etwas zu leicht genommen haben. Freilich rechneten wir dabei ganz wesentlich auf eine freundliche Mitwirkung der Brauer. Die ist uns auch namentlich Anfangs vielfach versprochen worden; aber nur von sehr vereinzelt Seiten her haben wir bis heute die gewünschte und wiederholt erbetene Theilnahme erhalten.

Versuchen wir heute einmal einen andern Weg als Reizmittel zur Mitwirkung. Sezen wir eine Biographie her. Der  
Dortmunder Adam  
bietet interessante Einzelheiten genug.

## DTABf-Text

- `<body>` ...
- `<p>`Als wir im Dezember 1862 unsere "Schule d. Br." in die Welt sandten,`<lb/>` gaben wir im Vorwort Kunde von unserem Vorhaben, die deutschen Lokalbiere`<lb/>`aus eigener Anschauung kennen zu lernen und Alles, was wir darüber erfahren`<lb/>`konnten, in diesen Kreisen mitzutheilen. Heute sehen wir ein, daß wir dieses`<lb/>` Unternehmen etwas zu leicht genommen haben. Freilich rechneten wir dabei`<lb/>`

# Beispiel 2: Daniel Sanders Briefe

Volltext	Absender	Empfänger	Absendeort	Empfangsort	Datum	Archiv
	Sanders, Daniel	Jüdische Gemeinde zu Strelitz	Berlin	Altstrelitz	1842-05-25	Neue Synagoge Berlin
	Sanders, Daniel	Jüdische Gemeinde zu Strelitz	Altstrelitz	Altstrelitz	1842-08-05	Neue Synagoge Berlin
	Sanders, Daniel	Oppenheim, Heinrich Bernhard	Altstrelitz	Heidelberg	1844-07-00	n/a
	Sanders, Daniel	Jüdische Gemeinde zu Strelitz	Altstrelitz	Altstrelitz	1844-08-16	Neue Synagoge Berlin
<b>D T A</b>	Sanders, Daniel	Schwegler, Albert	Altstrelitz	Tübingen	1845-01-19	UB Tübingen
	Sanders, Daniel	Barth, Johann Ambrosius	Altstrelitz	Leipzig	1846-04-26	Staatsbibliothek zu Berli
	Sanders, Daniel	Barth, Johann Ambrosius	Altstrelitz	Leipzig	1846-04-30	Staatsbibliothek zu Berli
	Sanders, Daniel	Hofmann, Friedrich	Altstrelitz	Hildburghausen	1846-06-09	Sanders-Archiv
	Sanders, Daniel	Barth, Johann Ambrosius	Altstrelitz	Leipzig	1846-06-19	Staatsbibliothek zu Berli
<b>D T A</b>	Sanders, Daniel	Hofmann, Friedrich	Altstrelitz	Hildburghausen	1846-08-22	Landesbibliothek Cobur
	Sanders, Daniel	Jüdische Gemeinde zu Strelitz	Altstrelitz	Altstrelitz	1847-01-10	Neue Synagoge Berlin
	Sanders, Daniel	Jüdische Gemeinde zu Strelitz	Altstrelitz	Altstrelitz	1847-04-06	Neue Synagoge Berlin
	Sanders, Daniel	Jüdische Gemeinde zu Strelitz	Altstrelitz	Altstrelitz	1847-08-06	Neue Synagoge Berlin
<b>D T A</b>	Sanders, Daniel	Hofmann, Friedrich	Altstrelitz	n/a	1847-08-28	Landesbibliothek Cobur
	Grimm, Jacob	Sanders, Daniel	Berlin	Altstrelitz	1847-08-28	n/a

- <http://www.deutschestextarchiv.de/sanders-briefe/>

```
<correspDesc ref="www.dta.de/volger_sanders_1881">  
  <correspAction type="sent">  
    <persName ref="http://d-nb.info/gnd/117488739">Volger,  
    Georg Heinrich Otto</persName>  
    <placeName ref="http://www.geonames.org/2925533">Frankf  
    urt am Main</placeName>  
    <date when="1881-04-25"/>  
  </correspAction>  
  
  <correspAction type="received">  
    <persName ref="http://d-nb.info/gnd/119242044">Sanders,  
    Daniel</persName>  
  </correspAction>  
  ...  
</correspDesc>
```

