

# Homogene Aufbereitung Historischer Zeitungen: Richtlinien und Workflows des Deutschen Textarchivs

Susanne Haaf

Deutsches Textarchiv, BBAW – Berlin

*Matthias Boenig, Alexander Geyken, Susanne Haaf,  
Bryan Jurish, Christian Thomas,  
Frank Wiegand, Kay-Michael Würzner*





# Inhalt

---

1. Über das Projekt „Deutsches Textarchiv“
2. Problem Statement
3. Projekte und Workflows
4. Resultierende DTABf-Anpassungen für Zeitungen
5. Nutzung der Daten

# Das Deutsche Textarchiv

Anmelden (DTAQ) DWDS dexDB CLARIN-D

**DTA**   Texte ▼ Projekt ▼ Dokumentation ▼ Impressum

in den Titeldaten  im Korpus  in der Dokumentation [Hilfe](#)

Beispielanfragen: @versprechend && !viel \$!=Erkenntnis \$euche #has[textClassDTA,/Medizin/] #less\_by\_date[1790,1820]

## 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 Akademiemitgliedern 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 schreibweisentolerant über den gesamten jeweils verfügbaren Bestand durchsuchbar. Sämtliche Texte stehen zum Download zur Verfügung. [mehr ...](#)

---

### NEUIGKEITEN AUS DEM PROJEKT

*2. DTA- & CLARIN-D-Konferenz und Workshop: Textkorpora in Infrastrukturen für die Geistes- und Sozialwissenschaften, 17./18.11.2014, BBAW Berlin*

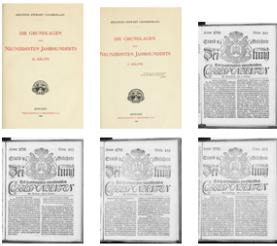
Die Projekte DTA und CLARIN-D laden ein zur kommenden zweiten gemeinsamen Konferenz zum Thema "Textkorpora in Infrastrukturen für die Geistes- und Sozialwissenschaften". Begleitet wird die Konferenz von einem CLARIN-D-Workshop zum Arbeitspaket 5 „Sprachressourcen und Dienste“. Die Veranstaltungen finden am 17. und 18. November 2014 im Einsteinsaal der Berlin-Brandenburgischen Akademie der Wissenschaften, Jägerstr. 22/23, in Berlin statt. [mehr ...](#)


---

### Das DTA in Zahlen

- 1553 Werke
- 461370 digitalisierte Seiten
- 110390283 fortlaufende Wortformen
- 771056527 Zeichen (Unicode)
- 787 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

### Teilprojekte

- DTA-Erweiterungen (DTAE)
- CLARIN-D: Kurationsprojekt 1 der F-AG 1
- Metadatenformular zur Aufnahme einzelner

DFG

2007-2014



<http://www.deutschestextarchiv.de>



# Das Deutsche Textarchiv

**D T A**

- Grundlage für ein Referenzkorpus zur Entwicklung der neuhochdeutschen Sprache (ca. 1600–1900)
- rd. 1500 Texte unterschiedlicher Disziplinen und Textsorten in Text und Bild, hohe Textgenauigkeit (Double Keying)
- Automatisierte linguistische Analyse (Lemmatisierung, Tokenisierung, POS-Tagging, Orthographische Normierung)

**D T A Bf**

- Richtlinien für die quellennahe Transkription
- Strukturelle XML-Annotation nach TEI/P5

**D T A Q**

- Richtlinien und Vorverarbeitung der Bilder
- Web-basierte Qualitätssicherung

**D T A E**

- Integration historischer Textdaten aus anderen Projektkontexten (z.B. Historische Zeitungen)
- Kuration und Zusammenführung verstreuter Textressourcen
- Grundlage: DTA-Infrastruktur zur Nachnutzung
- Dabei Anpassung des DTA-Workflows entsprechend den speziellen Projektanforderungen

# Text-Bild-Ansicht im DTA

Aviso. Relation oder Zeitung. Wolfenbüttel, 1609.

Bild: 0007 << vorherige Seite

nächste Seite >>

Suche im Werk

Avifa  
Relation oder Zeitung.  
Was sich begeben vnd  
zugetragen hat/ in Deutsch: vnd Welsch-  
land/ Spanien/ Niederlandt/ Engellandt/ Franck-  
reich/ Vngern/ Osterreich/ Schweden/ Polen/  
vnd in allen Provintzen/ in Ost: vnd  
West Jndien etc.So alhie den 15. Januarij angelant.Gedruckt im Jahr/ 1609.

Informationen zum Werk

- Titeldaten
- Verfügbarkeit Text (TEI-XML-, HTML-, TCF-, E-Book-Fassung): Distributed under the Creative Commons Attribution-NonCommercial 3.0 Unported (German) License.
- Nutzungsbedingungen

Download dieses Werks

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

Metadaten zum Werk

TEI-Header • CMDI • Dublin Core

Ansichten dieser Seite

- gerendertes HTML (normierte Zeichen)
- gerendertes HTML (Originalzeichen)
- zugrunde liegendes XML/TEI
- Reintext
- normalisierte Orthographie
- Faksimile in hoher Auflösung

Feedback

Sie haben einen Fehler gefunden? Dann können Sie diesen über unsere Qualitätssicherungsplattform DTAQ melden.

Kommentar zur DTA-Ausgabe

Thomas Gloning: Bereitstellung der Texttranskription (2013-03-18T14:17:31Z)  
Bitte beachten Sie, dass die aktuelle Transkription (und Textauszeichnung)

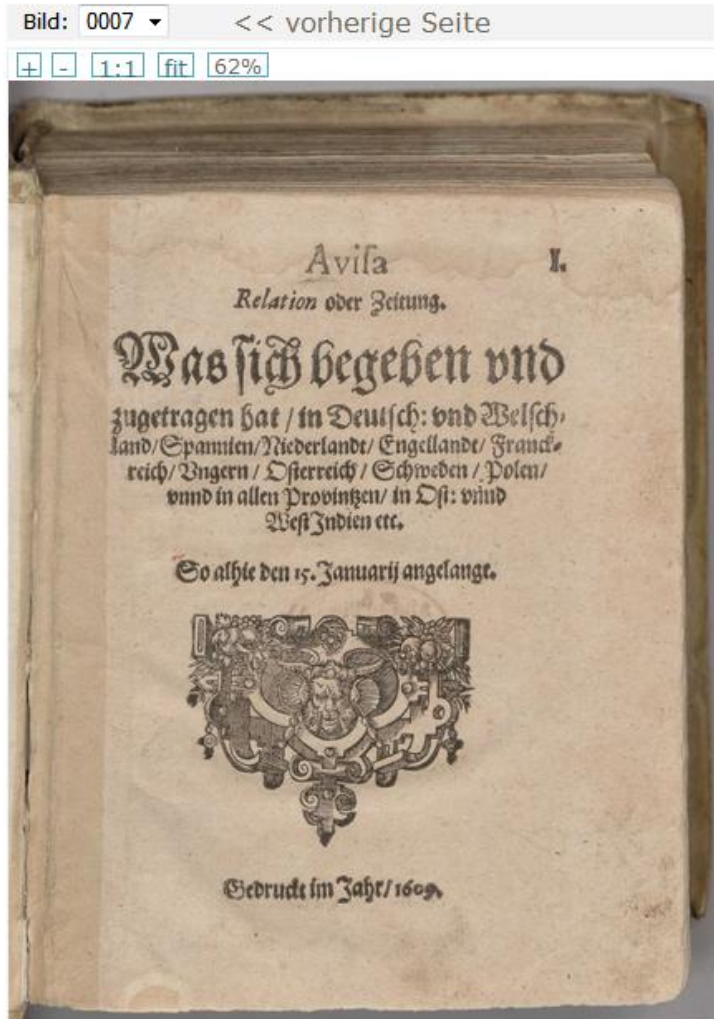
Bild: 0007 1:1 fit 71%

Avifa  
Relation oder Zeitung.  
Was sich begeben vnd  
zugetragen hat / in Deutsch: vnd Welsch-  
land/ Spanien/ Niederlande/ Engellande/ Franck-  
reich/ Vngern / Osterreich / Schweden / Polen/  
vnd in allen Provinzen/ in Ost: vnd  
West Jndien etc.So alhie den 15. Januarij angelange.Gedruckt im Jahr/ 1609.

