



Brandenburgisches Zentrum
für Medienwissenschaften

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

22.-23. Januar 2026
Potsdam



Brandenburgisches Zentrum
für Medienwissenschaften

Herzlich Willkommen!



Fachhochschule Potsdam
University of
Applied Sciences



FILMUNIVERSITÄT
BABELSBERG
KONRAD WOLF



 **Fraunhofer**
FOKUS



GESELLSCHAFT
FÜR INFORMATIK

 **TECHNISCHE
UNIVERSITÄT
BERLIN**



 **FREIE
UNIVERSITÄT
BERLIN**

Gefördert durch:



Bundesministerium
für Forschung, Technologie
und Raumfahrt



Finanziert von der
Europäischen Union
NextGenerationEU

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)

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?

CC BY-SA 3.0



Wer sind wir?





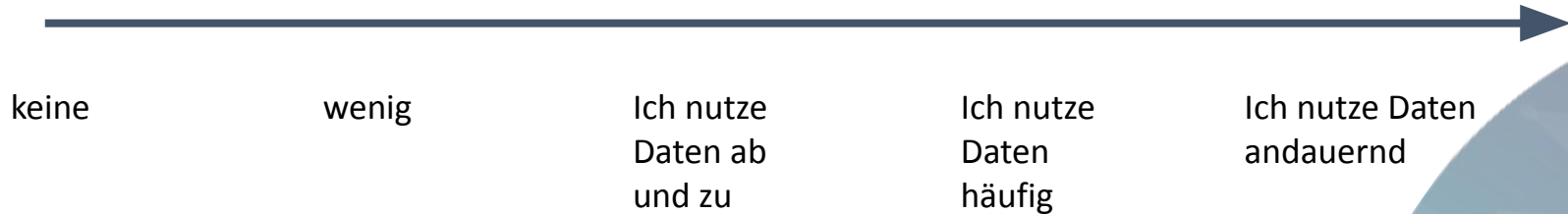
Wer sind wir?

Meine Disziplin ist:



Wer sind wir?

Erfahrung mit Daten (in der Forschung)





(Forschungs-)Daten

Was sind meine Forschungsdaten?

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 Forschungsergebnissen zu unterstützen. Dabei werden in jedem Schritt des datenbasierten Forschungsprozesses (vgl. Datenfluss) notwendige Kompetenzen vermittelt.
Hilfreich sind (Vor-)Kamfrasse in grundlegenden Konzepten von Forschungsdatenmanagement und Open X sowie Wissen um Anwendungsebenen von Daten.



Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Biemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, 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?



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

Korpusbildung

20.sonst.Person	•Hornemann_Klaus [Mischung]
21.sonst.Person	•Otto_Joachim [Szenografie]
22.sonst.Person	•Kraft_Uwe [Prod.-lfg]
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 Filmkopie © MAX FILM 1992 © 2007 Filmgalerie 451, für diese DVD-Ausgabe enthält: "Stilles Land" (1991), "Konsquenzen - Peter, 25" (1987)
Schlagwort	•Spielfilm
	•Wende/A
Stichwörter	Wendezeit
Zusammenfassung	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 ungarbanalen Hindernissen; nur die Assistentin Claudia, in die Kai s kluge Marketing-Ratschläge unterbreitet. (Filmportal)