**correspSearch**    Start    Suche    Daten    Über correspSearch    API    Mitmachen    Deutsch ▼

Briefpartner:     Rolle:  +

Startdatum:     Enddatum:     Ort:     Funktion des Ortes:  ▼    Verfügbarkeit:  ▼

↻ Weitere Optionen

1 von 13 Treffern in 1 Publikation

📄 Suchergebnis herunterladen (beta)

10. April  
1809

Von Friedrich Creuzer an Jacob Grimm

In: Steig, Reinhold: Zeugnisse zur Pflege der deutschen Literatur in den Heidelberger Jahrbüchern. In: Neue Heidelberger Jahrbücher 11,2 (1902), S. 180–284. [Digitalisat verfügbar unter <http://digital.slub-dresden.de/id32725617Z>].

↻ Brief Nr. 19

2. Mai  
1809

Von Friedrich Creuzer an Jacob Grimm

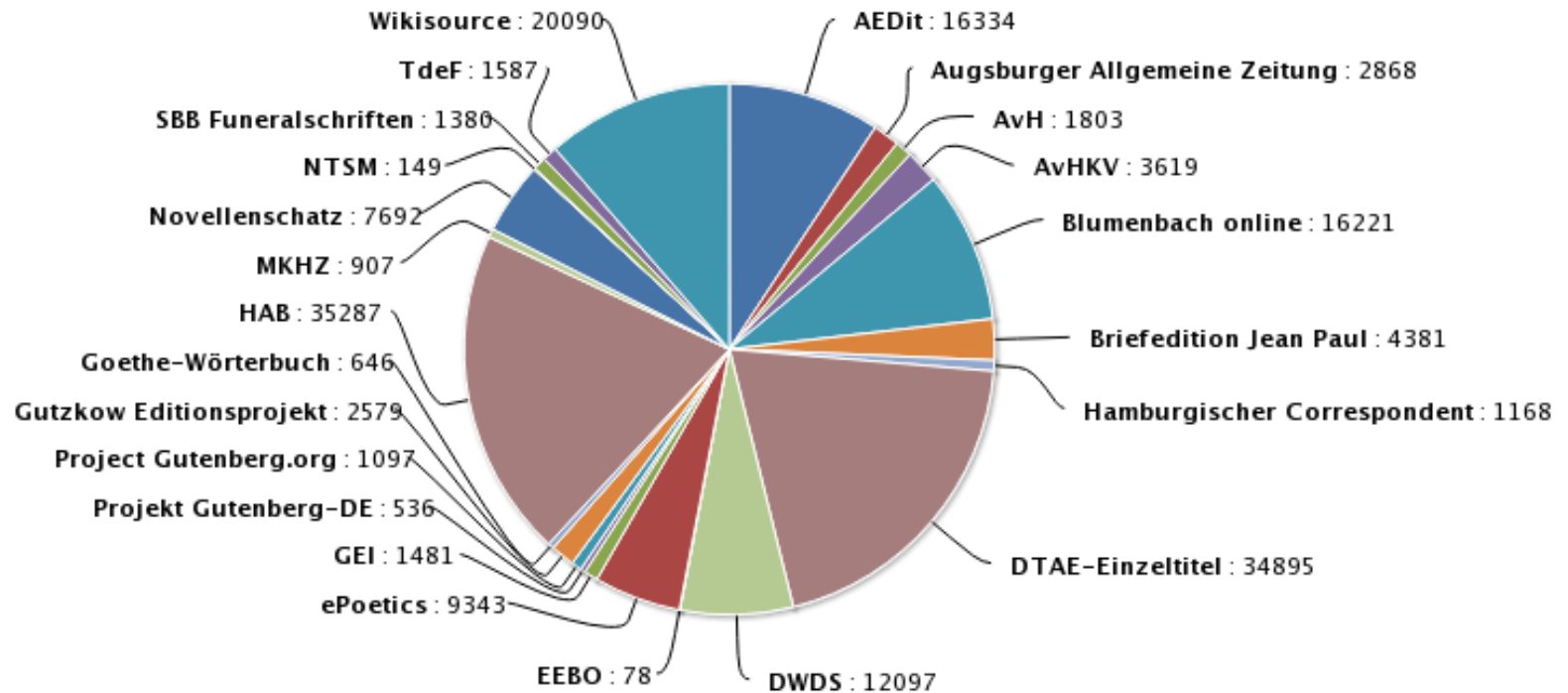
In: Steig, Reinhold: Zeugnisse zur Pflege der deutschen Literatur in den Heidelberger Jahrbüchern. In: Neue Heidelberger Jahrbücher 11,2 (1902), S. 180–284. [Digitalisat verfügbar unter <http://digital.slub-dresden.de/id32725617Z>].

