



Qualitätssicherung im Deutschen Textarchiv

Frank Wiegand, Deutsches Textarchiv

3. DGI-Konferenz – Frankfurt am Main, 8. Mai 2014

- DFG-Projekt (2007–2015) an der BBAW Berlin
- Referenzkorpus für die gedruckte neuhochdeutsche Sprache von 1650 bis 1900; Erstausgaben; Double Keying und OCR
- alle Digitalisate in Text und Bild online frei verfügbar
- DTA-Basisformat (TEI/XML → HTML, Text, ePub, ...)
- ausführliche und verlässliche Metadaten
- automatisch linguistisch annotiert und durchsuchbar
- Partner in CLARIN-D (Infrastrukturen für Geistes- und Sozialwissenschaften, BMBF-gefördert)

www.deutschestextarchiv.de

Suche: ADJ* „Mensch“

[kuernberger_amerikamuede_1855:73]	Der	allmächtige	Schuft hat wahrscheinlich eine ...
[garve_sammlung_1779:166]	... und die Natur der	ersten	Dichter verschiedene Werke hervorgebracht ...
[moritz_goetterlehre_1791:160]	... wird, in hoher	dichterischer	Schönheit dar.
[hippel_lebenslaeufer0302_1781:128]	... Begräbniß versammelten sich die	besten	Sänger und Sängerinnen im ...
[tieck_phantasus02_1812:317]	... das da ist ein	böser	Bube , ein Satiriker ...
[hippel_lebenslaeufer01_1778:140]	... er todt, der	geschickte	Mann! Curland verliert ...
[berg_ostasien03_1873:291]	... Jesus, der himmlische	ältere	Bruder , ist wirklich ...
[winckelmann_kunstgeschichte02_1764:62]	Antiochus IV. der	jüngere	Sohn Antiochus des Großen ...
[arnimb_guenderode02_1840:137]	... und Himmel auf der	morschen	Leiter , aber die ...
[abschatz_gedichte_1704:226]	Geh/	frecher	Jäger/ geh' und ...
[dilthey_geisteswissenschaften_1883:12]	... nacharistotelische Metaphysik und ihr	subjektiver	Charakter 305
[iffland_jaeger_1785:111]		Lieber	Herr!
[justi_velazquez01_1888:169]	... Joseph, die gemüthliche	alte	Frau , der Junge ...
[burckhardt_cicerone_1855:871]	... sich die Strenge der	alten	Niederländer in eine milde ...
[treitschke_geschichte03_1885:379]	... 1825, wähten die	unbedingten	Anhänger Oesterreichs schon einen ...
[humboldt_aequinoktial02_1859:47]	... von Luft und Sonne	gebräunte	Weißer sein möchten, ...
[haller_anfangsgruende01_1759:621]	... übrigen hat auch ein	berühmter	Mann aus den Herzohren ...
[boelsche_liebesleben01_1898:138]	... belastet offenbar mit einem	festen	Erbe dieses Individuums, ...
[arnold_ketzerhistorie02_1700:155]	... verleumdungen der falsch genanten	Lutherischen	Prediger zu Amsterdam/ ...
[carus_zoologie_1872:505]	..., wie Linne von	armen	Eltern geboren, bezog ...
[blum_spatziergaenge01_1774:174]	..., oder von dem	schönen	Jäger Endymion gehört?
[benner_herrnhuterey02_1747:28]	... ein anders ist ein	herrnhutischer	Jünger oder Apostel, ...

Warum Qualitätssicherung?

- Textgenauigkeit:
 - Double Keying: 99,95 % (5 Fehler auf 10 000 Zeichen)
 - OCR (Fraktur): <99 % (siehe Google ngrams)
- Metadaten sind essentiell für Forschungsarbeit
- jeder soll mitmachen können, ohne Expertenkenntnisse („Crowd“)
- Änderungen müssen schnell sichtbar werden
- DTA ist kein Selbstzweck (DTA-Basisformat sichert Nachhaltigkeit und Interoperabilität)

- Transkriptionsfehler, z. B. *Unter**b**altung* → *Unter**h**altung*
- Annotationsfehler, z. B. `<stage>` fehlt bei Bühnenanweisung im Drama
- Druckfehler, z. B. „Ich **u**ud meine **e** Hund.“
- Präsentationsfehler (XSLT, HTML, Javascript)
- Fehler im Workflow, z. B. falsch beschnittene Bilder
- Fehler in den Metadaten
- Fehler in der linguistischen Analyse:
„Mafchiene“ → „Maschiene“ Ma/GN#Schiene[_NN]

- webbasiertes Framework/Annotationstool
- Textdigitalisate in seitenweiser Darstellung, verschiedene Ansichten
- Annotation fehler-/zweifelhafter Passagen (Ticketsystem, Benutzergruppen)
- Korrekturlesen, direkt korrigieren
- linguistische Tools integriert
 - orthographische Normalisierung
 - Lemmatisierung, Part-of-Speech-Analyse