[http://www.deutschestextarchiv.de/anonym\\_aviso\\_1609/7](http://www.deutschestextarchiv.de/anonym_aviso_1609/7)

# Textstrukturierung nach DTA-Basisformat

Aviso. Relation oder Zeitung. Wolfenbüttel, 1609.



```

Bild: 0007 << vorherige Seite
nächste Seite >>
+ - 1:1 fit 62%
<TEI>
  <text>
    <pb facs="#f0007"/>
    <body>
      <div n="1">
        <head> <hi rendition="#c"><hi
rendition="#aq">Avisa</hi><note place="right">I.</note>
<lb/><hi rendition="#i"><hi
rendition="#aq">Relation</hi></hi> oder Zeitung.<lb/>
Was sich begeben vnd<lb/>
zugetragen hat/ in Deutsch: vnd Welsch-<lb/>
land/ Spanien/ Niederlandt/ Engellandt/ Franck-<lb/>
reich/ Vngern/ Osterreich/ Schweden/ Polen/<lb/>
vnd in allen Provintzen/ in Ost: vnd<lb/>
West Jndien etc.</hi> </head><lb/>
        <head> <hi rendition="#c">So alhie den <date>15.
Januarij</date> angelangt.</hi> </head><lb/>
        <milestone rendition="#hr" unit="section"/><lb/>
        <docDate> <hi rendition="#c">Gedruckt im Jahr/
1609.</hi> </docDate><lb/>
      </div>
    </body>
  </text>
</TEI>

```

[http://www.deutschestextarchiv.de/anonym\\_avis\\_1609/7](http://www.deutschestextarchiv.de/anonym_avis_1609/7)

- TEI-P5 Format für die semantische und typographische Auszeichnung historischer gedruckter Texte
- Elemente für die Textstrukturierung → Teilmenge der TEI; festgelegtes Vokabular von Attributen und Werten
- Elemente für die Metadatenstrukturierung
- Ziel:
  - Gute Abdeckung der behandelbaren Phänomene
  - Keine Ambiguitäten für die Textauszeichnung
  - Vollständige TEI-Konformität
- Kontinuierliche Anpassung an neue Phänomene
- Spezielle Richtlinien für besondere Textsorten (z.B. Funeralschriften, Zeitungen)
- DTA ist Mitglied des *TEI Simple* Advisory Boards

# Qualitätssicherung im DTA



zuletzt gelesen · Hilfe · Zufallsseite

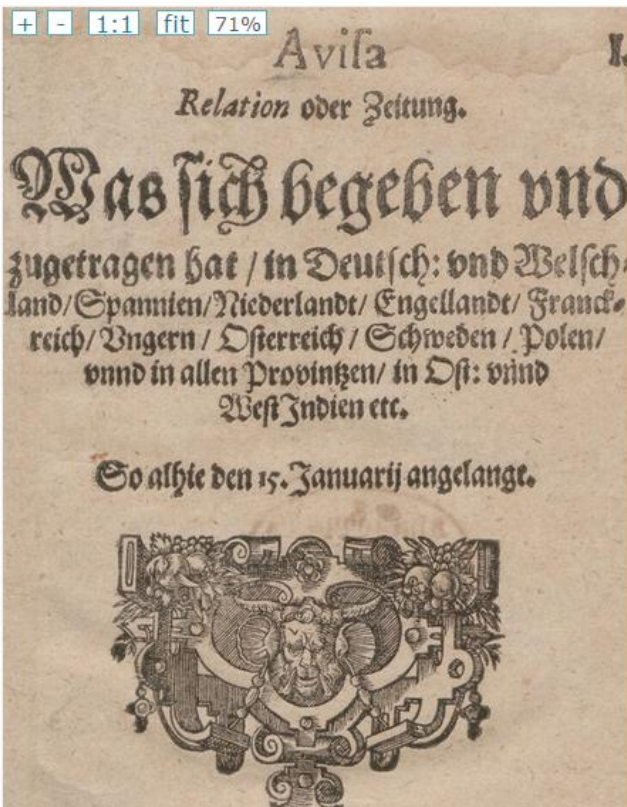
SusanneHaaf | [Admin](#) | [Profil](#) | [ausloggen](#)

## anonym\_avis\_1609 (DTAE)

offene Tickets: 2 (0 ganzes Buch)  
Stand: 2014-08-25 13:37:30

Text Text/Bild   
0 - 1 - 412 0 - 0 - 413

Bild: 0007 << vorherige Seite



nächste Seite >>

Avisa  
Relation oder Zeitung.  
Was sich begeben vnd  
zugetragen hat / in Deutsch: vnd Welschland / Spanien / Niederlande / Engelland / Franckreich / Vngern / Osterreich / Schweden / Polen / vnd in allen Provinzen / in Ost: vnd West Indien etc.  
So alhie den 15. Januarij angelant.  
Gedruckt im Jahr / 1609.

Buchdaten

DTA-Informationen  
Metadaten  
Ansichten (Webversion)  
nächstes Ticket

Korrekturstatus ?

Text von mir kontrolliert ?

Text/Bild von mir kontrolliert ?

Darstellung von mir kontrolliert

TEI-XML von mir kontrolliert

Tickets für diese Seite ?

neu: Ticket

Suche

?  
  +

Anmerkung zu dieser Seite anlegen

<http://www.deutschestextarchiv.de/dtaq/book/view/30366?p=7>

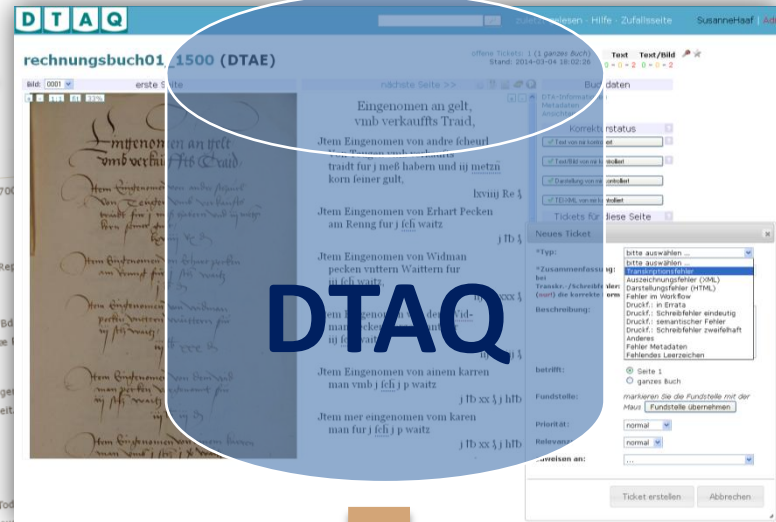




# Qualitätssicherung im DTA

HISTORIOGRAPHIE  
 Arnold, Gottfried: Unpartheyische Kirchen- und Ketzer-Historie. Bd. 2 (T. 3/4). Frankfurt (Main), 1704.  
 Burckhardt, Jacob: Die Cultur der Renaissance in Italien. Ein Versuch. Basel, 1860.  
 Chladni, Johann Martin: Allgemeine Geschichtswissenschaft. Leipzig, 1752.  
 Curtius, Ernst: Alterthum und Gegenwart. Gesammelte Reden und Vorträge. Bd. 1. Berlin, 1875.  
 Dahlmann, Friedrich Christoph: Geschichte der französischen Revolution bis auf die Stiftung der Republik. Hamburg, (1833).  
 Droysden, Franz Gustav: Geschichte Alexanders des Großen. Hamburg, (1833).  
 Droysden, Franz Gustav: Geschichte der russischen Revolution bis auf die Einführung der Verfassung. Hamburg, (1833).  
 [Gednee, Christian: ...]  
 Gundling, Nicolaus Hieronymus: Discovris über weyl. Herrn D. ...  
 a., 1733.  
 Harez, Peter: ...  
 Heeren, Arnold: ...  
 [Herder, Johann Gottfried: ...]  
 [Lettus, Henricus]: Der Liefländischen Chronik Erster Theil. Halle, 1747.  
 Lettus, Henricus: Der Liefländischen Chronik Erster Theil. Halle, 1747.  
 Marx, Karl: Der achtzehnte Brumaire des Louis Bonaparte. 2. Aufl. Hamburg, 1869.  
 Mommsen, Theodor: Römische Geschichte. Bd. 1: Bis zur Schlacht bei Pydna bis auf Sulla's Tod.  
 Mommsen, Theodor: Römische Geschichte. Bd. 2: Von der Schlacht bei Pydna bis auf Sulla's Tod.  
 Mommsen, Theodor: Römische Geschichte. Bd. 3: Von Sulla's Tode bis zur Schlacht von Thapsus.  
 Mommsen, Theodor: Römische Geschichte. Osnabrück, 1768.  
 Moser, Justus: Osnabrückische Geschichte. Osnabrück, 1768.  
 Müller, Karl Otfried: Die Dorier. Vier Bücher. Bd. 1. Breslau, 1824.  
 Müller, Karl Otfried: Die Dorier. Vier Bücher. Bd. 2. Breslau, 1824.  
 Müller, Karl Otfried: Die Dorier. Vier Bücher. T. 1. Berlin, 1811.  
 Niebuhr, Barthold Georg: Römische Geschichte. T. 2. Berlin, 1812.  
 Niebuhr, Barthold Georg: Römische Geschichte. T. 2. Berlin, 1812.  
 Ranke, Leopold von: Deutsche Geschichte im Zeitalter der Reformation. Bd. 1. Berlin, 1835.  
 Ranke, Leopold von: Deutsche Geschichte im Zeitalter der Reformation. Bd. 2. Berlin, 1835.  
 Ranke, Leopold von: Deutsche Geschichte im Zeitalter der Reformation. Bd. 2. Berlin, 1835.

Texte des DTA-Kernkorpus



DTAQ

Hamburgischer Correspondent<sup>(?)</sup>  
 Werke anzeigen ...

- Staats- und Gelehrte Zeitung des Hamburgischen unpartheyischen Correspondenten, Nr. 27, 17. Februar 1801.
- Staats- und Gelehrte Zeitung des Hamburgischen unpartheyischen Correspondenten, Nr. 23, 10. Februar 1801.
- Staats- und Gelehrte Zeitung des Hamburgischen unpartheyischen Correspondenten, Nr. 25, 13. Februar 1801.
- Staats- und Gelehrte Zeitung des Hamburgischen unpartheyischen Correspondenten, Nr. 21, 6. Februar 1801.
- Staats- und Gelehrte Zeitung des Hamburgischen unpartheyischen Correspondenten, Nr. 28, 18. Februar 1801.
- Staats- und Gelehrte Zeitung des Hamburgischen unpartheyischen Correspondenten, Nr. 26, 14. Februar 1801.