↻ Brief Nr. 21

1. AEDit Frühe Neuzeit (DFG)
2. Augsburger Allgemeine Ztg (Eigenleistung/CLARIN)
3. AvH auf Reisen (Eigenleistung/CLARIN)
4. AvH-Kosmos (Eigenleistung/CLARIN)
5. Blumenbach (Akademienprojekt)
6. Briefedition Jean Paul (Akademienprojekt)
7. Digitexte (Univ. Gießen)
8. Dingler (DFG)
9. ePoetics (BMBF)
10. Erbkam (Akademienprojekt)
11. Erfahrungsseelenkunde (Akademienprojekt)
12. GEI-Digital (DFG)
13. Grenzboten (DFG)
14. Gutenberg-DE (Kuration CLARIN)
15. Gutenberg.org (Kuration CLARIN)
16. Gutzkow Editionsprojekt (University of Exeter)
17. Goethe-Wörterbuch (Akademienprojekt)
18. Hamburgischer Correspondent (NRW-Landesförderung)
19. HAB Oberhofprediger (DFG)

20. Joachim v. Sandrart (DFG)
21. MKHZ (IDS-Mannheim)
22. Novellenschatz (Volkswagen-Stiftung)
23. Nürnberger Texte des späten Mittelalters (DFG)
24. Peter Schlemihl (Dennerlein, Stipendium BAW)
25. Ratgeber für ein gutes Leben (ECHO, MPI)
26. Sanders Briefwechsel (Einzelwissenschaftler)
27. SBB Funeralschriften (DFG)
26. Texte der ersten Frauenbewegung (Univ. Gießen, Einzelwissenschaftlerin)
27. TextGrid/Zeno-Subkorpus (Kuration CLARIN-D)
28. Wikisource-Subkorpus (Kuration CLARIN-D)
29. Wolfenbütteler Digitale Bibliothek (WDB) (CLARIN-D)
30. Div. Einzeltexte (ca. 200 Einzeltexte: George Adams, 1785 - )

nach Seiten gruppiert





# Nachnutzung des DTA

Mehrwert 1: einheitlich recherchieren

Mehrwert 2: Downloadpakete

1. Transliteration („lange s“ (ʃ) in ein „rundes s“ )
  2. Zurückführung auf eine phonetische Repräsentation (bei unbekanntem Wörtern), z. B. die phonetische Form für die Schreibweisen „Theyl“, „Thayl“, „Teyl“, -> [tail] -> *Teil*.
  3. Graphematische Ersetzungsregeln: z. B. „gläuben“ als Variante von „glauben“. Für jede historische Wortform das „ähnlichste“ moderne Wort automatisch ermittelt.
  4. Ausnahmelexikon  
→ Lemmatisierung der nhdt. Form mit TAGH
- 
- Bryan Jurish (2010), "[More than words: using token context to improve canonicalization of historical German.](#)" *Journal for Language Technology and Computational Linguistics*, 25(1):23-40

- CAB kann als HTTP-basierter Webservice in folgender Weise genutzt werden:
  - projektintern: Indizierung der (DTA-)Texte
  - in der Verarbeitungskette von CLARIN-D ([WebLicht](#)) in Zusammenhang mit anderen CLARIN-Web-Services
  - direkt zur Normalisierung:  
<http://www.deutschestextarchiv.de/cab/>