Parallele Ansicht: Bild – HTML

marperger_handelsdiener_1715 (CN)

offene Tickets: 3 (0 ganzes Buch)
Stand: 2014-03-10 14:41:07

Text Text/Bild 
0 - 0 - 636 0 - 0 - 636

Bild: 0005 << vorherige Seite

nächste Seite >>

  1:1 fit 50%



**Paul Jacob Marpergers/
Königl. Polnischen und Chur-Säch-
sischen Hof- und Commerciens-Raths/ wie
auch Mit-Glieds der Königl. Preussischen So-
cietät der Wissenschaften/
Getreuer und Geschickter
Handels-Diener/**

In welchem vornemlich
Was ein Handels-Diener sey/
was derselbe vor Qualitäten und
Wissenschaften an sich haben müsse/ dann
auch von dem dergleichen Handels-Dienern im
Fall ihres Wohl- oder Ubel-Verhaltens zukommen-
den Recht/ gehandelt/ durchgehends aber ein reicher
Vorrath allerhand heilsamer Lehren und Unterrich-
tungen/ gegeben wird/ deren sich Lehr-begierige
Handels-Diener sowohl zu Haus/ als auf Rei-
sen mit sonderbahren Nutzen bedienen
können.

Nürnberg und Leipzig/
zu finden bey Peter Conrad Monath.
An. 1715.

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

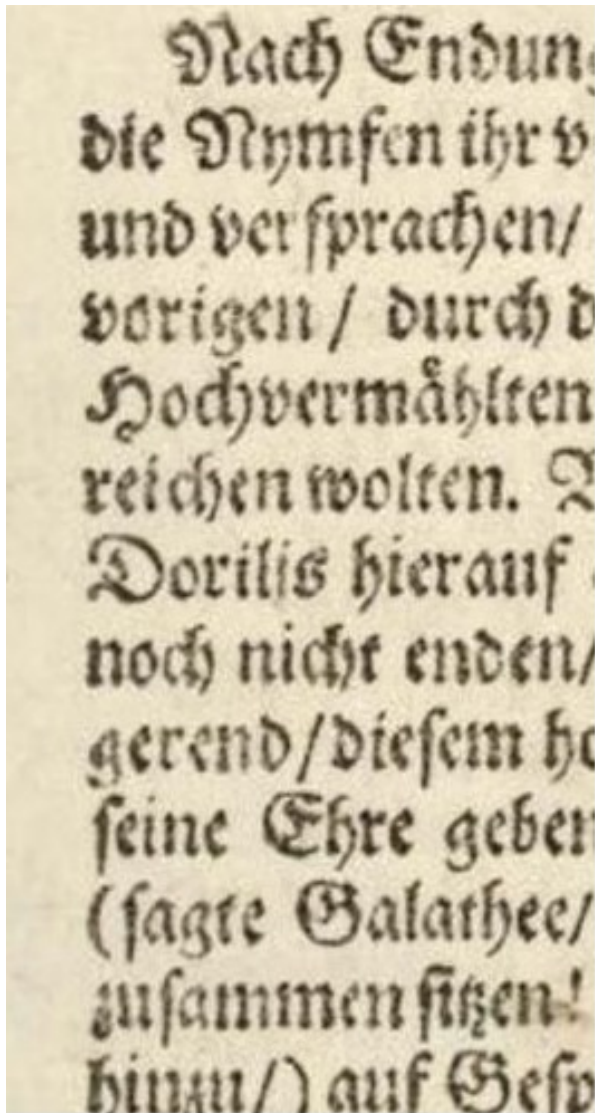
Suche

 Anmerkung zu dieser Seite anlegen

- Text und Bild vergleichen
oder:
- Text lesen, bei Auffälligkeiten: Blick aufs Bild
- je nach Anforderungen, Kenntnisstand
- vermerkbar, nachvollziehbar
- Qualität wird messbar

Ticket anlegen (Transkriptionsfehler)

dort in den Elyfer-Wäldern.



Nach Endung
die Nymfen ihr ve
und **verprochen** /
vorigen / durch die
Hochvermählten
reichen wolten. W
Dorilis hierauf an
noch nicht enden,
gerend / diesem ho
seine Ehre geben.
(sagte Galathee /
zusammen sitzen!
hinzu /) auf Gespr
wird / so ist vonnö
zum Haupt und G
de. Es braucht kei
Nymfe! sagte Flor

TEI-XML von n

Neues Ticket

***Typ:**

***Zusammenfassung:**
bei Transkr.-
/Schreibfehler: (**nur!**) die
korrekte Form

betrifft: Seite 86
 ganzes Buch

Fundstelle: *markieren Sie die Fundstelle mit der Maus*

Priorität:

Relevanz:

zuweisen an:

- Textgrundlage ändern ohne XML-Kenntnisse
- direkt im Browser (WYSIWYG)
- Freischaltung „auf Zuruf“
- Versionsverwaltung via `git`:
 - Änderungen sind dokumentiert
 - // // nachvollziehbar
 - // // kontrollierbar

```
22 <milestone rendition="#hr" unit="section"/><lb/>
23 <div type="diaryEntry" n="2">
24   <head>Gute Nachrichten aus
25     <placeName>Stuttgart</placeName>.</head><lb/>
26   <p>
27     <persName>Graf Leutrum</persName>, der bisherige
28     Intendant des Theaters, hat seine Entlassung
29     erhalten.<lb/>
30     Man soll den Todten nichts Böses nachsagen, aber
31     ein Exemplar, wie der Exintendant,<lb/>
32     ist selten zu finden. <persName>Graf
33     Leutrum</persName> ist derselbe Mann, der, als
34     einst <persName>Immermann</persName> in<lb/>
35     <placeName>Stuttgart</placeName> einige Tage
36     verweilte, und Jemand ihm die Nachricht brachte,
37     <persName>Immermann</persName><lb/>
38     sei da, darauf antwortete: Verhüten Sie, daß er
39     mich besucht, ich kann ihn nicht auf-<lb/>
40     treten lassen, alle Gastspiele sind schon vergeben.
41     Der edle Graf glaubte, <persName>Immermann</persName>
42     <lb/>sei ein reisender Schauspieler. Es ist in
43     <placeName>Paris</placeName> kein einziges Theater,
44     auch nicht unter<lb/>
45     den Boulevardstheatern, wo der Director
46     <persName>Alexander Dumas</persName>, oder
47     <persName ref="http://d-nb.info/gnd/11859737X">Edgar
48     Quinet</persName> für ei-<lb/>
49     nen Schauspieler nähme. Die deutschen Hofbühnen sind
50     mit solchen Chefs stark gesegnet.<lb/>
51     Und man will dann einen Aufschwung des Theaters! Die
52     Stuttgarter Bühne kann sich<lb/>
53     zu ihrem neuen Intendanten Glück wünschen, als
54     solcher ist Baron Taubenheim ernannt<lb/>
55     worden: ein Mann voll Kenntnisse, Geschmack und nobler
56     Gesinnung. Es ist dieß der-<lb/>
```

Sonderzeichen

ı	ı	ı	ı	ı	ı	ı	ı
À	à	Á	á	Â	â	Æ	æ
É	é	Ê	ê	Ë	ë	Ï	ï
Ò	ò	Ó	ó	Ô	ô	Œ	œ
Ù	ù	Ú	ú	Û	û	ü	ü
,	'	„	“	›	‹	»	«
½	⅓	⅔	¼	¾	⅕	⅔	⅜
⅙	⅚	⅛	⅞	⅜	⅝	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	#ln	#blue	#red	#et		
-#c-	-#right					

CTRL+E für oXygens „wrap tag“-Funktion benutzen.

Aktionen

validieren	speichern	verwerfen
------------	-----------	-----------

Hinweise

[Dokumentation DTA-Basisformat](#)

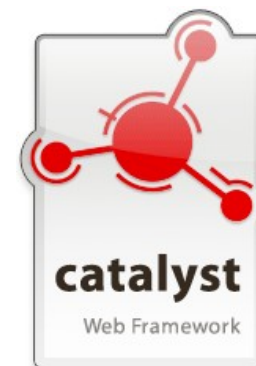
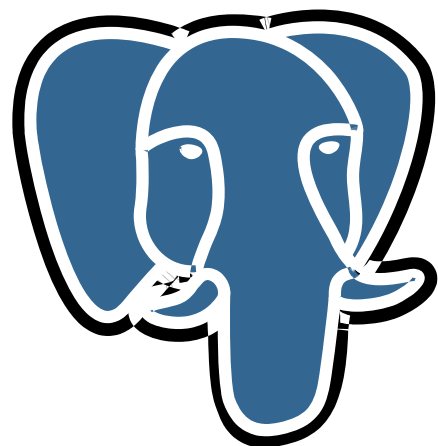
- Framework für oXygen XML Editor
 - für „größere“ Änderungen
 - bei neuen Transkriptionsprojekten
 - DTABf mit verschiedenen Levels integriert
 - direkte Visualisierung von Annotationen
 - frei zum Download verfügbar:
 - www.deutschestextarchiv.de/doku/software#dtaox

- Im Einsatz seit Juni 2011
- 1 991 Werke, 554 479 Seiten, 931 Mio. Zeichen
- ca. 64 500 Tickets angelegt (ca. 58 700 gelöst)
- ca. 32 200 Seiten Korrektur gelesen

- ca. 400 aktive Nutzer
- DTAQ ist nach Anmeldung frei zugänglich

- HAB Wolfenbüttel (Theatra, Oberhofprediger)
- AEDit Frühe Neuzeit (Leichenpredigten)
- SuUB Bremen („Die Grenzboten“)
- Digitale Faust-Edition
- Blumenbach Online
- CLARIN-D-Partner (Kurationsprojekt F-AG 1)
- „Privatanwender“ (Nachfahren J. J. v. Littrows, Jean-Paul-Freunde, ...)

DTAQ nutzt Open Source



Danke!

- <http://www.deutschestextarchiv.de/dtaq>
- @textarchiv
- +DeutschestextarchivDe
- wiegand@bbaw.de