



z^e M

Brandenburgisches Zentrum
für Medienwissenschaften

(Meta-)Daten bei Film, Video und bewegtem Bild – Recherche, Bereinigung, Visualisierung mit OpenRefine und Excel

22.-23. Januar 2026
Potsdam



\mathbf{z}^e_m

Brandenburgisches Zentrum
für Medienwissenschaften

Herzlich Willkommen!



Fachhochschule Potsdam
University of
Applied Sciences



FILMUNIVERSITÄT
BABELSBERG
KONRAD WOLF



GESELLSCHAFT
FÜR INFORMATIK



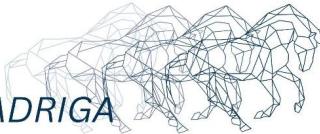
Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt

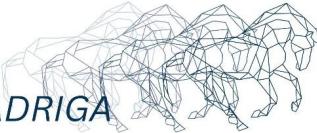


**Finanziert von der
Europäischen Union**
NextGenerationEU



Zeitplan Tag 1

10.00 Uhr	Begrüßung
10.15 Uhr	Einführung
11.15 Uhr	Vorstellung Fallstudie Filmuniversität
11.30 Uhr	Pause
11.45 Uhr	Metadaten Datenrecherche
12.50 Uhr	Mittagspause (Seoul Garden)
14.00 Uhr	Datenbereinigung: Warum und wann? Beispielprojekte Datenbereinigung mit OpenRefine (OER) (Pausen nach Bedarf)
17.15 Uhr	Tagesabschluss
17.45 Uhr	Gemeinsames Abendessen (Genusswerkstatt)



QUADRIGA

Disziplinen

- Digital Humanities
- Verwaltungswissenschaft
- Informatik
- Informationswissenschaft

Datentypen

- Text
- Tabelle
- Bewegtes Bild

Entwicklung von
Open Educational Resources (OER)



Wer sind wir?





Handout

https://quadriga-dk.github.io/Metadaten_Film_Workshop_2026/

Wer sind wir?



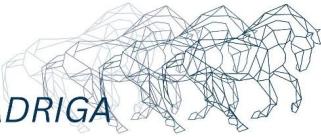
Deutschlandkarte: Von C. Busch, Hamburg - Eigenes Werk, CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=10031875>



Wer sind wir?



Berlinkarte: Von TUBS - Diese Datei enthält Elemente, die von folgender Datei entnommen oder adaptiert wurden; CC BY-SA 3.0, <https://commons.wikimedia.org/w/index.php?curid=20837459>
Potsdamkarte: Von TUBS - Eigenes Werk mittels: Stadt Potsdam: Arbeitslosigkeit in der Landeshauptstadt Potsdam 2006. in der Wayback Machine, OpenStreetMap Data (© OpenStreetMap und Mitwirkende, CC-BY-SA); 2021 revisions: Stadt Potsdam: Statistischer Jahresbericht 2019, CC BY-SA 2.0, <https://commons.wikimedia.org/w/index.php?curid=105681978>



Wer sind wir?



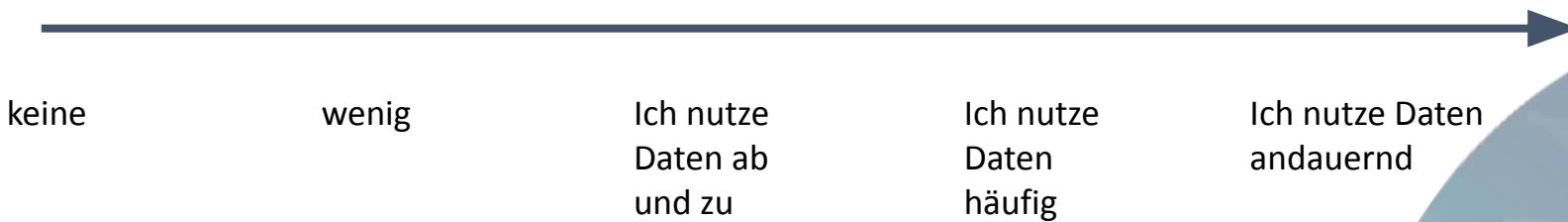
Meine Disziplin ist:



QUADRIGA

Wer sind wir?

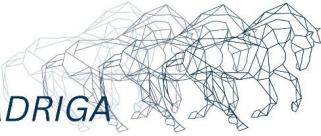
Erfahrung mit Daten (in der Forschung)





(Forschungs-)Daten

Was sind meine Forschungsdaten?



Datenkompetenz

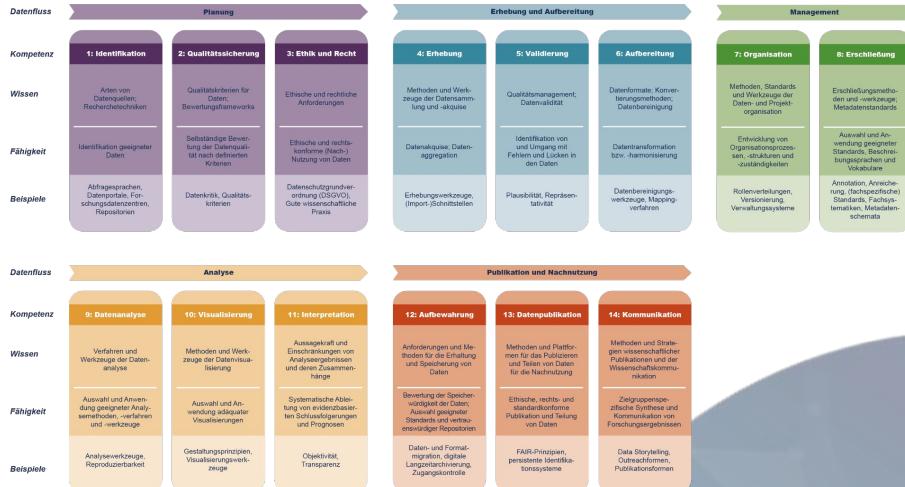
Was ist Datenkompetenz?

„Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner“
(Ridsdale et al., 2015, p. 8)

QUADRIGA Datenkompetenzframework

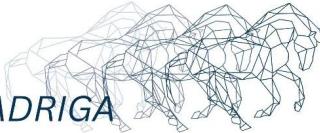
Das QUADRIGA Datenkompetenzframework verfolgt ebenfalls das Ziel, die wissenschaftliche Integrität und das Reflexionsvermögen zu fördern und eine grundlegend offene Haltung (gemäß dem Leitprinzip so offen wie möglich und so geschlossen wie nötig) für das Teilen und die Nutzung von Forschungsdaten zu unterstützen. Dafür werden in jedem Schritt des datenbasierenden Forschungsprozesses (vgl. Datenfluss) notwendige Kompetenzen vermittelt.

Hinweis: (Von-)Kenntnisse in gründlegenden Konzepten von Forschungsdatenmanagement und Open X sowie Wissen um Anwendungsbereiche von Daten.