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.
az1ddzzzc||37

Datenträger
||||ba|||||
    
```



Untitled-1 - quick_viz

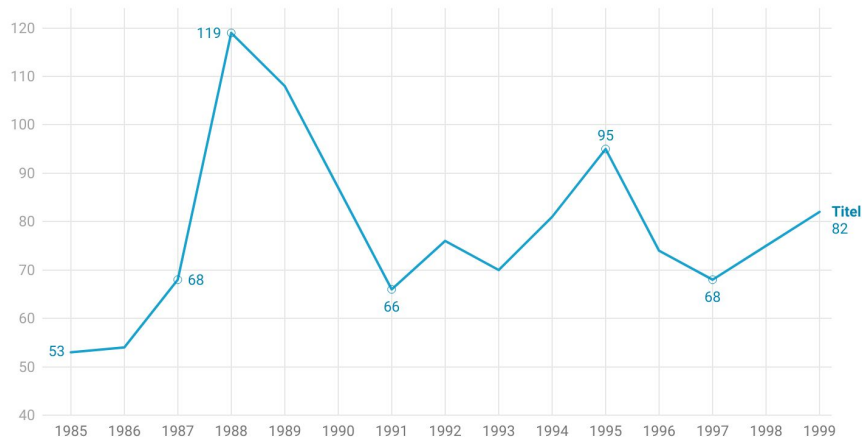
	verantw._Pers._1	verantw._Pers._2	verantw._Pers._3
1			
2	Titel		
3			
4	Nacht in Leipzig, Eine	Olga Mielke	
5			
6	Leipzig 1998		
7			
8	Brief blieb ungeöffnet, Der		
9			
10	Liste war unvollständig, Die	Danny Weber	
11			
12	Niemand nahm den Auftrag mit	Waltraud Schulz	
13			
14	Als der Winter kein Ende nahm	Hoang Minh Tuan	
15			
16	Ruhe vor dem Wort, Die	Xenia Abendroth	
17			
18	Atlas der verlorenen Städte, Der		
19			
20	Zwischen Frühschicht und Flucht	Benjamin Christ	
21			
22	Als der Lehrer streikte	Frieda Eggert	
23			
24	Trennlinie	Volker Uhl	
25			
26	Abmeldung war nicht nötig	Benjamin Christ	
27			
28	Im Schatten der Falte	Simone Schwarz	
29			
30	Heft 6 Ein Versuch	Merten Oertel	
31			
32	Melancholie der leeren Schulhöfe, Die	Rosemarie Quade	
33			
34	Treppengeländer lose		
35			
36	Niemand sang die Hymne		

Quelle:
Screenshot Filmarchiv Datensatz

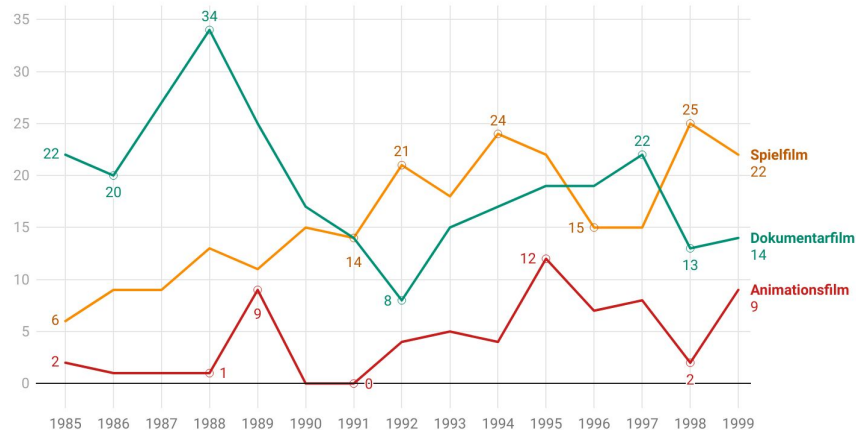
Quelle:
Screenshot bereinigter Filmarchiv Datensatz

Auswertung von Metadaten

Titel pro Jahr 1985-1999



Gattungen pro Jahr 1985-1999





Pause



Metadaten

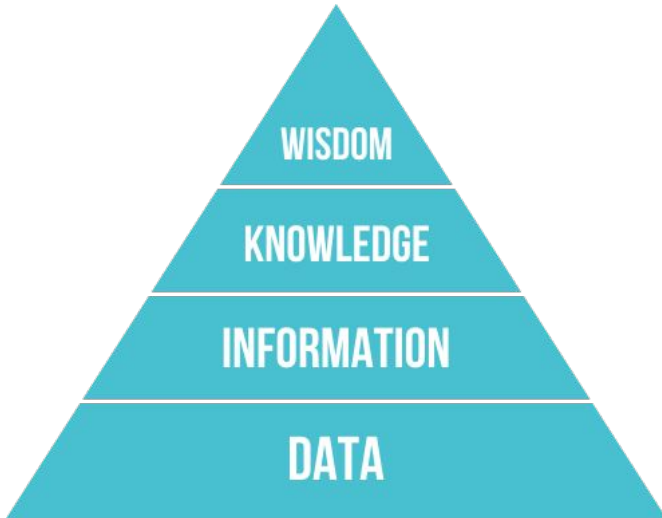
- Was sind (Meta-)Daten?
- (Typische) Eigenschaften von Metadaten



Metadaten

Was sind (Meta-)Daten? – Daten

Was sind (Meta-)Daten? – Daten



INTANGIBLE		TANGIBLE	