Texte aus dem Modul DTA-Erweiterungen

Double Keying

Zoning; Richtlinien



DTA



# DTA Zahlen

	Moritz, Karl Philipp	Über die bildende Nachahmung des Schönen	1788	Wissenschaft	Philosophie
	Bräker, Ulrich	Lebensgeschichte und natürliche Ebentheur des Armen Mannes im Tockenburg. Herausgegeben von H. H. Füßli	1789	Belletristik	Autobiographie
	Gerstner, Franz Joseph Ritter von	Einleitung in die statische Baukunst	1789	Wissenschaft	Architektur
	Gleditsch, Johann Gottlieb	Vermischte botanische und oekonomische Abhandlungen	1789	Wissenschaft	Biologie
	Gleditsch, Johann Gottlieb	Vermischte botanische und oekonomische Abhandlungen	1789	Wissenschaft	Biologie
	Gleditsch, Johann Gottlieb	Vermischte botanische und oekonomische Abhandlungen	1789	Wissenschaft	Biologie
	Gotter, Friedrich Wilhelm	Die Erlaubnisse der Kunst	1789	Belletristik	Drama
	Schiller, Friedrich	Der Geist der Sprache	1789	Belletristik	Roman
	Schiller, Friedrich	Was heißt Kritik in der Poesie?	1789	Wissenschaft	Historiographie
	Goethe, Johann Wolfgang von	Faust	1789	Belletristik	Drama
	Goethe, Johann Wolfgang von	Torquato	1789	Belletristik	Drama
	Hugo, Gustav	Lehrbuch der Jurisprudenz	1789	Wissenschaft	Jura
	Kotzebue, August von	Mensch und Natur	1789	Belletristik	Drama
	Moritz, Karl Philipp	Anton	1789	Belletristik	Roman
	Forster, Georg	Ansicht der Natur	1789	Wissenschaft	Geographie
	Forster, Georg	Ansicht der Natur	1790	Wissenschaft	Geographie
	Gall, Franz Joseph	Philosophische Medizin	1789	Wissenschaft	Medizin
	Klinger, Friedrich Maximilian	Fausts	1789	Belletristik	Roman
	Moritz, Karl Philipp	Götter	1789	Wissenschaft	Philosophie
	Sailer, Johann Michael	Kurzgefasste Erinnerungen an junge Prediger	1791	Gebrauchsliteratur	Theologie
	Büsch, Johann Georg	Theoretisch-Praktische Darstellung der Handlung in deren mannigfaltigen Geschäften	1792	Gebrauchsliteratur	Ökonomie
	Büsch, Johann Georg	Theoretisch-Praktische Darstellung der Handlung in deren mannigfaltigen Geschäften	1792	Gebrauchsliteratur	Ökonomie
	Günther, Karl Gottlob	Europäisches Völkerrecht in Friedenszeiten nach Vernunft, Verträgen und Herkommen mit Anwendung auf die deutschen Reichsstände	1792	Wissenschaft	Jura

**DTA Webseite ([www.deutschestextarchiv.de](http://www.deutschestextarchiv.de))**

**1 552 Werke frei verfügbar (CC BY-NC)**

**461 053 Digitalisierte Seiten**

**~ 110 M Tokens**

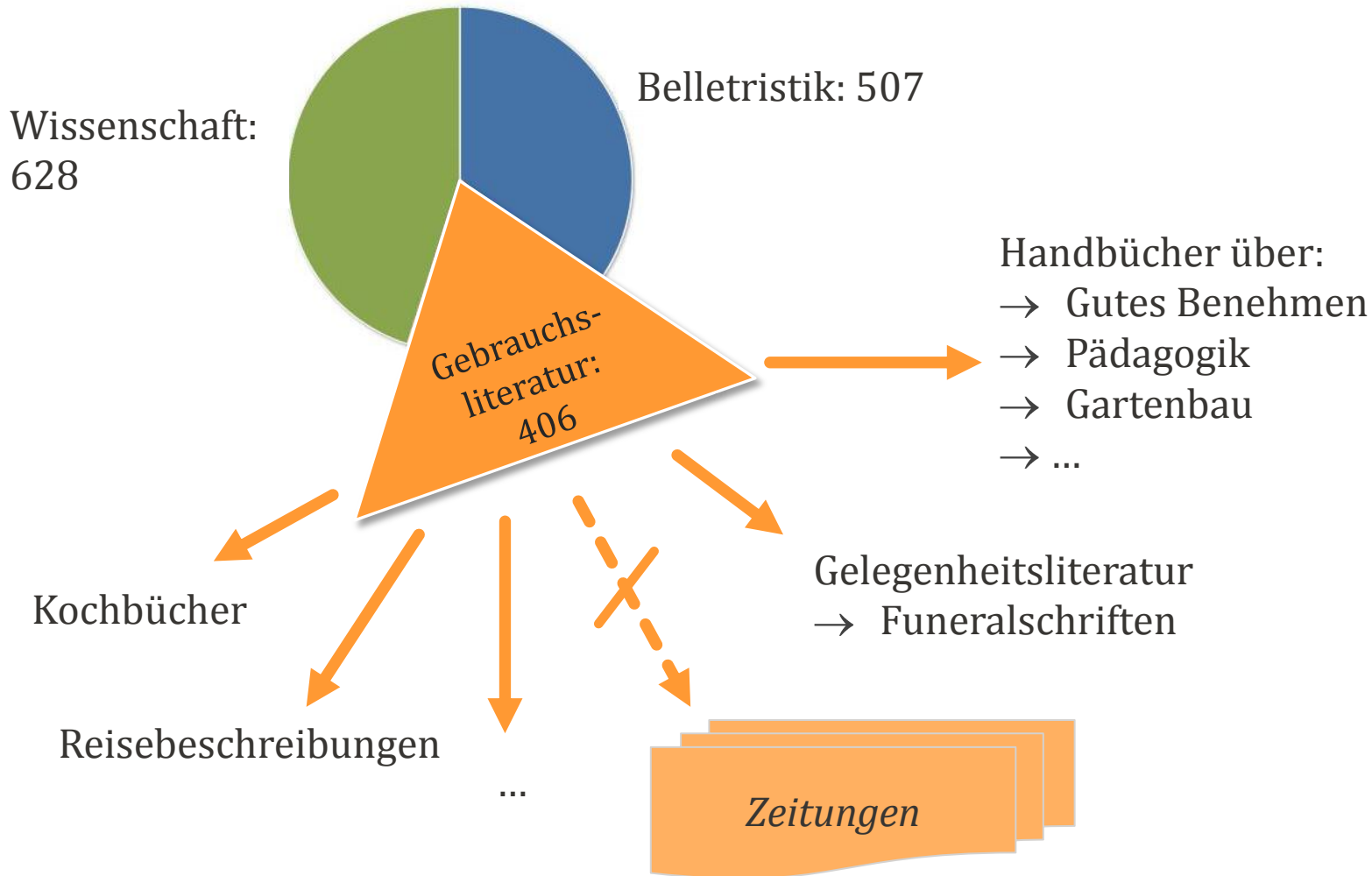
**DTA Korpora insgesamt :**

**~2 400 Werke**

**~ 200 M Tokens**

# Problem Statement

November 2014: 1552 Werke verfügbar über die DTA-Webseite



# Historische Zeitungen für das DTA: Laufende Projekte

---

1. Mannheimer Korpus historischer Zeitungen (18.–20. Jh.)
2. Neue Rheinische Zeitung (1848/49)
3. Hamburgischer Correspondent (1712–1851)
4. Die Grenzboten (1841–1922)

- Verschiedene DTAE-Workflows
- Anpassungen des DTA-Basisformats

→ Ziel: DTA-konforme Homogenisierung der Daten, um Interoperabilität innerhalb des DTA-Korpus zu erzielen