"Kleid" / D\*/DTA Query Lizard

kaskade.dwds.de/dstar/dta/dstar.perl?fmt=expand-html&q=Kleid&x=Token

Home Query Lizard JSON Text Help

### Base Query

Query Term(s): Kleid

Expander: Token

### Expansions:

<input checked="" type="checkbox"/> <a href="#">Cleider</a>	<input checked="" type="checkbox"/> <a href="#">Kleider=</a>	<input checked="" type="checkbox"/> <a href="#">Kleydern</a>	<input checked="" type="checkbox"/> <a href="#">kleid</a>
<input checked="" type="checkbox"/> <a href="#">Cleidis</a>	<input checked="" type="checkbox"/> <a href="#">Kleidere</a>	<input checked="" type="checkbox"/> <a href="#">Kleydes</a>	<input checked="" type="checkbox"/> <a href="#">kleider</a>
<input checked="" type="checkbox"/> <a href="#">Cleyder</a>	<input checked="" type="checkbox"/> <a href="#">Kleideren</a>	<input checked="" type="checkbox"/> <a href="#">Kleyds</a>	<input checked="" type="checkbox"/> <a href="#">kleideren</a>
<input checked="" type="checkbox"/> <a href="#">KLEID</a>	<input checked="" type="checkbox"/> <a href="#">Kleidergen</a>	<input checked="" type="checkbox"/> <a href="#">Kleydt</a>	<input checked="" type="checkbox"/> <a href="#">kleidern</a>
<input checked="" type="checkbox"/> <a href="#">Klaid</a>	<input checked="" type="checkbox"/> <a href="#">Kleidern</a>	<input checked="" type="checkbox"/> <a href="#">Kleyt</a>	<input checked="" type="checkbox"/> <a href="#">kleides</a>
<input checked="" type="checkbox"/> <a href="#">Klaider</a>	<input checked="" type="checkbox"/> <a href="#">Kleidern/</a>	<input checked="" type="checkbox"/> <a href="#">chlaider</a>	<input checked="" type="checkbox"/> <a href="#">kleids</a>
<input checked="" type="checkbox"/> <a href="#">Klaidern</a>	<input checked="" type="checkbox"/> <a href="#">Kleides</a>	<input checked="" type="checkbox"/> <a href="#">claider</a>	<input checked="" type="checkbox"/> <a href="#">kleit</a>
<input checked="" type="checkbox"/> <a href="#">Klayd</a>	<input checked="" type="checkbox"/> <a href="#">Kleidi</a>	<input checked="" type="checkbox"/> <a href="#">clayder</a>	<input checked="" type="checkbox"/> <a href="#">kleitt</a>
<input checked="" type="checkbox"/> <a href="#">Klayder</a>	<input checked="" type="checkbox"/> <a href="#">Kleids</a>	<input checked="" type="checkbox"/> <a href="#">cleider</a>	<input checked="" type="checkbox"/> <a href="#">kleyd</a>
<input checked="" type="checkbox"/> <a href="#">Klaydern</a>	<input checked="" type="checkbox"/> <a href="#">Kleidt</a>	<input checked="" type="checkbox"/> <a href="#">cleyder</a>	<input checked="" type="checkbox"/> <a href="#">kleyder</a>
<input checked="" type="checkbox"/> <a href="#">Klaydt</a>	<input checked="" type="checkbox"/> <a href="#">Kleidörn</a>	<input checked="" type="checkbox"/> <a href="#">cleydern</a>	<input checked="" type="checkbox"/> <a href="#">kleyderen</a>
<input checked="" type="checkbox"/> <a href="#">Kleder</a>	<input checked="" type="checkbox"/> <a href="#">Kleit</a>	<input checked="" type="checkbox"/> <a href="#">klaid</a>	<input checked="" type="checkbox"/> <a href="#">kleydern</a>
<input checked="" type="checkbox"/> <a href="#">Kleeder</a>	<input checked="" type="checkbox"/> <a href="#">Kleiter</a>	<input checked="" type="checkbox"/> <a href="#">klaider</a>	<input checked="" type="checkbox"/> <a href="#">kleydes</a>
<input checked="" type="checkbox"/> <a href="#">Kleid</a>	<input checked="" type="checkbox"/> <a href="#">Kleyd</a>	<input checked="" type="checkbox"/> <a href="#">klaidern</a>	<input checked="" type="checkbox"/> <a href="#">kleydis</a>
<input checked="" type="checkbox"/> <a href="#">Kleid'</a>	<input checked="" type="checkbox"/> <a href="#">Kleyde</a>	<input checked="" type="checkbox"/> <a href="#">klayd</a>	<input checked="" type="checkbox"/> <a href="#">kleyds</a>
<input checked="" type="checkbox"/> <a href="#">Kleide</a>	<input checked="" type="checkbox"/> <a href="#">Kleyder</a>	<input checked="" type="checkbox"/> <a href="#">klayder</a>	<input checked="" type="checkbox"/> <a href="#">kleyt</a>
<input checked="" type="checkbox"/> <a href="#">Kleider</a>	<input checked="" type="checkbox"/> <a href="#">Kleyderen</a>	<input checked="" type="checkbox"/> <a href="#">klaydt</a>	<input checked="" type="checkbox"/> <a href="#">verkleid</a>