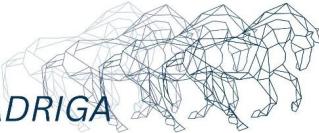
Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetrich, B. (2015). Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report (S. 123) [Report]. Dalhousie University. <https://doi.org/10.13140/RG.2.1.1922.5044>





**Welche Veränderungen sind in den studentischen Filmen
der Filmuniversität Babelsberg KONRAD WOLF bzw. der
Hochschule für Film und Fernsehen (HFF) "Konrad Wolf"
Potsdam zur Wendezeit erkennbar?**

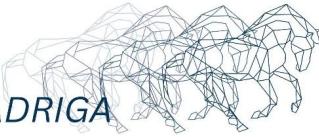




Korpusbildung

- Eingrenzung des Untersuchungszeitraums: 1985 bis 1999
- Recherche im Katalog
- Konkretisierung und Operationalisierung der Fragestellung
- Analyse des Datenmodells

20.sonst.Person	Hornemann, Klaus [Mischung]
21.sonst.Person	Otto, Joachim [Szenografe]
22.sonst.Person	Kraft, Uwe [Prod.-Ilg]
Titel	Stilles Land
Zusatz	Diplomfilm (Regie)
Personenang.	Regie: Andreas Dresen
Verweisung	AT: Provinztheater
Ort	Potsdam ; Berlin ; Leipzig
Verlag/Prod.	Hochschule für Film und Fernsehen "Konrad Wolf" : Max Film
Jahr	1991/1992
Umfang/Format	[35mm] (ca. 98:00min) ; color ; COMOPT
Allgem. Anmerkung	Koproduktion HFF / Max Film Wolfgang Pfeiffer Filmproduktion (Die 35mm Filmkopie liegt nicht an der HFF vor. Eine Filmkopi © MAX FILM 1992 © 2007 Filmgalerie 451, für diese DVD-Ausgabe enthält: "Stilles Land" (1991), "Konsquenzen - Peter, 25" (1987)
Schlagwort	Spielfilm
Stichwörter	Wende/A
Zusammenfassung	Wendezeit Andreas Dresens Debüt spielt in den bewegten Herbsttagen d Kai soll dort zum ersten Mal inszenieren – ausgerechnet "Wart auf die Situation. Ein Ensemblemitglied ist indessen zur ungar banalen Hindernissen; nur die Assistentin Claudia, in die Kai s kluge Marketing-Ratschläge unterbreitet. (Filmportal)



Korpusbildung

Datenerhebung und Bereinigung mittels *OpenRefine* und *Python*

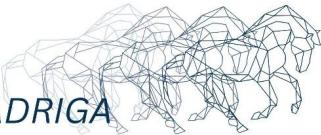
```
*****  
  
Satzkennung  
00000nM2.01200024-----h  
  
FMT  
MH  
  
ID  
HF0102406  
  
Erfass.dat  
20000814  
  
Korr.dat  
20161014  
  
Reg. ID  
HF0102406  
  
Code. Ang.  
az1ddzzc |37  
  
Datenträger  
|||||ba|||||
```



1	1	verantw._Pers._1	verantw._Pers._2	verantw._Pers._3
2	Titel			
3	Nacht in Leipzig, Eine	Olga Mielke		
4	Leipzig 1998			
5	Brief blieb ungeöffnet, Der			
6	Liste war unvollständig, Die	Danny Weber		
7	Niemand nahm den Auftrag mit	Waltraud Schulz		
8	Als der Winter kein Ende nahm	Hoang Minh Tuan		
9	Ruhe vor dem Wort, Die	Xenia Abendroth		
10	Atlas der verlorenen Städte, Der			
11	Zwischen Frühlicht und Flucht	Benjamin Christ		
12	Als der Lehrer streikte	Frieda Eggert		
13	Trennung	Volker Uhl		Col 4: -----
14	Abmeldung war nicht nötig	Benjamin Christ		
15	Im Schatten der Falte	Simone Schwarz		
16	Heft 6 Ein Versuch	Merten Oertel		
17	Melancholie der leeren Schulhöfe, Die	Rosemarie Quade		
18	Treppengeländer löse			
19	Niemand sang die Hymne			

Quelle:
Screenshot Filmarchiv Datensatz

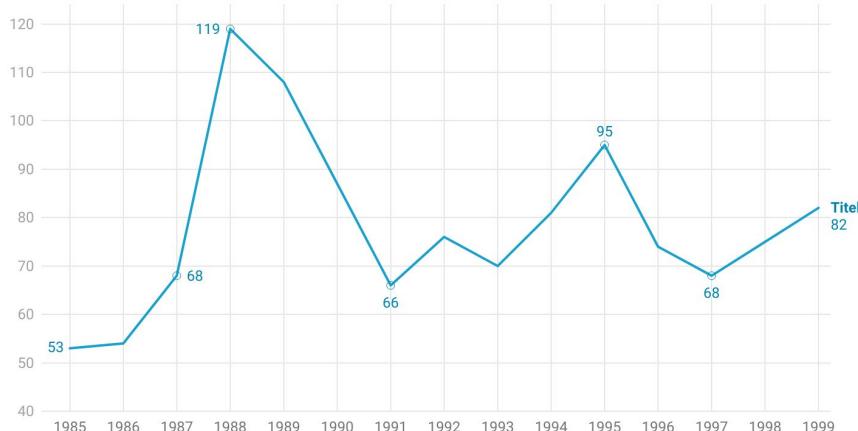
Quelle:
Screenshot bereinigter Filmarchiv Datensatz



Auswertung des Korpus

Auswertung von Metadaten

Titel pro Jahr 1985-1999

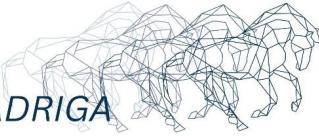


Gattungen pro Jahr 1985-1999





Pause



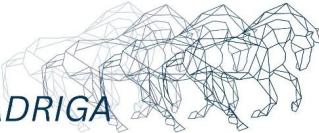
Metadaten

- Was sind (Meta-)Daten?
- (Typische) Eigenschaften von Metadaten



Metadaten

Was sind (Meta-)Daten? – Daten



Metadaten

Was sind (Meta-)Daten? – Daten

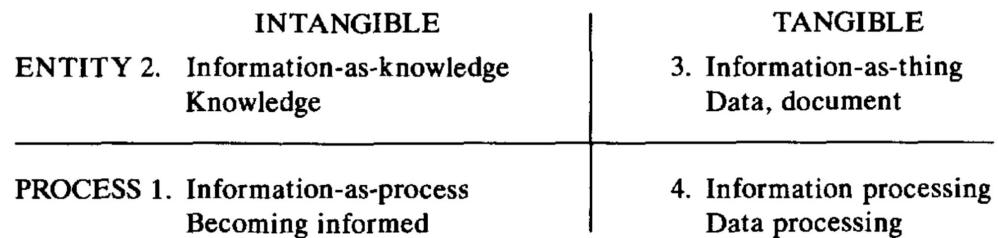
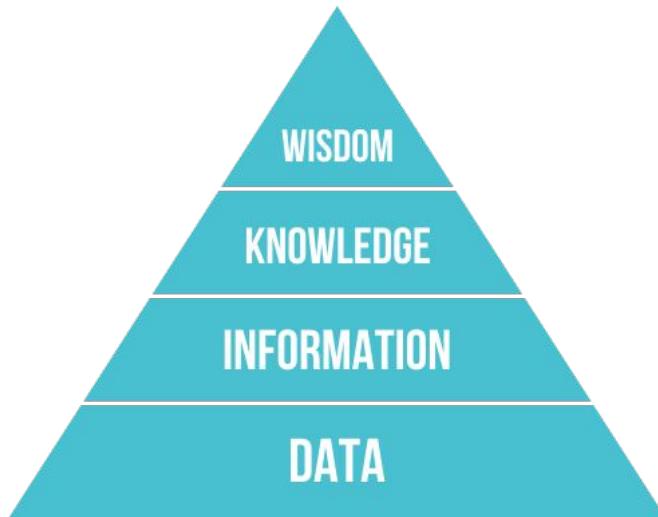
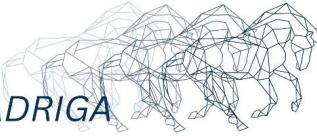


FIG. 1. Four aspects of information.