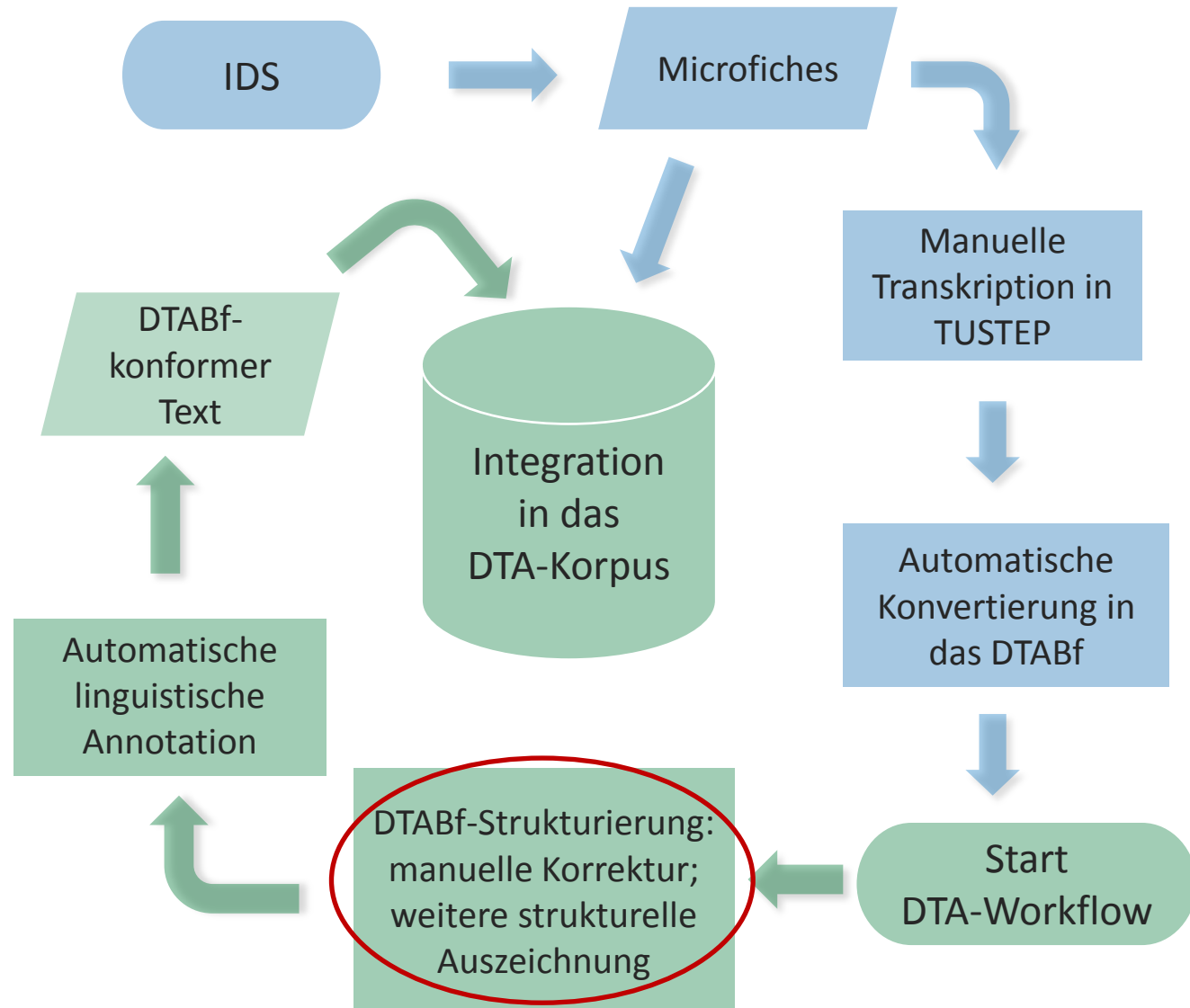




# MKHZ: Workflow

Mannheimer Korpus  
Historischer  
Zeitungen und  
Zeitschriften

- 652 vollständige Ausgaben
- aus 21 verschiedenen Zeitungen und Zeitschriften
- des 18. bis 20. Jahrhunderts



# MKHZ: Nachbearbeitung im DTA

Strukturierung  
der Titelei

```
<front>
  <div>
    <pb n="25" facs="jpeg/Weltgeschichte/Weltgeschichte0001.jpg"/>
    <p rendition="#c"><hi rendition="#g"><lb/> Der allerneuesten</hi><hi
rendition="#fr"><lb/> Europäischen</hi><hi rendition="#fr"><lb/> Welt= und
Staats=Geschichte</hi><hi rendition="#aq"><lb/> II</hi>. <hi rendition="#fr"><hi
rendition="#g">Theil</hi>.</hi></p>
    <p rendition="#c">
      <hi rendition="#aq">No. <lb/> <figure/>IV.</hi>
    </p>
    <p rendition="#c"><lb/> 2. Woche. <lb/> Freytags den 10. Jenner, <hi
rendition="#aq">MDCCLXIV.</hi> </p>
```

Strukturierung  
und  
Klassifizierung  
der Artikel

```
  </div>
</front>
<body>
  <p rendition="#c"><lb/> Schluß derer Gedancken von Neuen=Jahrs-<lb/>Zetteln.</p>
  <p><hi rendition="#in"><lb/> W</hi>Jr beschliessen unsere, einer viel weitläufftigern
Aus-<lb/>führung würdige Abhandlung, nur mit einem eintzi-<lb/>gen Exempel, woraus wir die
ohnedem tägliche Wahr-<lb/>heit, daß mehr Weibs= als Manns=Personen ster-<lb/>ben, zu Tage
legen. Es ist von einem Ort hergenommen, wo <lb/> eine sehr <hi rendition="#aq">frequente
Academie flori</hi> ret, und also um so viel wich-<lb/>tiger vor unsern Satz. Diesemnach
sind zu Halle</p>
```

Korrektur  
fehlerhafter  
Strukturen

```
  <table>
    <row>
      <cell><lb/> Anno</cell>
      <cell>1729.</cell>
      <cell>gestorben</cell>
      <cell>501.</cell>
      <cell>Manns=Personen.</cell>
```

# MKHZ: Vervollständigung der Metadaten

## Metadaten

<b>Titel:</b>	Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil
<b>Untertitel:</b>	No. LVIII, 30. Woche
<b>Aut.-Daten:</b>	
<b>Ersch.-Jahr:</b>	1744
<b>Verlag:</b>	Johann Michael Funk
<b>Ort:</b>	Erfurt (Thüringen)
<b>erschienen in:</b>	Allerneueste Europäische Welt- und Staats-Geschichte, II. Theil
<b>Auflage:</b>	Erstausgabe

## zugehörige Werke

- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. IV, 2. Woche, Erfurt (Thüringen), 10. Januar 1744.
- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. V, 3. Woche, Erfurt (Thüringen), 13. Januar 1744.
- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. VI, 3. Woche, Erfurt (Thüringen), 17. Januar 1744.
- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. XII, 6. Woche, Erfurt (Thüringen), 7. Februar 1744.
- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. XVI, 8. Woche, Erfurt (Thüringen), 21. Februar 1744.
- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. XV, 9. Woche, Erfurt (Thüringen), 24. Februar 1744.
- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. XIX, 10. Woche, Erfurt (Thüringen), 2. März 1744.
- Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. XX, 10. Woche, Erfurt

## Informationen

<b>Quelle:</b>	MKHZ
<b>Publikationstyp:</b>	Zeitschrift/Zeitung
<b>Umfang:</b>	8 Scans ca. 16609 Zeichen ca. 2320 Tokens /[[:a1num:]]/ ca. 1173 Oberflächentypes
<b>Schriftart:</b>	Fraktur
<b>Genre:</b>	Gebrauchsliteratur :: Zeitung
<b>im DTA seit:</b>	2014-01-13 19:21:00
<b>zuletzt geändert:</b>	2014-05-13 16:28:46
<b>Verfügbarkeit:</b>	Text (TEI-XML-, HTML-, TCF-, E-Book-Fassung): Dieses Werk steht unter der „Creative Commons Namensnennung - Nicht kommerziell 3.0 Deutschland Lizenz“ (CC BY-NC). Weitere Informationen: Nutzungsbedingungen.
<b>Grundlage dieses Digitalisats:</b>	Institut für Deutsche Sprache, Mannheim: Bereitstellung der Bilddigitalisate und TEI Transkription Mikrofilmarchiv der deutschsprachigen Presse e.V., Dortmund: Bereitstellung der den Bilddigitalisaten zugrunde liegenden Microfilmaufnahmen Peter Fankhauser: Transformation von TUSTEP nach TEI P5. Transformation von TEI P5 in das DTA-Basisformat. Susanne Haaf: Artikelstrukturierung nach DTA-Basisformat. CLARIN-D: Langfristige Bereitstellung der DTA-Ausgabe



# MKHZ: Integration in das DTA

## nn\_weltgeschichte0258\_1744 (MKHZ)

offene Tickets: 0 (0 ganzes Buch)  
Stand: 2014-05-13 16:28:46

Text Text/Bild

Bild: 0001 : 457 erste Seite

1:1 fit 62%

457 (457) 365  
Der allerneuesten  
Europäischen  
Welt- und Staats-Geschichte  
II. Theil.

No. LVIII.

30. Woche.  
Freytags den 24. Julii, MDCCXLIV.

**Neuigkeiten von Teutschland.**  
An will die positive Nachricht haben, daß an dem  
Ruffischen Hof der Marquis di Borra obllig frey ge-  
sprochen sey, ja daß demselben so gar der St. An-  
dreas-Orden solte zugeschiedt worden seyn. Singe-  
gen hat man nach eben diesen Briefen sagen wollen, es sey dem Mar-  
quis de Chetardie der Hof und das Ruffische Reich verbotthen wor-  
den. Unter dem Articul von Rußland berichten wir ein mehreres  
davon.  
Vom Tod des Hrn. General Mengels sind folgende öffentli-  
che Nachrichten bekandt worden, die wir hier wiederholten. Man  
erzehlt, dieser brave Herr habe sich von seinem Jäger bereden lassen,  
daß er ihm erlaubt, zweymahl auf die jenseits Rheins gestandene  
Französische Wachen etwa 40. Schritte von ihm zu feuern, und  
anzügliche Reden auszustossen. Die Frankosen aber hätten so derb  
2. Th. 30. Woche. M m ge.

nächste Seite >>

```
<front>
  <pb facs="#f0001" n="457"/>
  <titlePart rendition="#c" type="main">
    <hi rendition="#g">Der allerneuesten</hi><lb/>
    <hi rendition="#fr">Europäischen<lb/>
    welt- und Staats-Geschichte</hi><lb/>
    <hi rendition="#aq">II</hi>
  .
  <hi rendition="#fr">
  <hi rendition="#g">Theil</hi>
  .</hi>
</titlePart>
<titlePart rendition="#c" type="sub">
  <hi rendition="#aq">No.<lb/>
  <figure/>
  LVIII.</hi>
</titlePart><lb/>
<titlePart rendition="#c" type="sub">30. Woche.
</titlePart><lb/>
<docDate rendition="#c">Freytags den 24. Julii,
<hi rendition="#aq">MDCCXLIV.</hi>
</docDate><lb/>
</front>
<body>
  <div type="jPoliticalNews">
    <div type="jArticle">
      <head>Neuigkeiten von Teutschland.</head><lb/>
      <p>
        <hi rendition="#in">M</hi>
        An will die positive Nachricht haben, daß an dem<lb/>
        Ruffischen Hof der
```

XML-Ansicht

Diese Ansicht stellt das zugrunde liegende XML dieser Seite dar.