ENTITY 2.	Information-as-knowledge Knowledge	3.	Information-as-thing Data, document
PROCESS 1.	Information-as-process Becoming informed	4.	Information processing Data processing

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 \(DOI\)](#)

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)

Was sind (Meta-)Daten? – Metadaten

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



Eigenschaften von Metadaten

- Eher knapp
- Stark strukturiert
- Vorgegebene Vokabulare
- Metadatenschemata

Eigenschaften von Metadaten

- Eher knapp
- Stark strukturiert
- Vorgegebene Vokabulare

- Metadatenschemata

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



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

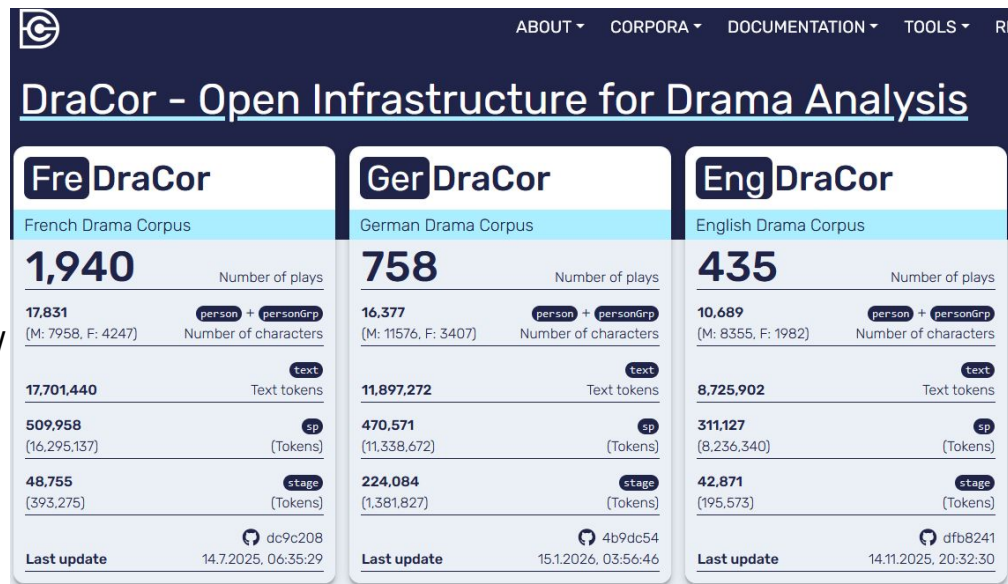
- Bibliografische Angaben
 - Titel
 - Autor*in
 - Textform: Bühnenstück/Roman..
 - Genre
 - Sprache
 - Publikationsdatum
 - Publikationsort
 - (Werk/Expression/Manifestation/Exemplar)

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

zenodo

FID Media

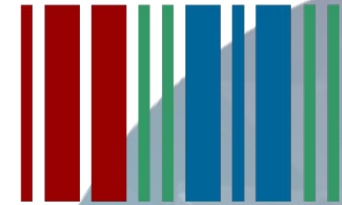
ORCID

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>)

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

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



Beispiele

- “[1985]”
- “1985-1986”
- “1885”

Heuristiken

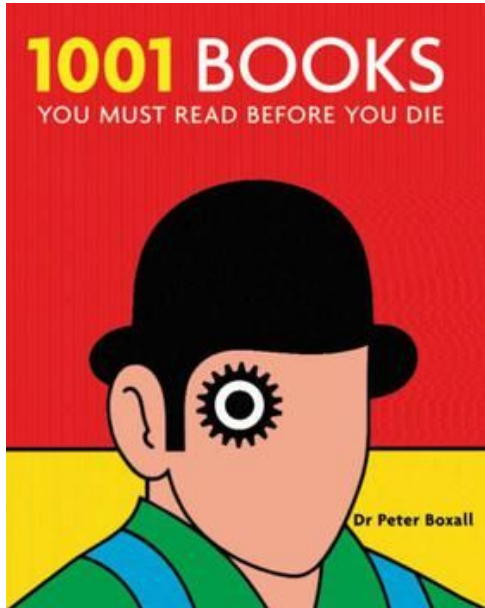
- Niemals einen fehlenden Wert durch 0 ersetzen.
- Mehr als X (bspw. 2 oder 3) Standardabweichungen vom Mittelwert entfernt

Dokumentation