Buckland, M. K. (1991). Information as thing. Journal of the American Society for Information Science, 42(5), 351–360.
[https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5<351::AID-ASI5>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5<351::AID-ASI5>3.0.CO;2-3)
Digital Object Identifier (DOI)



Metadaten

Was sind (Meta-)Daten? – Daten

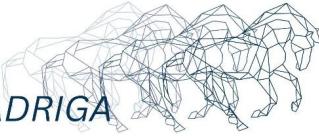
- (Mess-)Werte, die (mit einer gewissen Struktur) festgehalten sind
- Forschungsdaten:
„Unter digitalen Forschungsdaten verstehen wir dabei alle digital vorliegenden Daten, die während des Forschungsprozesses entstehen oder ihr Ergebnis sind.“ (Kindling und Schirmbacher, 2013, p. 130)



Metadaten

Was sind (Meta-)Daten? – Metadaten

- Daten über Daten
- Beschreibend, strukturell, administrativ, ...
- Was?, Wie?, Wer?, ...

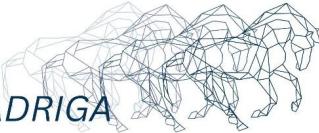


Metadaten

Eigenschaften von Metadaten

- Eher knapp
- Stark strukturiert
- Vorgegebene Vokabulare

- Metadatenschemata



Metadaten

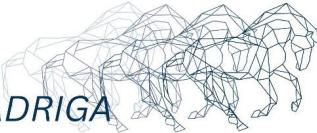
Eigenschaften von Metadaten

- Eher knapp
- Stark strukturiert
- Vorgegebene Vokabulare

- Metadatenschemata

Metadaten werden zu (Forschungs-)Daten, wenn man sie so benutzt.





QUADRIGA

Metadaten Film

- Filmografische Angaben
 - Filmtitel
 - Entstehungsjahr
 - Entstehungsland
 - Länge
 - Material
 - Beteiligte Personen
 -
- Frage der Vollständigkeit und Verlässlichkeit der Angaben
- Rolle des Entstehungskontextes
- Eingeschriebene Wertvorstellungen und Weltsichten



Metadaten Text

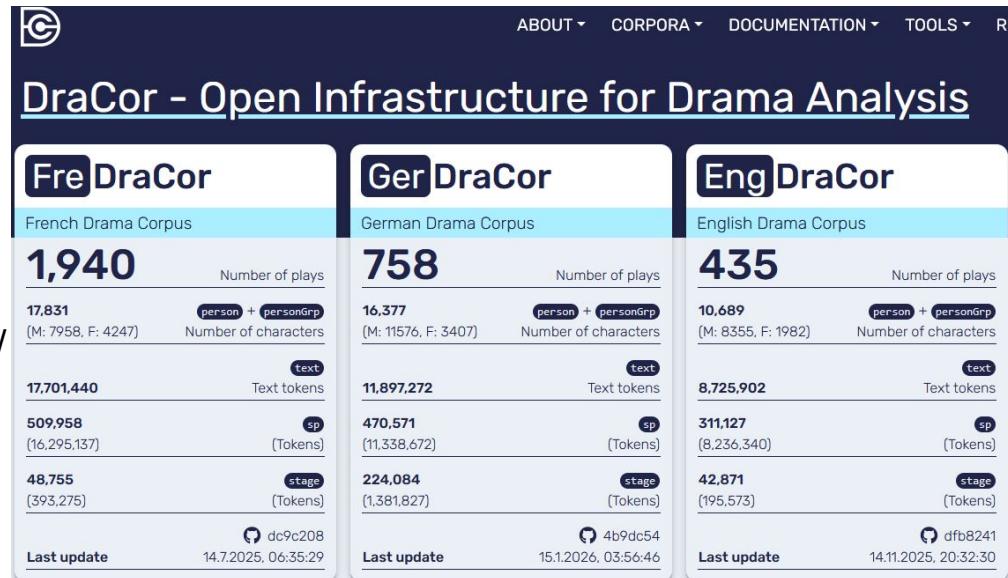
- Bibliografische Angaben
 - Titel
 - Autor*in
 - Textform: Bühnenstück/Roman...
 - Genre
 - Sprache
 - Publikationsdatum
 - Publikationsort
 - (Werk/Expression/Manifestation/
Exemplar)

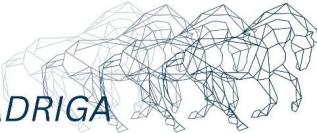
Dramatische Metadaten:

Anzahl der Figuren

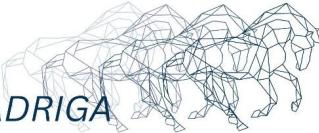
Anzahl der Akte / Tokens

Co-Occurrence Netzwerkdaten





- Bibliographische Angaben
 - Werke, Autor*innen, Affiliation, Institutionen, Themen, Herausgeber, Förderer, Zugänglichkeit und Lizenzen...
- Forschung zu fachspezifischen Publikationskulturen
 - Wie werden Werke veröffentlicht (als Artikel, Monografie, online)?
 - Wer veröffentlicht (z. B. Geschlechterverteilung, Institutionen, Anzahl der Autor*innen)?
 - Wie wird Forschung kommuniziert (z.B. Blogs, Social Media, usw.)?
- Forschung zu Erschließung
 - Wie wurden Werke erschlossen und bereitgestellt (z. B. Biases bei den Erschließungswerkzeugen wie die GND)?



- Literaturdatenbanken und andere Verzeichnisse
- Repositories
- Social Media, Websites und Blogs
- ...

Web of ScienceTM

Scopus[®]

 **OpenAlex**

UNIVERSITÄTSBIBLIOTHEK

zenodb

FID Media

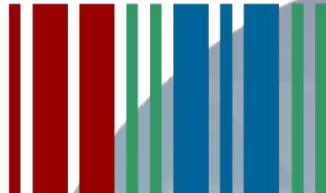
ORCID

Connecting research and researchers

 **DOAJ** OPEN GLOBAL TRUSTED



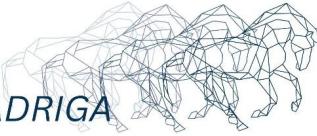
GND


WIKIDATA



Handout

https://quadriga-dk.github.io/Metadaten_Film_Workshop_2026/



Teilt Euch in Gruppen auf zur Recherche in verschiedenen Datenbanken:

- IMDb (<https://www.imdb.com/de/>)
- omdb (<https://www.omdb.org/de/de>)
- OFDb (<https://www.ofdb.de/>)
- TMDB (<https://www.themoviedb.org/?language=de>)
- filmportal.de (<https://www.filmportal.de/>)
- DEFA-Stiftung (<https://www.defa-stiftung.de/filme/filme-suchen/>)
- Dokumentarfilmgeschichte
(<http://www.db.dokumentarfilmgeschichte.de/search.php>)



Rechercheaufgabe

Recherchiert in der gewählten Datenbank nach der Person “Konrad Wolf” und dessen Filme. Achtet dabei insbesondere auf folgende Aspekte:

- Wie sind die Seiten der Datenbank aufgebaut?
- Nach was kann wie recherchiert werden?
- Welche Metadaten sind abrufbar?
- Wie verlässlich erscheinen die abgerufenen Informationen?
- Wie kann mit den Metadaten weiter gearbeitet werden?
- Sind Informationen zu den rechtlichen Aspekten der Metadaten vorhanden?

Bearbeitungszeit: ca. 15 Minuten



Diskussion der Ergebnisse

Gemeinsame Diskussion zu den Recherchen

- Wie unterscheiden sich die einzelnen Datenbanken und Rechercheergebnisse?
- Welche Datenbanken sind für welche Zwecke besonders geeignet?



Mittagspause



Datenbereinigung

- Warum Bereinigen?
- Fehler-Typen
- Heuristiken
- Dokumentation
- Kleine Auswahl von Tools



Warum Bereinigen?