Buchdaten

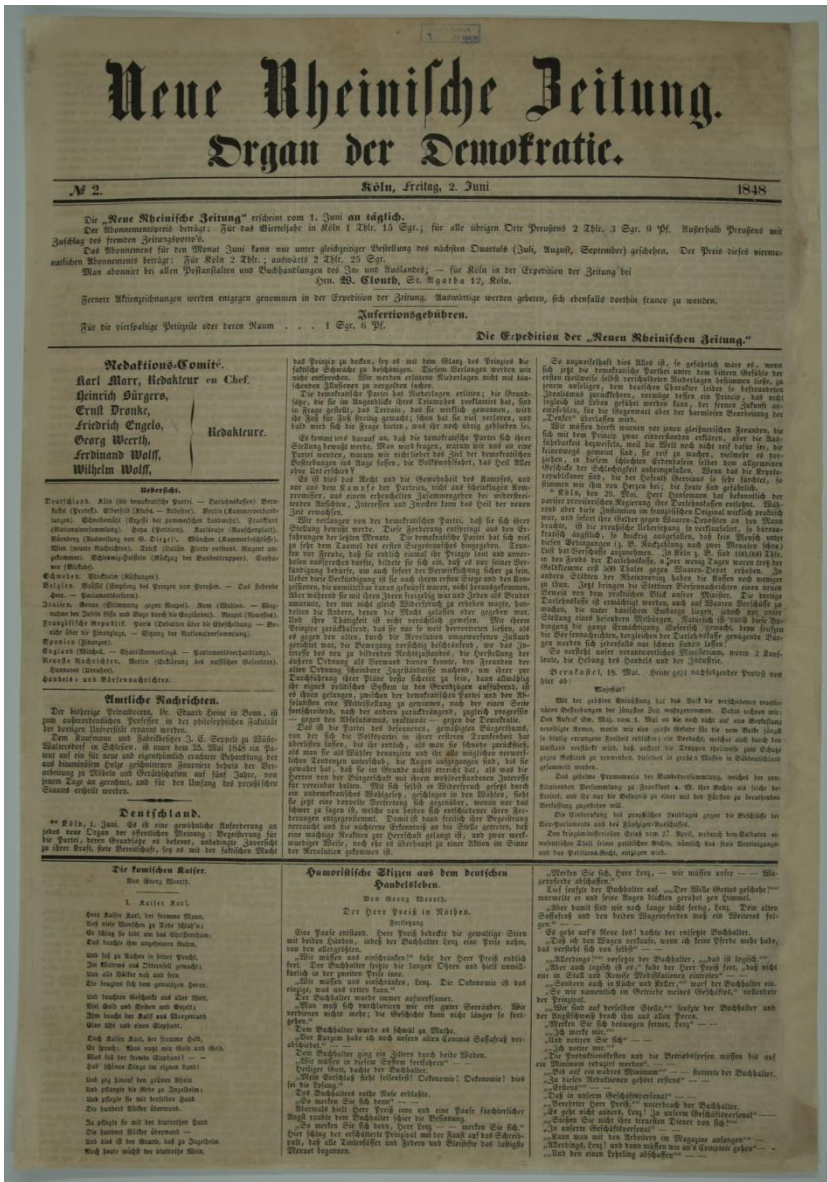
DTA-Informationen  
Metadaten  
Ansichten

Suche

Der allerneuesten Europäischen Welt- und Staats-Geschichte II. Theil. No. LVIII, 30. Woche, Erfurt (Thüringen), 24. Juli 1744. <http://www.deutschestextarchiv.de/dtaq/book/show/30555>

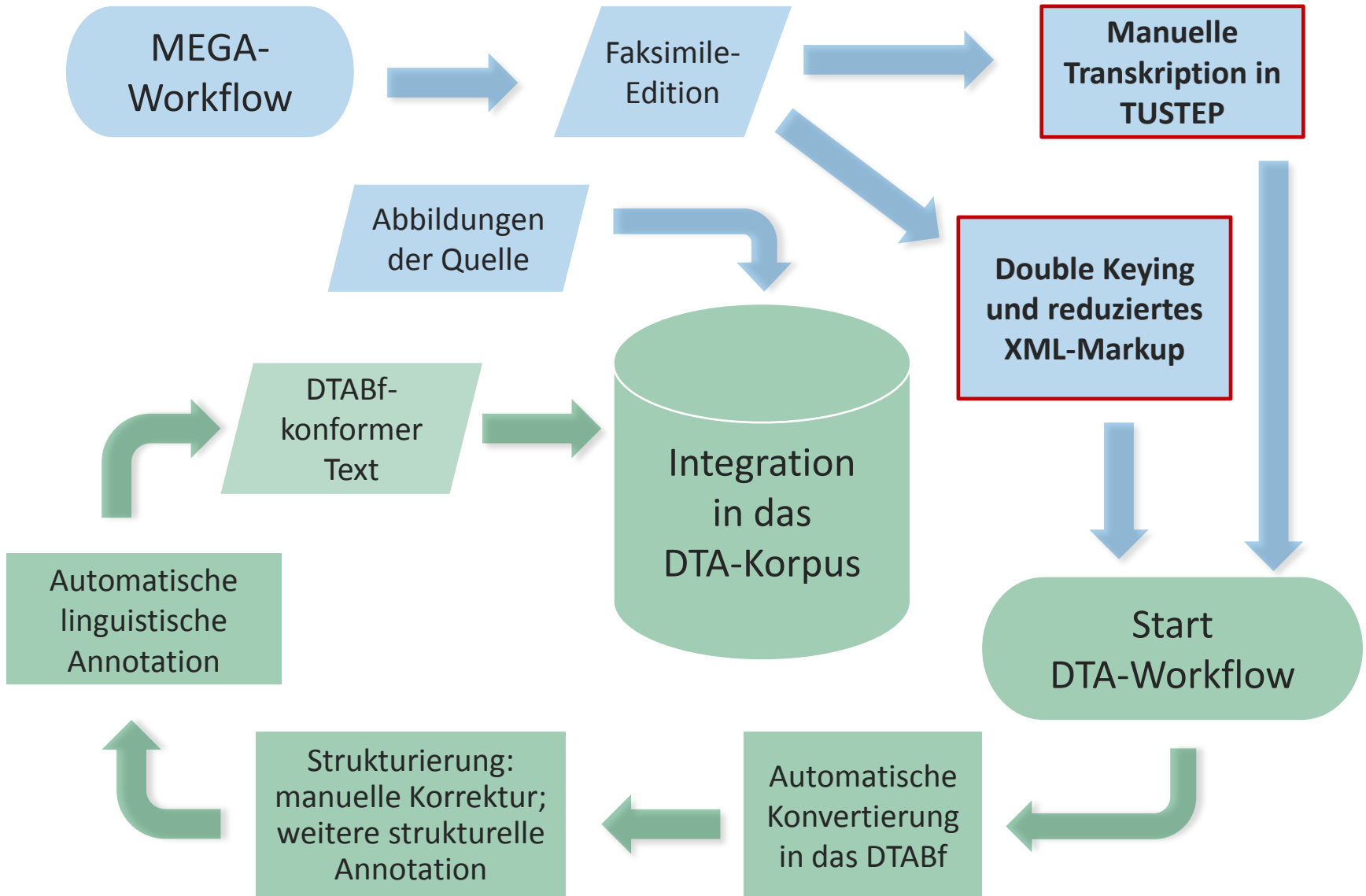
# 2. Neue Rheinische Zeitung (NRhZ)

- 301 Ausgaben (1718 Seiten)
- 1. Juni 1848 - 19. Mai 1849
- Vollständig digitalisiert im Rahmen des Editionsprojahens *Marx-Engels-Gesamtausgabe an der BBAW*



S. Haaf: Homogene Aufbereitung Historischer Zeitungen

# NRhZ: Workflow



# NRhZ: Nachbereitung im DTA

```
<d1><g>Deutschland</g></d1>
<artikel><d2><koz="**"/> <g>Köln</g>, 1. Juni.</d2></artikel>
<artikel><d2><koz="**"/> <g>Köln</g>, den 29. Mai.</d2>
<p>Herr Hansemann hat bekanntlich der pariser provisorischen Regierung ihre Darlehnskassen entlehnt. Während aber diese Institution im französischen Original wirklich praktisch war, und sofort ihre Gelder gegen Waaren-Depositien an den Mann brachte, ist die preußische Uebersetzung [...]
Darlehnskasse genügende Bürgen werden sich jedenfalls nur schwer finden lassen!</p>
<p>So versteht unser verantwortliches Ministerium, worin 2 Kaufleute, die Hebung des Handels und der Industrie.</p>
</artikel>
```

```
<div type="jPoliticalNews">
  <head><hi rendition="#g">Deutschland</hi></head>
  <div type="jArticle" xml:id="ar002_002">
    <head><bibl><author>**</author></bibl>
      <hi rendition="#g">Köln</hi>, 1. Juni.</head>
      <gap reason="copyright"/>
    </div>

  <div type="jArticle" xml:id="ar002_003">
    <head><bibl><author>*</author></bibl>
      <hi rendition="#g">Köln</hi>, den 29. Mai.</head>
    <p>Herr Hansemann hat bekanntlich der pariser provisorischen Regierung ihre Darlehnskassen entlehnt. Während aber diese Institution im französischen Original wirklich praktisch war, und sofort ihre Gelder gegen Waaren-Depositien an den Mann brachte, ist die preußische Uebersetzung [...]
    Darlehnskasse genügende Bürgen werden sich jedenfalls nur schwer finden lassen!</p>
    <p>So versteht unser verantwortliches Ministerium, worin 2 Kaufleute, die Hebung des Handels und der Industrie.</p>
  </div>
```