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

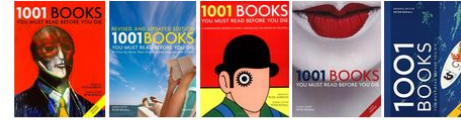
Kleine Auswahl von Tools

- <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

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: 126,596, 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
★★★★★ | |

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

Facet / Filter **Undo / Redo** 10 / 10 **1,318 rows**

Refresh **Reset all** **Remove all** Show as: rows records Show: 5 10 25 50 100 500 1000 rows

sex or gender change **Cluster**

4 choices Sort by: name count

female 282
male 1029
non-binary 1
trans man 1
(blank) 5

Facet by choice counts

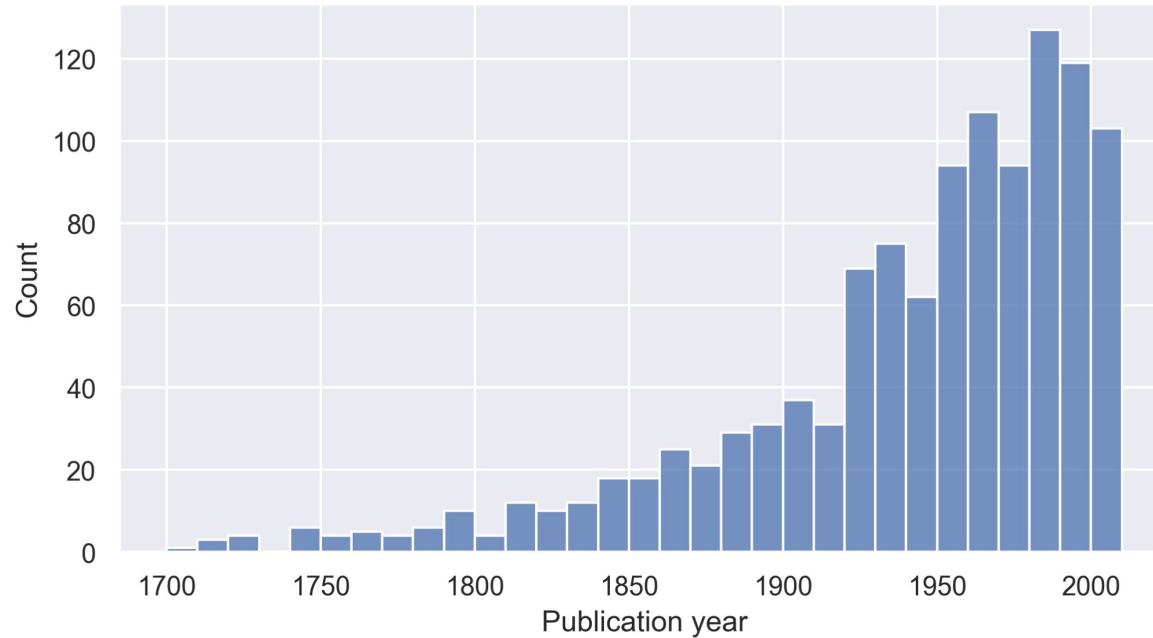
Author change **Cluster**

769 choices Sort by: name count

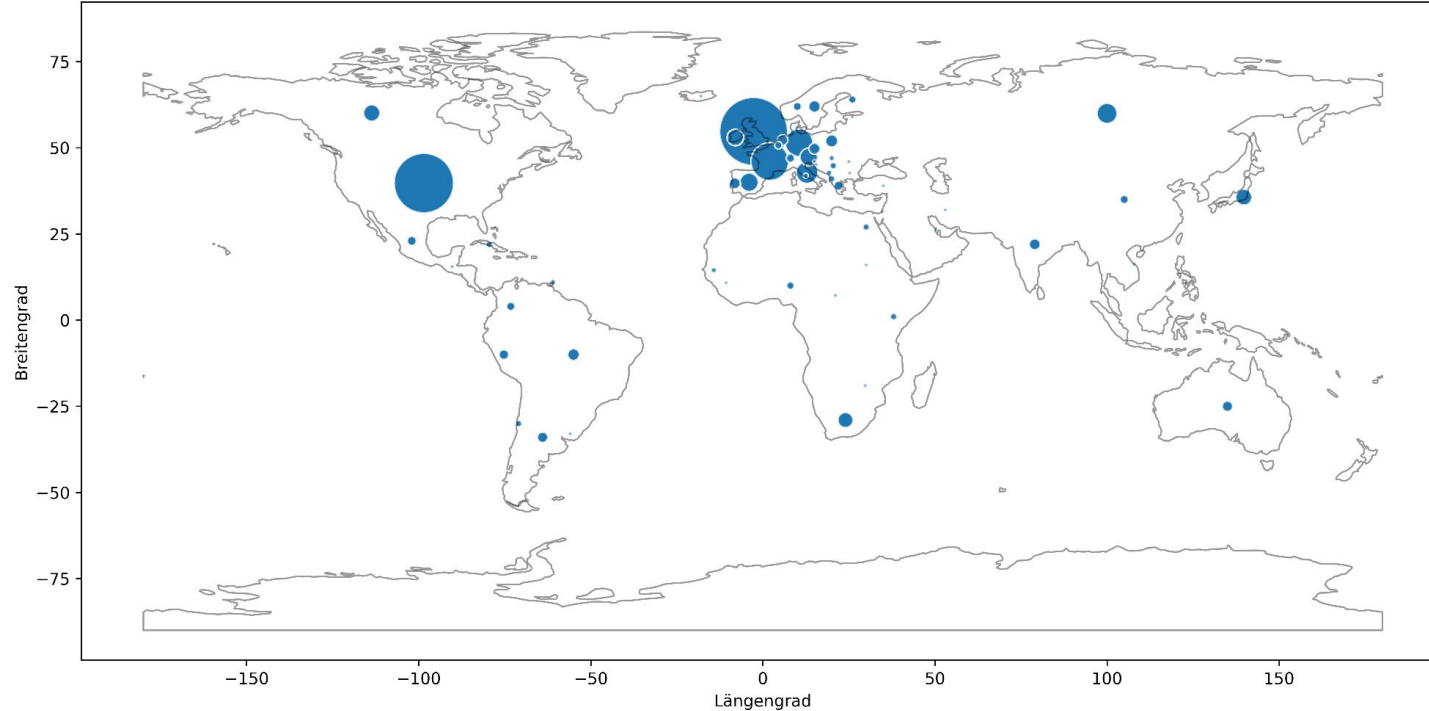
Coetzee, J.M. 10
Dickens, Charles 10
Woolf, Virginia 9
Beckett, Samuel 8
DeLillo, Don 8
Greene, Graham 8
McEwan, Ian 8
Roth, Philip 8
Auster, Paul 7
Ballard, J.G. 7
Bellow, Saul 7

	All	Book Title	Author	List	Origin	Wils	natic	Period	List	Work Wikid	Author Wikid	sex or gend	Work	Auth	Worl	Auth	
1.	Aesop's Fables	Aesopus	2) Deleted 2008		174	Greek	pre-1700s			Q865902 Choose new match	Facet		871410	1246432	52	100	
2.	Metamorphoses	Ovid	2) Deleted 2008		368	Roman	pre-1700s			Q184742 Choose new match	Text filter	few match	868100	1059672	61	134	
3.	Chaireas and Kallirhoe	Chariton	2) Deleted 2008		1136		pre-1700s			Q2972092 Choose new match	Edit cells	few match	Split into several columns...				