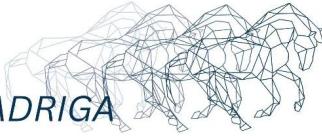
- falsch, Fehler
- verzerrt/verfälscht Auswertungen



Fehler-Typen

- Fehlende Werte (Missings)
- Falsche Werte, Fehler
- Ausreißer
- Vereinheitlichung

Muss-Fehler vs. Kann-Fehler



Datenbereinigung

Beispiele

- “[1985]”
- “1985-1986”
- “1885”

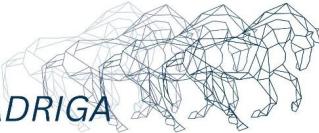


QUADRIGA

Datenbereinigung

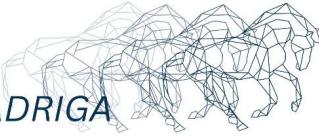
Heuristiken

- Niemals einen fehlenden Wert durch 0 ersetzen.
- Mehr als X (bspw. 2 oder 3) Standardabweichungen vom Mittelwert entfernt



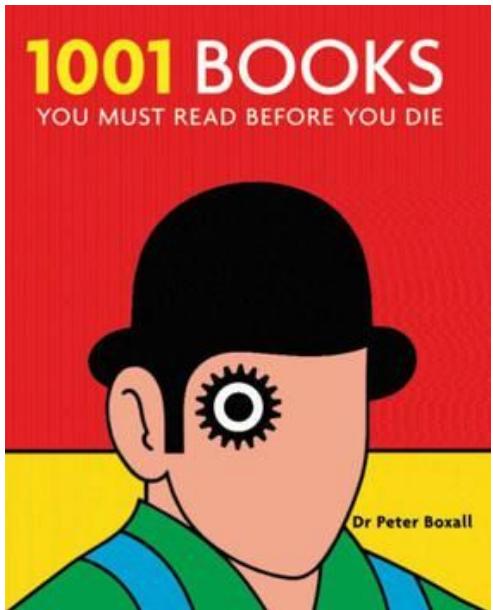
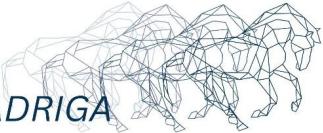
Dokumentation

- Bereinigung ohne Dokumentation ist grundsätzlich falsch / unwissenschaftlich.
- Dokumentiert werden muss:
 - Was wurde manipuliert (gelöscht, interpoliert, ...)?
 - Warum wurde diese Entscheidung getroffen?



Kleine Auswahl von Tools

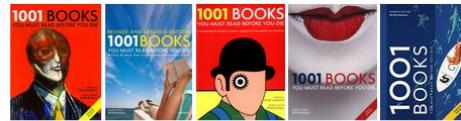
- <https://openrefine.org>
- <https://www.tableau.com/products/prep>
- <https://www.talend.com/products/data-preparation/>
- Programmiersprachen haben typischerweise Programmbibliotheken für die Bereinigung
 - Python
 - R
 - Julia
 - ...



Beispielprojekt: 1001 Books

goodreads Home My Books Browse Community Search books

Listopia



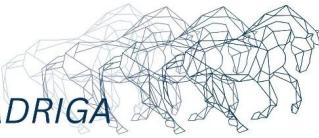
The books on Boxall's list, which is found in the 5 editions of the published book with a **TOTAL NUMBER OF 1318** books.

These books are mostly **NOVELS**. That is why there are no holy books, Shakespeare, etc. **THIS LIST IS COMPLETE. DO NOT ADD ANY BOOKS AND ALSO DO NOT REMOVE ANY.** In case of doubt post a comment here and the people maintaining this list will take a look at it! The list can be found at <https://docs.google.com/spreadsheet/c...>

flag

All Votes Add Books To This List

1		To Kill a Mockingbird by Harper Lee 4.26 avg rating – 6,886,100 ratings score: 155,720, and 1,590 people voted	Want to Read Rate this book
2		Pride and Prejudice by Jane Austen 4.30 avg rating – 4,803,282 ratings score: 126,592, and 1,302 people voted	Want to Read Rate this book
3		1984 by George Orwell 4.20 avg rating – 5,461,826 ratings score: 125,791, and 1,298 people voted	Want to Read Rate this book
4		The Lord of the Rings by J.R.R. Tolkien 4.54 avg rating – 730,839 ratings score: 92,710, and 958 people voted	Want to Read Rate this book
5		The Great Gatsby by F. Scott Fitzgerald 3.93 avg rating – 5,917,002 ratings score: 87,311, and 921 people voted	Want to Read Rate this book
6		Jane Eyre by Charlotte Brontë 4.16 avg rating – 2,323,658 ratings score: 84,224, and 885 people voted	Want to Read Rate this book



Beispielprojekt: 1001 Books

OpenRefine 1001 books plus relevance tsv [Permalink](#)

Facet / Filter Undo / Redo 10 / 10

Refresh Reset all Remove all

sex or gender change
4 choices Sort by: name count Cluster
female 282
male 1029
non-binary 1
trans man 1
(blank) 5
Facet by choice counts

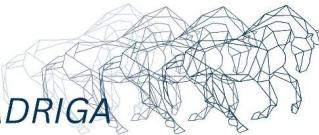
Author change
769 choices Sort by: name count Cluster
Coetzee, J.M. 10
Dickens, Charles 10
Woolf, Virginia 9
Beckett, Samuel 8
DeLillo, Don 8
Greene, Graham 8
McEwan, Ian 8
Roth, Philip 8
Auster, Paul 7
Ballard, J.G. 7
Bellow, Saul 7

1,318 rows

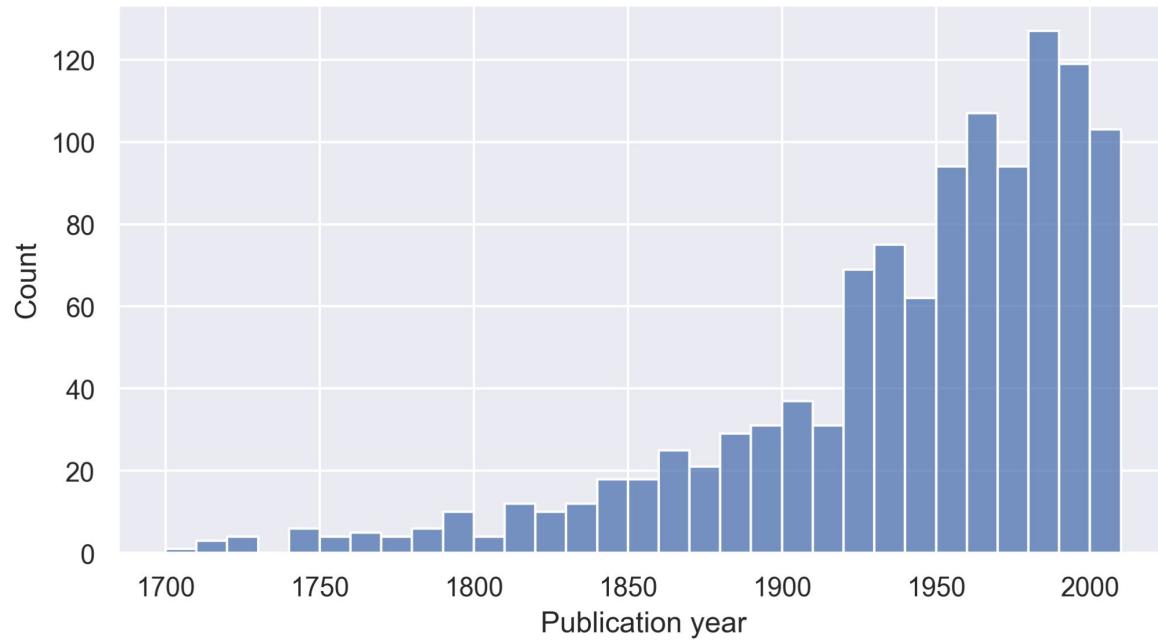
Show as: rows records Show: 5 10 25 50 100 500 1000 rows