# NRhZ: Nachbereitung im DTA

```
<div type="jArticle" xml:id="ar005_006" next="#ar005_006b">
  <head><bibl><author>*</author></bibl><hi rendition="#g">Berlin,</hi> 2. Juni.</head>
  <p>Die demokratische Partei nimmt hier von Tag zu Tag an Macht und Einfluß zu. Eine Menge von
    Umständen hat unsere halb eingeschüchterte Bürgerschaft dem Volk genähert und eine neue
  [...]
  régime die Landwehr zu fanatisiren suchte, um sie als Mittel zur Verdrängung der Bürgerwehr
  und als Gegengewicht gegen sie zu gebrauchen. Er mußte im Stillen</p>
</div>
</div>
<div type="jFeuilleton">
  <div type="jArticle" xml:id="ar005_007">
    <head>Holländische Reisen <bibl>von<lb/>
      <author>Georg Weerth</author>.</bibl></head>
    <lg type="poem">
      <head>I.</head>
      <lg n="1">
        <l>Die bunten Flaggen winken —</l><lb/>
  [...]
<pb n="0018"/>
<div type="jArticle" xml:id="ar005_006b" prev="#ar005_006">
  <p>
    erröthen, wenn er hörte, wie die von ihm so verachteten Wiener nicht nur die
    Berliner Märzrevolution durch einen Volkssieg hervorgerufen, sondern auch Schlag
    auf Schlag jeden Reaktionsversuch durch siegreiche Aufstände vereitelt hatten.
    Alles das hat seine Wirkung gehabt. Die Berliner Bürgerschaft hat den
    Konservatismus, den ihr die Schrecken der Handelskrisis aufgedrängt, fahren
    lassen; sie spricht sich entschieden gegen die Regierung aus und schließt sich
    mehr und mehr der demokratischen Partei an.
  </p>
</div>
```



# NRhZ: Nachbereitung im DTA

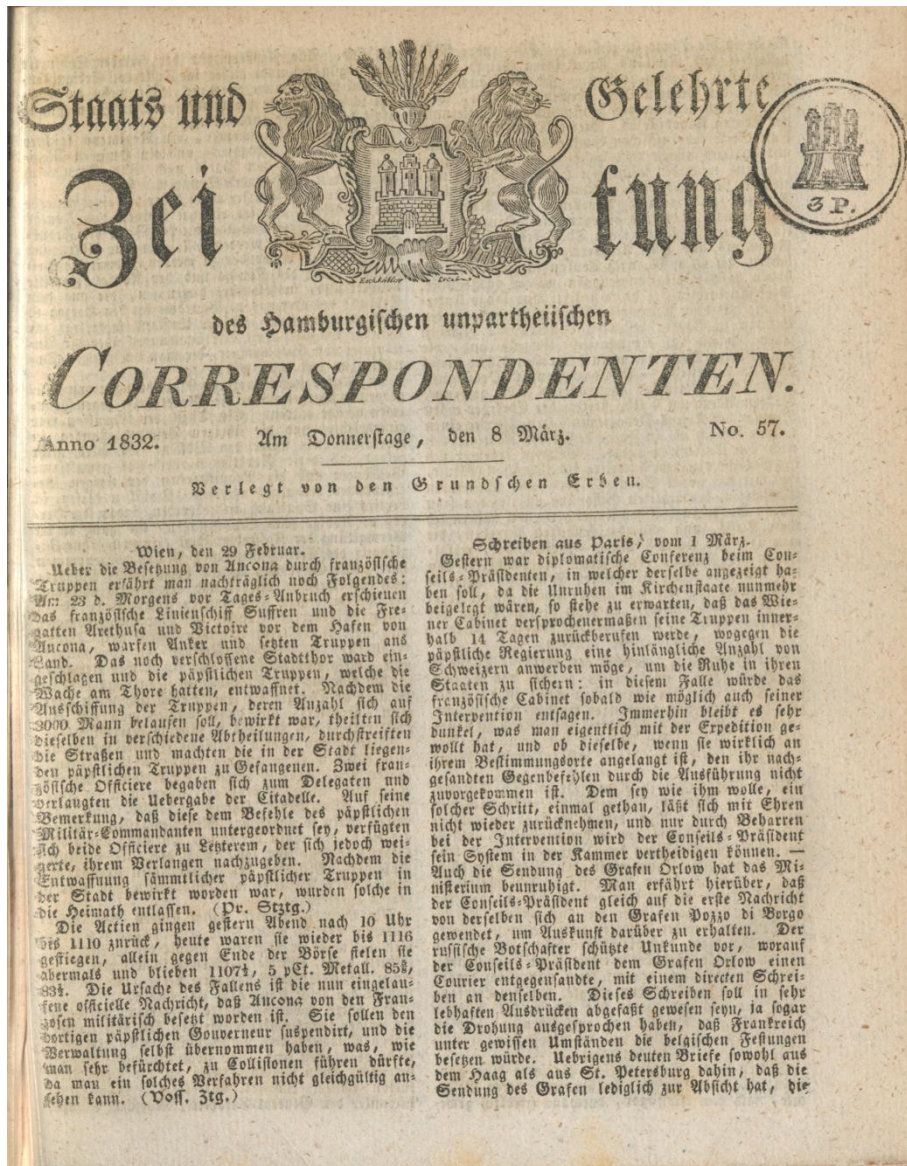
[\*] Berlin, 2. Juni.

Die demokratische Partei nimmt hier von Tag zu Tag an Macht und Einfluß zu. Eine Menge von Umständen hat unsere halb eingeschüchterte Bürgerschaft dem Volk genähert und eine neue Verbrüderung Beider vorbereitet. Der Berliner Bürger hat bei der Revolution theilweise mitgefochten, er ist stolz auf sie, und mußte erbittert werden durch die Art, wie Herr Camphausen sie am 30. Mai desavouirte, durch den Verfassungs-Entwurf, durch die heimliche Ausräumung des Zeughauses. Er mußte Feuer und Flammen sprühen, wenn er jeden Morgen in der Vossischen oder Spenerschen die zahllosen wuthschnaubenden Adressen aus hundert kleinen märkischen, pommerschen, westpreußischen, lausitzischen und sächsischen Oertern und Oertchen las, Adressen, in denen die Barrikadenkämpfer des 18. März, die noch dazu Berliner waren, von miserablen Provinzialen, von den Pfahlbürgern, Bauern, Beamten und Krautjunkern der norddeutschen Sahara mit den beleidigendsten Ausbrüchen reaktionärer Tollwuth überschüttet wurden. Er mußte sich aufs Aeüßerste verletzt fühlen, wenn er sah, wie die Partei des ancien régime die Landwehr zu fanatisiren suchte, um sie als Mittel zur Verdrängung der Bürgerwehr und als Gegengewicht gegen sie zu gebrauchen. Er mußte im Stillen

[zum nächsten Teil \(005.xml#ar005\\_006b\) ...](#)