<http://kaskade.dwds.de/dstar/dta/lizard.perl?q=Kleid>

Downloads gesamt: 2044 (01/2018)

287 dta\_kernkorpus

180 dta\_komplett

139 dta\_komplett\_1600-1699

134 dta\_kernkorpus\_gebrauchsliteratur

132 dta\_komplett\_gebrauchsliteratur

127 dta\_kernkorpus\_1800-1899

125 dta\_komplett\_1800-1899

125 dta\_kernkorpus\_belletristik

124 dta\_komplett\_wissenschaft

121 dta\_kernkorpus\_wissenschaft

...

<http://www.deutschestextarchiv.de/download/>



# Weitere Perspektiven des DTA für die Digitalen Geisteswissenschaften

- Linguistik:
  - DTA als Grundlage für ling. Veröffentlichungen
  - Ergebnisse der linguistischen Analyse, CAB als Webservice
- Literaturwissenschaft:
  - Digitalisierte Textausgaben, Werke im Kontext
- Editionswissenschaft:
  - Saubere Textvorlagen für die Edition
- Geschichtswissenschaft:
  - Begriffsgeschichte (mit DiaCollo, einem Werkzeug zur Ermittlung diachroner Kollokationen)

- dynamisch wachsend durch Kooperationsprojekte (4268 Werke, 256 Mio. Tokens, davon 120 Mio. Kernkorpus)
- Qualitätssicherung: 1307 angemeldete Nutzer in DTAQ, der kollaborativen Qualitätssicherungsplattform des DTA
- ~60 wissenschaftliche Publikationen über das DTA (Linguistik, Zeitgeschichte, Kultur- und Literaturwissenschaften)

<http://www.deutschestextarchiv.de/clarin-kooperationen>



- DTA: 2007-2016 (DFG-gefördert)
- Seit 2017: Teil von CLARIN  
(„Nutzungsphase“ (9/2016-8/2020))
  - Zentrum Sprache (BBAW) ist Koordinator des Kompetenzbereichs „historische Daten“
- DTA-Plattform kann technisch betrieben werden:
  - Texte nachhaltig in der CLARIN-Infrastruktur (CLARIN-Servicezentrum, Persistenz und Downloadmöglichkeiten)
  - Schnittstellen (Suche, DTAQ, CAB)

- DTA als aktives Archiv hochqualitativer historischer Texte (projektübergreifend!)
- Interoperabilität durch Best-Practice XML/TEI- basiertes DTA-Basisformat
- (möglichst) breite Nachnutzung
- Nachhaltigkeit durch CLARIN

## Kontakt

{clarin|dta}@bbaw.de

## Links

- DTA-Web: [www.deutschestextarchiv.de](http://www.deutschestextarchiv.de) + DTAQ: [~/dtaq](#)
- DTA-Basisformat: [www.deutschestextarchiv.de/doku/basisformat](http://www.deutschestextarchiv.de/doku/basisformat)
- DiaCollo: <http://kaskade.dwds.de/dstar/dta/diacollo/>
- BBAW als CLARIN-Zentrum: <http://clarin.bbaw.de>
- Projekt OCR-D: [www.ocr-d.de](http://www.ocr-d.de)

## Mehr

Publikationen: [www.deutschestextarchiv.de/doku/publikationen](http://www.deutschestextarchiv.de/doku/publikationen)

Blog: <http://sprache.hypotheses.org>

Twitter: [@textarchiv](#)