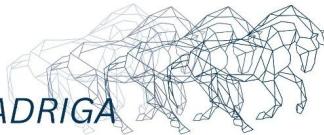
#	Book Title	Author	List	Origin	Wils.	natic	Period	List	Work Wikid	Author Wiki	sex or gend	Work	Authc	Work	Auth		
1.	Aesop's Fables	Aesopus	2) Deleted 2008		174	Greek	pre-1700s		Q865902 Choose new match	Facet	male	871410	1246432	52	100		
2.	Metamorphoses	Ovid	2) Deleted 2008		368	Roman	pre-1700s		Q184742 Choose new match	Text filter	new match	868100	1059672	61	134		
3.	Chaireas and Kallirhoe	Chariton	2) Deleted 2008		1136		pre-1700s		Q2972092 Choose new match	Edit cells	new match						
4.	An Ethiopian Romance	Heliodorus	2) Deleted 2008	Aithiopika	1164	Syrian	pre-1700s		Q2657860 Choose new match	Edit column	Split into several columns...						
5.	The Thousand and One Nights	Anonymous	1) core list		525	Arabian	pre-1700s		Q5258 Choose new match	Transpose	Join columns...						
6.	The Tale of the Bamboo Cutter	Anonymous	3) Added 2008		873		pre-1700s		Q923593 Choose new match	Sort...	Add column based on this column...						
7.	The Tale of Genji	Murasaki, Shikibu	3) Added 2008		1082		pre-1700s		Q5269 Choose new match	View	Add column by fetching URLs...						
8.	Romance of the Three Kingdoms	Luo, Guanzhong	3) Added 2008		1158		pre-1700s		Q70806 Choose new match	Reconcile	Add columns from reconciled values...						
9.	The Water Margin	Shi, Nai'an	3) Added 2008		414		pre-1700s		Q70827 Choose new match	Choose new match	male	Q264517 Choose new match					
10.	Tirant lo Blanc	Marforrel, Joanot	3) Added 2008		1280		pre-1700s		Q1777502 Choose new match	Choose new match	male	Q1777502 Choose new match					
11.	La Celestina	Rojas, Fernando de	3) Added 2008		1154		pre-1700s		Q559667 Choose new match	Choose new match	male	Q363836 Choose new match					
12.	Amadis of Gaul	Montalvo, Garcí Rodríguez de	3) Added 2008		1195		pre-1700s		Q538705 Choose new match	Choose new match	male	Q538705 Choose new match					
13.	The Life of Lazarillo de Tormes	Anonymous	3) Added 2008	La vida del Lazarillo de Tormes	1001		pre-1700s		Q425075 Choose new match	Choose new match	male	Q23468 Choose new match					
14.	Gargantua and Pantagruel	Rabelais, François	1) core list		1055	French	pre-1700s		Q822290 Choose new match	Choose new match	male	545871 Choose new match	657493	39	90		
15.	The Lusiads	Camões, Luís de	3) Added 2008		293		pre-1700s		Q781696 Choose new match	Choose new match	male	Q590 Choose new match					
16.	Monkey: Journey to the West	Wu, Cheng'en	3) Added 2008		294		pre-1700s		Q70784 Choose new match	Choose new match	male	Q228689 Choose new match					
17.	Euphues: The Anatomy of Wit	Lyly, John	2) Deleted 2008		1297	English	pre-1700s		Q5410723 Choose new match	Choose new match	male	1758091 Choose new match	130947	66	50		
18.	The Unfortunate Traveller	Nashe, Thomas	1) core list		1200	English	pre-1700s		Q261627 Choose new match	Choose new match	male	21599 Choose new match	47647	3	25		
19.	Thomas of Reading	Deloney, Thomas	3) Added 2008		1307		pre-1700s		Q443042 Choose new match	Choose new match	male	Q443042 Choose new match	13649	39830	2	24	
20.	The Golden Ass	Apuleius, Lucius	1) core list		465	Roman	pre-1700s		Q127500887 Choose new match	Choose new match	male	Q1530410 Choose new match		3466	0	7	
21.	Don Quixote	Cervantes Saavedra, Miguel de	1) core list		840	Spanish	pre-1700s		Q1044767 Choose new match	Choose new match	male	Q170512 Choose new match		302843	259405	31	71
22.	The Travels of Persiles and Sigismunda	Cervantes Saavedra, Miguel de	3) Added 2008	Los trabajos de Persiles y	1123	Spanish	pre-1700s		Q2634153 Choose new match	Choose new match	male	Q5682 Choose new match		4291046	2272986	123	176

javascript:()



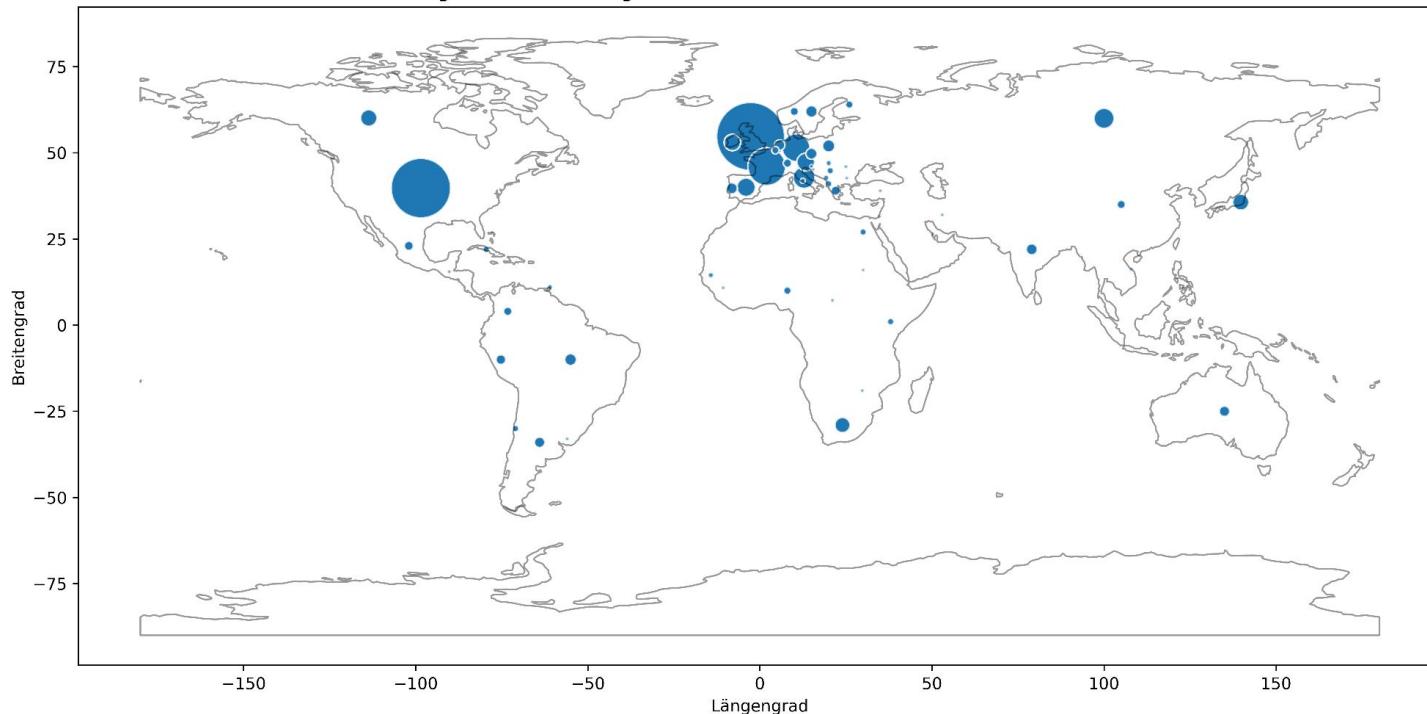
Beispielprojekt: 1001 Books





Beispielprojekt: 1001 Books

Geografische Verteilung der Werke nach Geodaten aus GND und Wikidata





Beispielprojekt correspSearch

correspSearch

Start Suche Daten Über correspSearch Mitmachen API quoteSalute Deutsch

Briefpartner:in

Briefpartner:in

Erwähnte Person (beta)