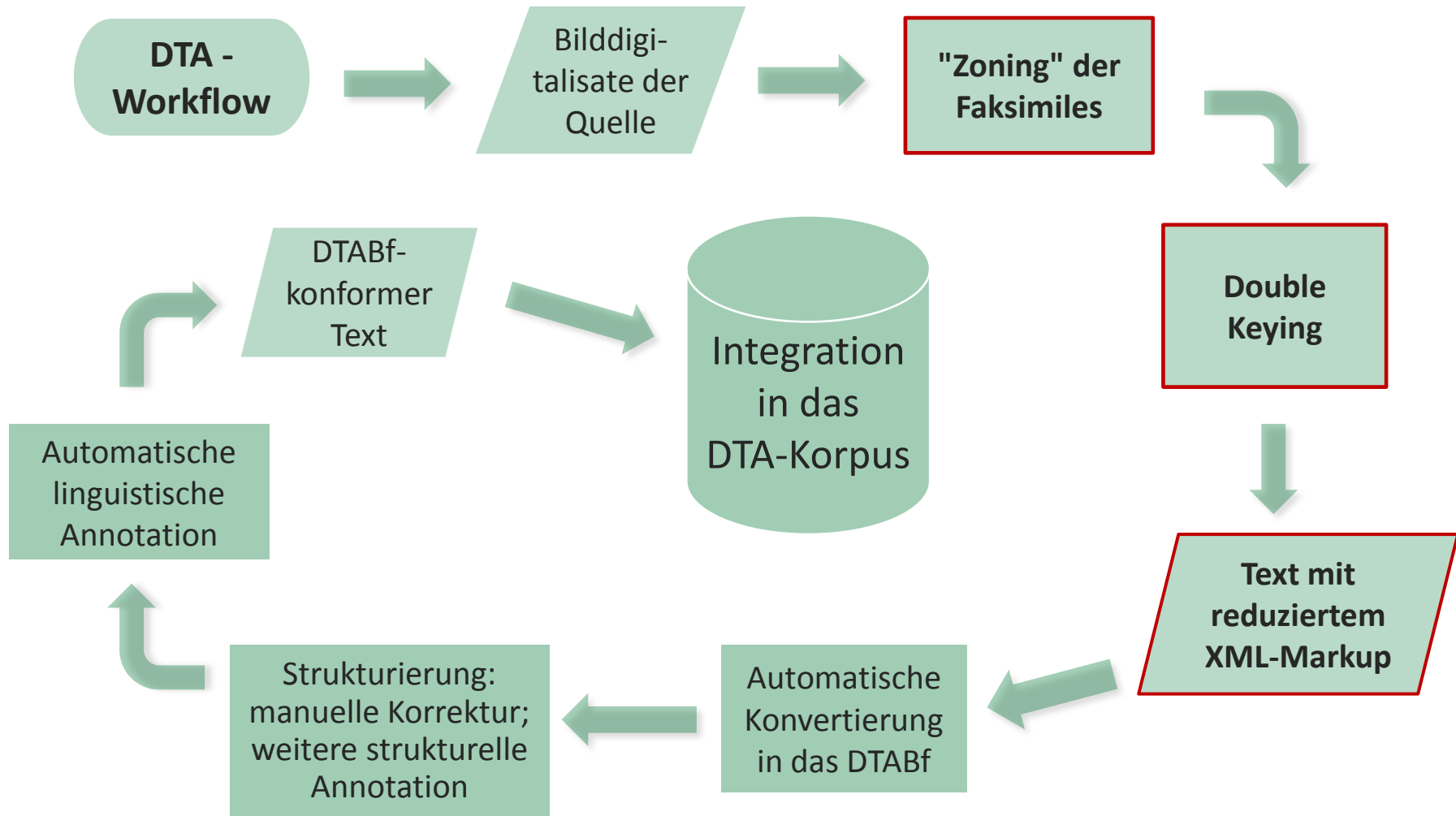
@xml:id	#ar005_006
@type	article
@facs	0017

# 3. Hamburgischer Correspondent



- *Staats und Gelehrte Zeitung des Hamburgischen unpartheiischen Correspondenten* und deren Vorgänger
- Digitalisierung von 320 ausgewählten Ausgaben
- 1712–1851
- Projekt an der Universität Paderborn (Prof. Dr. Britt-Marie Schuster), DTA ist Partner
- von Beginn an DTA-Workflow

# Hamburgischer Correspondent: Workflow



- Nachnutzung des Workflows für den Aufbau des DTA-Kernkorpus
- DTA-Qualitätssicherung von Beginn an



# Hamburgischer Correspondent: Qualität der Bildvorlagen

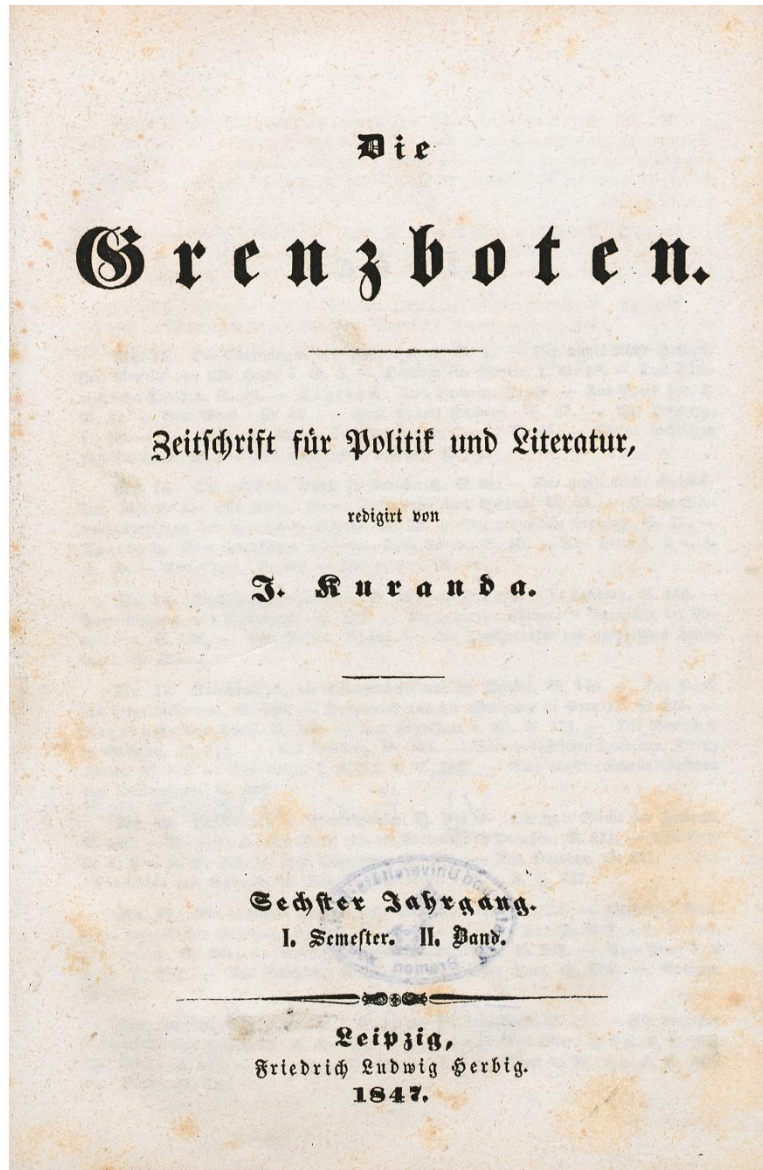


Wien, den 29 Februar.  
 Ueber die Besetzung von Ancona durch französische Truppen erzählt man nachträglich noch Folgendes: Am 22 d. Monats vor Tages-Anbruch erschienen Schreien aus Paris, vom 1 März. Gesehn war diplomatische Konferenz beim Con- seils-Präsidenten, in welcher derselbe anesiegt ha- ben soll, da die Kränken im Kirchenraate ummehr das Wort zu führen vermögen, das das Wite-

Monuments-Preis: Klotzschs Mart (für 1841, Postämter fünfzehn Mart). Infections-Gebühr: fünf S. Schillinge für die Preis-Beile. Einzelne Nummern dieser Zeitung sind täglich in der Expedition besorgen für Einen Schilling zu haben.

Paris, den 26 Juni.  
 (Mit der Raubpost über Antwerpen nach Amsterdam.)  
 Der Marine-Minister hat neue Instruktionen an die Häfen wegen Besetzung der Schiffe an den auf den Werften liegenden Schiffen erteilt, so wie Befehl, die seegefertigen nach der Levante zu schicken. Zu London haben diese Befehle einige Sen- sation gemacht.  
 Mehrere Blätter sprechen noch immer von einem Plane, die Kammer aufzulösen, indessen soll der Kö- nig persönlich gegen eine solche Maßregel eingenom- men.  
 glaubt man, daß Lord Palmerston deshalb stets auf seine Politik bestehe und den Absicht der Londoner Unterhandlungen hinsichtlich, weil er in den italienischen und marchonischen Anhängern die Zahl des Kaiserthums zu erkennen glaubt. — Was Frankreichs Stellung zum Papste in diesem Augenblicke betrifft, so ist be- hauptet worden, der gegenwärtige Minister der aus- wärtigen Angelegenheiten habe ihm sehr geantwortet, um ihn vor der englischen Politik zu warnen. Aber jeder, der den Gang der diplomatischen Geschäfte nie eingermessen kennt, weiß, daß eine so direkte Correspondenz zwischen dem Minister und dem Pa- pste in diesem Punkte für die nationale Partei der Türkei schmerzhaft und dieser Lage das französische Ge- dächtniß lebendiger werden, indem man anfragt, wie weit dies genügt sey, mit England gemeinschaft- lich die landwirthliche Annuveretten zu erörtern und die alte türkische Herrschaft wieder einzuleben. Von französischer Seite soll zwar entschieden erwidert worden seyn, Frankreich werde zwar im vorliegenden Falle nicht die Kontingenzen in ihrer Praxation zu förmlicher Unabhängigkeit unterstützen, aber auch seine Mitwirkung zur Wiederherstellung der Ruhe in Konstantinopel nicht verweigern, welche für alle

## 4. Die Grenzboten



- Erschienen zwischen 1841 und 1922; Wochenzeitschrift
- Gesamtumfang: 270 Bände, 180 000 Seiten
- Vollständig digitalisiert von der Staats- und Universitätsbibliothek Bremen  
→ OCR
- DFG-Projekt (SUUB, DTA) zur automatischen Korrektur und Strukturierung des OCR Output
- Ansprechpartner DTA: Kay-Michael Würzner, Christian Thomas

# Die Grenzboten: Semi-Automatische DTABf-Strukturierung

## grenzboten\_179382\_282158 (DTAE)

offene Tickets: 39 (1 ganzes Buch)  
Stand: 2014-05-20 12:21:42

Text Text/Bild   
2 - 196 - 126 0 - 0 - 324

Bild: 0005 << vorherige Seite

+ - 1:1 fit 100%



nächste Seite >>

## Die Grenzboten.

### Blätter für Deutschland und Belgien.

Redigirt von  
**J. Kuranda.**

Unter Mitwirkung der deutschen Schriftsteller:  
Karl Andree, Berthold Auerbach, Karl Beck, Baron A. von Bülow,  
Theodor Creizenach, Lorenz Diefenbach, F. Dingelstedt, I. Fester, Ludwig  
August Frankl, Carl Gutzkow, Heinrich Heine, J. Kaufmann, Heinrich  
Koenig, Gustav Kühne, Heinrich Laube, Hermann Marggraf, H. Merz,  
Julius Mosen, Theodor Mügge, R. E. Prutz, L. Schefer, H. Schiff,  
G. Schirges, Theodor Schliephake, Baron v. Sternberg, I. Venedey, Van  
Hasselt, A. Weill, Ernst Willkomm.