4.	An Ethiopian Romance	Heliodorus	2) Deleted 2008	Althiopika	1164	Syrian	pre-1700s			Q2657860 Choose new match	Edit column	few match	Join columns...				
5.	The Thousand and One Nights	Anonymous	1) core list		525	Arabian	pre-1700s			Q8258 Choose new match	Transpose		Add column based on this column...				
6.	The Tale of the Bamboo Cutter	Anonymous	3) Added 2008		873		pre-1700s			Q923593 Choose new match	Sort...		Add column by fetching URLs...				
7.	The Tale of Genji	Murasaki, Shikibu	3) Added 2008		1062		pre-1700s			Q8269 Choose new match	View		Add columns from reconciled values...				
8.	Romance of the Three Kingdoms	Luo, Guanzhong	3) Added 2008		1158		pre-1700s			Q70606 Choose new match	Reconcile		Rename this column...				
9.	The Water Margin	Shi, Nai'an	3) Added 2008		414		pre-1700s			Q70827 Choose new match		Q264517	male	Remove this column			
10.	Tirant lo Blanc	Marcorell, Joanot	3) Added 2008		1280		pre-1700s			Q559667 Choose new match		Q1777502	male	Move column to beginning			
11.	La Celestina	Rojas, Fernando de	3) Added 2008		1154		pre-1700s			Q583705 Choose new match		Q363836	male	Move column to end			
12.	Amadis of Gaul	Montalvo, Garci Rodriguez de	3) Added 2008		1195		pre-1700s			Q452075 Choose new match		Q353841	male	Move column left			
13.	The Life of Lazarillo de Tormes	Anonymous	3) Added 2008	La vida del Lazarillo de Tormes	1001		pre-1700s			Q770895 Choose new match		Q723468	male	Move column right			