Orte

Beruf

Datum

Volltextsuche (beta)

Edition / Publikation

CMIF-Verzeichnis

Datenset

Amsterdam

Erweiterte Suche

1 2 3 4 ... > >>

Johann Georg Ludwig Möller → August Wilhelm von Schlegel

In: Digitale Edition der Korrespondenz August Wilhelm Schlegels
<https://august-wilhelm-schlegel.de/briefedigital/>

Zum Brief

25. Mai 1793

Berlin

Wilhelm von Humboldt → August Wilhelm von Schlegel

In: Digitale Edition der Korrespondenz August Wilhelm Schlegels
<https://august-wilhelm-schlegel.de/briefedigital/>

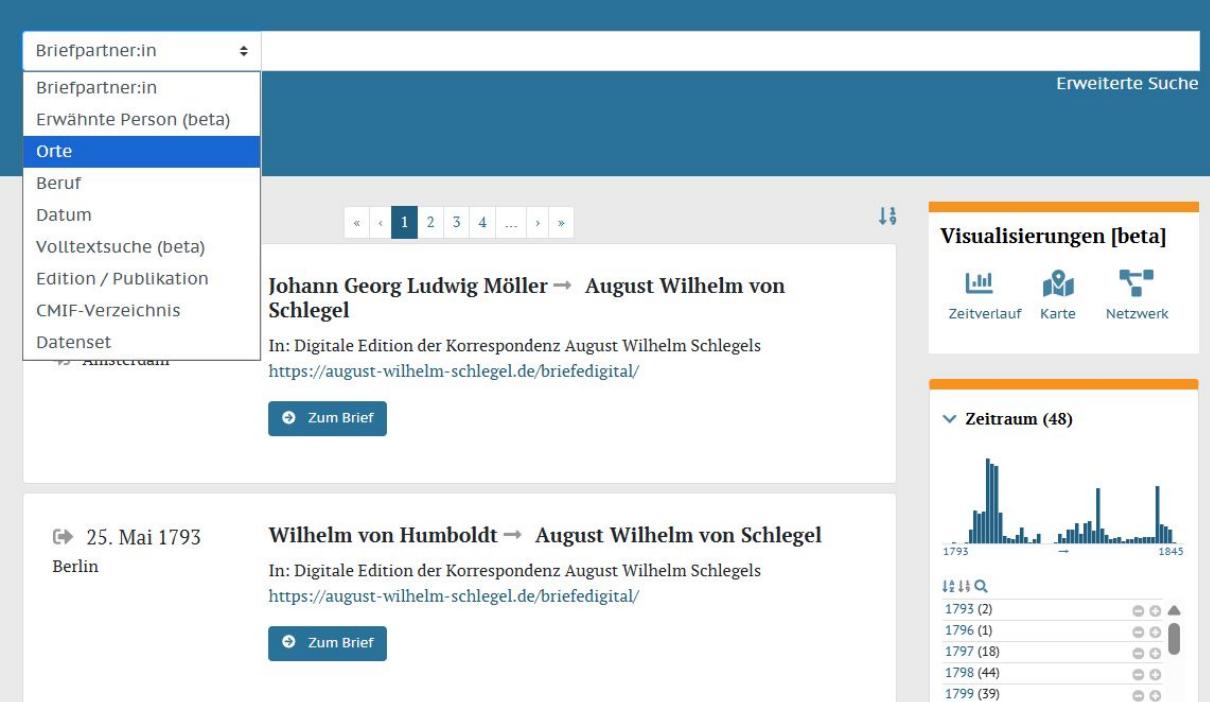
Zum Brief

Visualisierungen [beta]

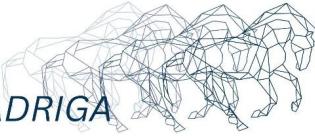
Zeitverlauf Karte Netzwerk

Zeitraum (48)

1793 (2)
1796 (1)
1797 (18)
1798 (44)
1799 (39)



<https://correspsearch.net/de/suche.html?p=http://d-nb.info/gnd/2004272-3&x=1&w=0>

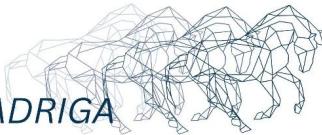


QUADRIGA

Beispielprojekt correspSearch

Datei Bearbeiten Format Ansicht Hilfe

- 1 11.10.1898 Postkarte von P. Klee (Bern) an H. Bloesch (Biel)
- 2 18.10.1898 Brief von P. Klee (München) an H. Bloesch (Bern)
- 3 30. - 31.10.1898 Brief von P. Klee (München) an H. Bloesch (Bern)
- 4 15. - 28.11.1898 Brief von P. Klee (München) an H. Bloesch (Bern)
- 5 28.12.1898 Brief von P. Klee (München) an H. Bloesch (Bern)
- 6 17.1. - 16.2.1899 Brief von P. Klee (München) an H. Bloesch (Bern), 1. Teil
- 7 31.1.1899 Postkarte von P. Klee (München) an H. Bloesch (Bern)
- 8 17.1. - 16.2.1899 Brief von P. Klee (München) an H. Bloesch (Bern), 2. Teil
- 9 1.3.1899 Brief von P. Klee (München) an H. Bloesch (Bern)
- 10 11.3.1899 Postkarte von P. Klee (München) an H. Bloesch (Bern)
- 11 10.6.1899 Postkarte von P. Klee (München) an H. Bloesch (Bern)
- 12 15.6.1899 Brief von P. Klee (München) an H. Bloesch (Bern)
- 13 8.7.1899 Postkarte von P. Klee (Burghausen) an H. Bloesch (Bern)
- 14 29.10.1899 Postkarte von P. Klee (München) an H. Bloesch (Bern)
- 15 2.12.1899 Postkarte von P. Klee (München) an H. Bloesch (Bern)
- 16 21.12.1899 Ansichtskarte von P. Klee u. a. (München) an H. Bloesch(Bern)
- 17 26.12.1899 Postkarte von P. Klee (München) an H. Bloesch (Bern)
- 18 20.1. - 1.2.1900 Brief von P. Klee (München) an H. Bloesch (Bern)
- 19 15.3.1900 Brief von P. Klee (München) an H. Bloesch (Bern)
- 20 13.4.1900 Ansichtskarte von P. Klee u. a. (Seeshaupt) an H. Bloesch(Bern)
- 21 Ende Mai 1900 Brief von P. Klee (Oberhofen) an H. Bloesch (Bern)
- 22 13.11.1900 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 23 15.12.1900 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 24 26.12.1900 Brief von P. Klee (München) an H. Bloesch (Paris)
- 25 10.1.1901 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 26 8.2.1901 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 27 19.2.1901 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 28 2.3.1901 Ansichtskarte von P. Klee u. a. (Starnberg) an H. Bloesch(Paris)
- 29 3.4.1901 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 30 28.4.1901 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 31 31.5.1901 Postkarte von P. Klee (München) an H. Bloesch (Paris)
- 32 28.9.1901 Briefkarte von P. Klee (Oberhofen) an H. Bloesch (Basel)
- 33 29.10.1901 Postkarte von P. Klee und Hermann Haller (Rom) an H. Bloesch (Bern)
- 34 26.11.1901 Ansichtskarte von P. Klee (Rom) an H. Bloesch (Bern)
- 35 9.12.1901 Ansichtskarte von P. Klee (Rom) an H. Bloesch (Bern)
- 36 14.12.1901 Ansichtskarte von P. Klee (Rom) an H. Bloesch (Bern)
- 37 23.12.1901 Brieffragment von P. Klee (Rom) an H. Bloesch (Bern)
- 38 21.1.1902 Ansichtskarte von P. Klee (Rom) an H. Bloesch (Bern)