Erster Jahrgang.

**Leipzig**  
Bei Fr. Ludw. Herbig.  
**1841.**

Buchdaten

DTA-Informationen  
Metadaten  
Ansichten  
nächstes Ticket

Korrekturstatus

Text von mir kontrolliert

FrankWiegand, 2014-02-03T16:54

Text/Bild von mir kontrolliert

Darstellung von mir kontrolliert

TEI/XML von mir kontrolliert

Tickets für diese Seite

Suche

DDC nicht verfügbar

+

[http://www.deutschestextarchiv.de/dtaq/book/show/grenzboten\\_179382\\_282158](http://www.deutschestextarchiv.de/dtaq/book/show/grenzboten_179382_282158)

# DTA-Basisformat – Auszeichnung von Zeitungen

## Inhaltsverzeichnis dieses Dokuments

- 1 Allgemeines
- 2 Einleitende Textstücke
  - 2.1 Der Titelpopf
  - 2.2 Einführende Informationen
- 3 Erfassung und Strukturierung des Textkörpers
  - 3.1 Allgemeine Hinweise zur Erfassung und
  - 3.2 Arten von Artikeln
    - 3.2.1 Berichte/Nachrichten
    - 3.2.2 Feuilleton
    - 3.2.3 Anzeigen
    - 3.2.4 Leserbriefe
    - 3.2.5 Sonstige Artikel
  - 3.3 Strukturierung innerhalb eines Artikels
  - 3.4 Verknüpfungen diskontinuierlicher Artikel
- 4 Abschließende Textteile
- 5 Beilagen

## I ALLGEMEINES

Die Auszeichnung von Zeitungstexten erfolgt grundlegend nach den Regeln des DTA-Basisformats. Die vorliegende Spezifikation bezieht sich lediglich auf die Auszeichnung von Zeitungen, die im DTA-Basisformat nicht berücksichtigt sind.

Grundsätzlich werden Zeitungsausgaben als in sich

## 3.2 Arten von Artikeln

### 3.2.1 Berichte/Nachrichten

Berichte und Nachrichten verschiedener Art, die in der Quelle unter einem thematischen Aspekt gruppiert wurden (i.d.R. Politische Nachrichten, Wirtschaftsnachrichten, Wetterberichte), werden in entsprechend typisierten `<div>`-Elementen der obersten Ebene zusammengefasst.

Die einzelnen Artikel innerhalb dieser Rubriken werden jeweils wiederum durch `<div>`-Elemente tieferer Ebene umschlossen, welche das Attribut-Wert-Paar `@type="jArticle"` erhalten.

Kategorie	@type-Wert	Hierarchie-Ebene (empfohlen)
Politische Nachrichten inkl. telegraphische Berichte	<code>jPoliticalNews</code>	1
Wirtschafts- und Finanznachrichten	<code>jFinancialNews</code>	1
Wetternachrichten	<code>jWeatherReports</code>	1
Artikel/Nachricht	<code>jArticle</code>	2   3   ...

```
<body>
  <div type="jPoliticalNews">
    <div type="jArticle">[Politische Nachricht]</div>
    <div type="jArticle">[Weitere politische Nachricht]</div>
  </div>
  <div type="jFinancialNews">
    <div type="jArticle">[Wirtschaftsnachricht]</div>
    <div type="jArticle">[Weitere Wirtschaftsnachricht]</div>
  </div>
  <div type="jWeatherReports">
    <div type="jArticle">[Wetternachricht]</div>
    <div type="jArticle">[Weitere Wetternachricht]</div>
  </div>
</body>
```

Sind die Artikel innerhalb einer Rubrik wiederum gruppiert (z.B. Gruppierung der Artikel nach Nationalstaaten innerhalb der Politischen Nachrichten, i.e. "Deutschland", "Frankreich", etc.), so

# DTABf-Lösungen für die Artikelstrukturierung

---

DTABf "Kern":

```
<div type="...">
```

- poem
- act
- scene
- letter
- advertisement
- preface
- contents
- index
- ...

DTABf für Zeitungen

(zusätzliche @type-Werte):

- jArticle
- jPoliticalNews
- jFinancialNews
- jFeuilleton
- jAnnouncements
- jEditorialStaff
- jExpedition

# Zugänglichkeit der Daten über das DTA

- Bereitstellung unter freien Lizenzen (CC by-XXX)
- Text-Bild-Ansicht in DTAQ und (nach Qualitätssicherung) auf der DTA-Webseite
- Download in unterschiedlichen Formaten (TEI-XML, HTML, ePub, TCF, CMDI, ...)
- OAI-PMH und Fedora Repository  
→ Integration in die CLARIN-D-Infrastruktur

## Suche im Werk

 [Hilfe](#)

## Ansichten für dieses Werk

- Text-Bild-Ansicht
- alle Faksimiles
- DTAQ (Qualitätssicherung)
- Lese- bzw. E-Book-Fassung

## Download

XML (TEI P5) · HTML · Text

TCF (text annotation layer)

TCF (tokenisiert, serialisiert, lemmatisiert, normalisiert)

## Metadaten

TEI-Header · CMDI · Dublin Core

## Statistiken

Scans:	158
Zeichen:	ca. 192 503
Tokens:	ca. 28 323
Oberflächentypes:	ca. 6 564

## Wortwolken

- Lemmata
- Lemmata (nur Nomen)
- Types
- Types (nur Nomen)

# Beispiel Korpusanfrage in Zeitungen

1: [hc_351910_1712:3]	... und dieser Stadt; Und hierdurch möchten wohl	<b>die grosse Dessenen/</b> welche man bereits formiret gehabt ...
2: [anonym_annus_1597:91]	... ob jhnen solch Gut vnd Gelt/ durch	<b>die Türckische Fusten</b> genommen widerumb geliffert wurde/ darbey ...
3: [anonym_aviso_1609:316]	... vnsera Deputirte von Brüssel hiedurch nach Holland/	<b>die verschobene Puncten</b> folgend zuuergleichen/ verreisert/ Nemlich ...
4: [hc_341510_1712:3]	... bis hieher/ und treiben auch bis an	<b>die Navarrische Gräntzen</b> ihre Contributiones ein/ so daß ...
5: [anonym_aviso_1609:187]	... Ständen zum höchsten entschuldigt/ daß sie vmb	<b>die geringste Berachtschlagungen</b> nit wissen/ darzu auch nit ...
6: [hc_341510_1712:4]	... daß die Feinde nichts unternehmen können/ nachdem	<b>die Englische Truppen</b> sich separiret/ und zwischen Tarragona ...
7: [hc_352608_1721:5]	Was	<b>die Geldrische Deputirten</b> wegen Ausbesserung ihrer Gräntz-Vestungen vorgetragen, ...
8: [anonym_annus_1597:100]	... Johann von Sachsen/ Administrator der Chur Sachsen	<b>die Voitländische Lehen</b> in Prag empfangen.
9: [hc_1202907_1789:7]	... erste Abtheilung desselben wird eigene Aufsätze, --	<b>die zweyte Recensionen</b> enthalten.
10: [hc_982006_1789:5]	Jetzt schickte der	<b>die feindliche Truppen</b> commandirende Seraskier, Jbrahim Pascha, ...
11: [anonym_aviso_1609:192]	MAN thut mit verlangen beyderseits Zusammenkunfft/	<b>die streitige Puncten</b> zuerörtern/ wie auch die Confirmation ...
12: [anonym_aviso_1609:313]	.../ vber vorige Resolution/ nemlich/ daß	<b>die interessirte Fürsten</b> parirn sollen/ erfolgen.
13: [hc_672704_1725:1]	... so Jhro Königl. Majest. gegenwärtig zu Friedens-Zeiten an	<b>die enroullirte Matrosen</b> jährlich zahlen lassen, die sich ...
14: [anonym_annus_1597:127]	... wegen jhnen die Maltesische/ wie gleichfals auch	<b>die Florentinische Galleen/</b> ein zeithero, mit beraubung ...
15: [anonym_aviso_1609:146]	... wie auch das Newstadter Rahthauß/ so wol	<b>die gewöhnliche Landtstuben</b> verboten worden/ seyn sie nichts ...
16: [anonym_aviso_1609:369]	... Lüttig viel Becker vnd Kolengreber angenommen/ auch	<b>die verfallene Wagenmülen/</b> so der Spinola vor 3. ...
17: [anonym_aviso_1609:332]	.../ vnnd Commission sich des Tytuls vber alle	<b>die vnirte Provintzien/</b> als Hertzog von Geldern/ ...

Kontakt:

[haaf@bbaw.de](mailto:haaf@bbaw.de)

Zugang zu den DTA-Korpora:

[www.deutschestextarchiv.de](http://www.deutschestextarchiv.de)

Jeder kann mitmachen!:

[www.deutschestextarchiv.de/dtaq](http://www.deutschestextarchiv.de/dtaq)

[www.deutschestextarchiv.de/dtae](http://www.deutschestextarchiv.de/dtae)

Literatur:

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