QUADRIGA

Beispielprojekt correspSearch

Facet / Filter Undo / Redo 270 / 270

Refresh Reset All Remove All

senderPlace change
34 choices Sort by: name count Cluster

Aschau b. Prien 1
Ascona 1
Beatenberg 1
Berlin 1
Bern 58
Bern? 1
Bidart 1
Braunschweig 3
Burghausen 1
Faug 1
Florenz 1

key change reset
0 — 190

senderDate change reset
NaN-NaN-NaN NaN:NaN:NaN — 1970-01-01 01:00:00
 Time Non-Time Blank Error
0 187 0 0

187 rows

Show as: rows records Show: 5 10 25 50 100 500 1000 rows

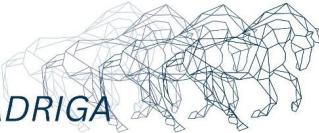
		key	senderDateText	senderDate	sender	senderID	senderPlace	senderPlaceID	addressee
1.	1	11.10.1898	1898-10-11	Paul Klee	http://d-nb.info/gnd/118562827				Hans Bloesch
2.	2	18.10.1898	1898-10-18	Paul Klee	http://d-nb.info/gnd/118562827				Hans Bloesch
3.	3	30. – 31.10.1898	[1898-10-30,1898-10-31]	Paul Klee	http://d-nb.info/gnd/118562827				Hans Bloesch
4.	4	15. – 28.11.1898	[1898-11-15..1898-11-28]	Paul Klee	http://d-nb.info/gnd/118562827				Hans Bloesch
5.	5	28.12.1898	1898-12-28	Paul Klee	http://d-nb.info/gnd/118562827				Hans Bloesch
6.	6	17.1. – 16.2.1899	[1899-01-17..1899-02-16]	Paul Klee	http://d-nb.info/gnd/118562827				Hans Bloesch
7.	7	31.1.1899	1899-01-31	Paul Klee	http://d-nb.info/gnd/118562827				Hans Bloesch
8.	8	17.1. – 16.2.1899	[1899-01-17..1899-02-16]	Paul Klee	http://d-nb.info/gnd/118562827	München	http://www.geonames.org/2867714	Hans Bloesch	
9.	9	1.3.1899	1899-03-01	Paul Klee	http://d-nb.info/gnd/118562827	München	http://www.geonames.org/2867714	Hans Bloesch	
10.	10	11.3.1899	1899-03-11	Paul Klee	http://d-nb.info/gnd/118562827	München	http://www.geonames.org/2867714	Hans Bloesch	

Facet Text facet
Text filter Numeric facet
Edit cells Timeline facet
Edit column Scatterplot facet
Transpose Custom text facet...
Sort... Custom Numeric Facet...
View Customized facets
Reconcile www.geonames.org/2867714
<http://www.geonames.org/2867714>



Datenbereinigung

Eure Erfahrung



OpenRefine

- Freies OpenSource Tool
- Einfach zu erlernen
- Großer Funktionsumfang
- Komplexe Bereinigungsschritte möglich



OpenRefine

Datenbereinigung mit OpenRefine

Main features		
 Faceting Drill through large datasets using facets and apply operations on filtered views of your dataset.	 Clustering Fix inconsistencies by merging similar values thanks to powerful heuristics.	 Reconciliation Match your dataset to external databases via reconciliation services.
 Infinite undo/redo Rewind to any previous state of your dataset and replay your operation history on a new version of it.	 Privacy Your data is cleaned on your machine, not in some dubious data laundering cloud.	 Wikibase Contribute to Wikidata, the free knowledge base anyone can edit, and other Wikibase instances.

And much more to discover in our documentation.



<https://quadriga-dk.github.io/Bewegtes-Bild-Fallstudie-2/intro.html>

The screenshot displays a digital humanities project titled "Studentische Filme an der Filmuniversität Babelsberg zur Wendezeit (1985-1999)". The interface includes a sidebar with navigation links for the study, a main content area with a word cloud, and a right sidebar for feedback and citation information.

Content Area:

Studentische Filme an der Filmuniversität Babelsberg zur Wendezeit (1985-1999)

A large word cloud centered around the word "Wende". Other prominent words include "Mauerfall", "Mauer", "Wahlen", "DDR", "Arbeiter", "Grenze", "Revolution", "Politik", "Jugend", "Kunst", "Spielfilm", "Zukunft", "Adaption", "Bahnhof", "Animationsfilm", and "Grenze".

Right Sidebar:

Fragen oder Feedback

Stellen Sie eine Frage ↗
Geben Sie uns Feedback ↗
Mit Ihren Rückmeldungen können wir unser interaktives Lehrbuch gezielt an Ihre Bedürfnisse anpassen.

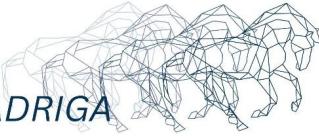
Zitierhinweis

```
@book{Schick_Studentische_2022,
  title = {Studentische Filme an der Filmuniversität Babelsberg zur Wendezeit (1985-1999)},
  author = {Schick, Thomas},
  year = {2022},
  version = {1.0.0-beta},
  doi = {10.5281/zenod.1655281},
  url = {https://quadriga-dk.github.io/Bewegtes-Bild-Fallstudie-2/intro.html},
  copyright = {Open Education},
  language = {de},
  keywords = {Open Education}
}
```

Schick, T., Winter, M., Loist, S. & Giesecke, L. (2025). Studentische Filme an der Filmuniversität Babelsberg zur Wendezeit (1985-1999).

Bottom Note:

Unterstützung steht der Elternhürde des Hochschul für Film und Fernsehen (HFF) "Konrad Wolf" über das



Was ist Hypothes.is?



- Open-Source-Tool zur **Web-Annotation**
- **Kommentieren und Markieren** von Webseiten & PDFs
- **Kollaboratives Arbeiten** mit Kommentarfunktion
- **Privat oder in Gruppen** nutzbar

The screenshot shows a comparison between the Hypothes.is interface and a text block from OpenRefine. A red arrow points from the text block to the interface.

Annotation direkt im Text

OpenRefine ist eine OpenSource Software für die
Bereinigung, Formatierung und Anreicherung von Datensätzen.

Annotations Page Notes

tn_1 Nützliches Tool für Datennalyse. Hat jemand Erfahrungen mit OpenRefine?
1 Reply

tn_2 Ja, habe ich schon genutzt. Besonders effektiv für das Bereinigen von unordentlichen Daten.
1d

tn_3 Einfach zu lernen und leistungsstark. Guter erster Schritt, um Daten aufzuräumen.
1d

tn_4 OpenRefine kann auch zur Datenanreicherung durch Verknüpfung mit externen Quellen genutzt werden.
1d

Post to: Gruppe



Technik-Check ⏱ 10-20 Minuten

1. **Website** via Link / Code aufrufen
2. **Hypothes.is Login** eingeben
3. **Test-Annotation** vornehmen

Vollständige Anonymität: Die Erhebung erfolgt **komplett anonym**. Die Daten dienen nur der **Qualitätssicherung und Weiterentwicklung** der QUADRIGA OER-Materialien.





Datenbereinigung mit Open Refine

Aufgabe:

- Kapitel 4.3 der Fallstudie bearbeiten
- mit Hypothes.is ggf. kommentieren bei Unklarheiten



Bearbeitet das Kapitel 4.3.2 “Vorbereitung und Import des Datensatzes”

(ohne “Die Importeinstellungen im Überblick”)

Bearbeitungszeit: ca. 10 Minuten



Bearbeitet das Kapitel 4.3.3 “Sichtung des Datensatzes”

Bearbeitungszeit: ca. 20 Minuten



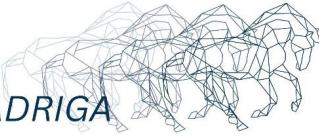
Bearbeitet das Kapitel 4.3.4 “Bereinigung der Spalten”

Bearbeitungszeit: ca. 20 Minuten



Pause





Bearbeitet das Kapitel 4.3.5 “Bereinigung der Einträge - Standardisierung der Jahresangaben”

Bearbeitungszeit: ca. 40 Minuten



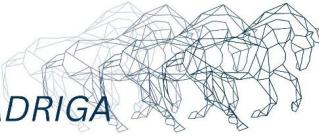
Optional: Bearbeitet das Kapitel 4.3.5 “Bereinigung der Einträge - Standardisierung der Titel”

Bearbeitungszeit: ca. 50 Minuten



Bearbeitet das Kapitel 4.3.6 “Export des Datensatzes”

Bearbeitungszeit: ca. 5 Minuten

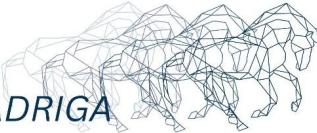


Feedback

Was nehmt ihr aus dem **Kapitel** mit?

Wie hat euch das **Annotieren** gefallen?





QUADRIGA

Wir möchten dein Feedback!

Hilf uns, unseren Workshop noch besser zu machen
– nimm dir kurz Zeit für die Umfrage!



(Meta-)Daten bei Film, Video und bewegtem Bild – Recherche, Bereinigung, Visualisierung mit OpenRefine und Excel

Poststand Fx. 23.1.2028

Ihr Feedback zum heutigen Tag

1. In welchen Disziplinen arbeiten Sie?

2. In welcher Phase Ihres Studiums / ihrer wissenschaftlichen Qualifikation befinden Sie sich derzeit?
 Masterstudium
 Promotions
 Postdocs
 Sonstiges, nämlich:

3. Was war Ihre Hauptmotivation am Workshop teilzunehmen?

4. Wie haben Sie den heutigen Tag erlebt?
 0 Ich habe voll gelitten
 1 Ich habe überhaupt nicht zu :2: 3: 4: 5 (stimme voll und ganz zu)
 2 Ich habe mich beim Lernen in dieser Workshop-Umgebung wohlgeföhlt
 3 Ich stimme überhaupt nicht zu :2: 3: 4: 5 (stimme voll und ganz zu)
 4 Ich kann das in diesem Workshop Gelernte unmittelbar anwenden
 5 Ich stimme überhaupt nicht zu :2: 3: 4: 5 (stimme voll und ganz zu)

5. Welche Disziplinen haben Ihnen besonders gut gefallen?
 ...
 ...
 ...
 ...
 ...

6. Welche Disziplinen interessieren Sie in Zukunft?
 ...
 ...
 ...
 ...
 ...

Viel Dank für Ihre Teilnahme!
Bitte die Karten in die Box am Ausgang einwerfen.
Herzlichen Dank!



Ablauf Tag 2

10.00 Uhr	Visualisierung mit Mark-Jan Bludau
	Pause
12.30 Uhr	Visualisierung mit Excel (OER) oder OpenRefine aus 4.3 weitermachen
13.30 Uhr	Abschluss
14.00 Uhr	Ende
14.30 Uhr	(Weiter-)Entwicklung von datengestützter eigener Forschung
16.00 Uhr	Ende



Updates & Feedback



Möchten Sie von uns zu Neuigkeiten kontaktiert oder zu Feedback und Erprobungen eingeladen werden?

<https://gesellschaftfuerinformatik.limesurvey.net/254548?lang=de>



Technik-Check ⏱ 10-20 Minuten

1. **Website** via Link / Code aufrufen
2. **Hypothes.is Login** eingeben
3. **Test-Annotation** vornehmen

Vollständige Anonymität: Die Erhebung erfolgt **komplett anonym**. Die Daten dienen nur der **Qualitätssicherung und Weiterentwicklung** der QUADRIGA OER-Materialien.





Datenbereinigung mit Open Refine

Aufgabe:

- Kapitel 5.2 der Fallstudie bearbeiten
- mit Hypothes.is ggf. kommentieren bei Unklarheiten



Übersicht Kapitel 5.2 “Visualisierungen zur Fallstudie”



Bearbeitet das Kapitel 5.2.1 “Import des Datensatzes in Excel”

(Windows in der Webansicht - MacOS Besonderheiten beachten)

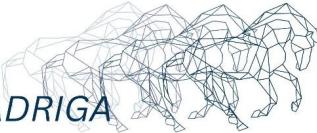
Bearbeitungszeit: ca. 15 Minuten



Bearbeitet das Kapitel 5.2.2 “Titel pro Jahr”

(die erste Pivot-Tabelle/-Grafik)

Bearbeitungszeit: ca. 20 Minuten



QUADRIGA

Visualisierung

Bearbeitet das Kapitel 5.2.3 “Anzahl der Gattungen pro Jahr”

(Wiederholung Pivot-Tabelle/-Grafik)

Bearbeitungszeit: ca. 15 Minuten



Visualisierung

Optional: Bearbeitet das Kapitel 5.2.4 “Projektarten pro Jahr”

(Übung möglichst ohne zu spicken)

Bearbeitungszeit: ca. 15 Minuten



Visualisierung

Erfahrungsaustausch Kapitel 5.2 “Visualisierungen zur Fallstudie”

Dauer: ca. 5 Minuten



Optional: Bearbeitet das Kapitel 5.2.6 “Anzahl der Schlagworte pro Jahr”

(Für die ganz Schnellen - PowerQuery)

Bearbeitungszeit: falls noch Zeit ist

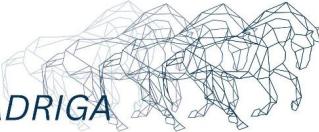


Feedback

Was nehmt ihr aus dem **Kapitel** mit?

Wie hat euch das **Annotieren** gefallen?





Wir möchten dein Feedback!

Hilf uns, unseren Workshop noch besser zu machen
– nimm dir kurz Zeit für die Umfrage!



QUADRIGA
(Meta-)Daten bei Film, Video und bewegtem Bild – Recherche, Bereinigung, Visualisierung mit OpenRefine und Excel
Postausk. Fx. 23.1.2028

Ihr Feedback zum heutigen Tag

1. In welchen Disziplinen arbeiten Sie?
 Materialstudium
 Promotions
 Postdocs
 Sonstiges, nämlich:

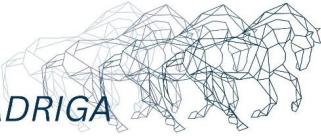
2. In welcher Phase ihres Studiums / ihrer wissenschaftlichen Qualifikation befinden Sie sich derzeit?
 Masterstudium
 Promotions
 Postdocs
 Sonstiges, nämlich:

3. Was war Ihre Hauptmotivation am Workshop teilzunehmen?

4. Wie haben Sie den heutigen Tag erlebt?
 0 Ich habe voll gelitten
 1 Ich habe überhaupt nicht zu :2: 3: 4: 5 (stimme voll und ganz zu)
 2 Ich habe mich beim Lernen in dieser Workshop-Umgebung wohlgeföhlt
 3 Ich stimme überhaupt nicht zu :2: 3: 4: 5 (stimme voll und ganz zu)
 4 Ich kann das in diesem Workshop Gelernte unmittelbar anwenden
 5 Ich stimme überhaupt nicht zu :2: 3: 4: 5 (stimme voll und ganz zu)

5. Wie haben Sie die Kolleg:innen (betr. wahrscheinlich) in Zukunft interessieren?
 Ich habe Ihnen besonders gut
 Ich habe Ihnen sehr gut
 Ich habe Ihnen gut
 Ich habe Ihnen schlecht
 Ich habe Ihnen sehr schlecht

Vielen Dank für Ihre Teilnahme!
Bitte die Karten in die Box am Ausgang einwerfen.
Herzlichen Dank!



Ideen für datenbasierte Forschung