14.	Gargantua and Pantagruel	Rabelais, François	1) core list		1055	French	pre-1700s			Q822290 Choose new match		Q131018	male	545871	657493	39	90
15.	The Lusiads	Camões, Luís de	3) Added 2008		293		pre-1700s			Q781898 Choose new match		Q590	male	222947	334243	37	86
16.	Monkey: Journey to the West	Wu, Cheng'en	3) Added 2008		294		pre-1700s			Q70784 Choose new match		Q228889	male	1758091	130947	66	50
17.	Euphuus: The Anatomy of Wit	Lyly, John	2) Deleted 2008		1297	English	pre-1700s			Q5410723 Choose new match		Q261627	male	21599	47647	3	25
18.	The Unfortunate Traveller	Nashe, Thomas	1) core list		1200	English	pre-1700s			Q1588078 Choose new match		Q443042	male	13649	39830	2	24
19.	Thomas of Reading	Deloney, Thomas	3) Added 2008		1307		pre-1700s			Q127500887 Choose new match		Q1530410	male		3466	0	7
20.	The Golden Ass	Apuleius, Lucius	1) core list		465	Roman	pre-1700s			Q1044767 Choose new match		Q170512	male	302843	259405	31	71
21.	Don Quixote	Cervantes Saavedra, Miguel de	1) core list		840	Spanis h	pre-1700s			Q480 Choose new match		Q5682	male	4291046	2272986	123	176
22.	The Travels of Persiles and Sigismunda	Cervantes Saavedra, Miguel de	3) Added 2008	Los trabajos de Persiles y	1123	Spanis h	pre-1700s			Q2634153 Choose new match		Q5682	male	27263	2272986	13	176

javascript:0



Geografische Verteilung der Werke nach Geodaten aus GND und Wikidata



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

Erweiterte Suche

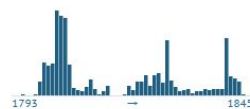
1
2
3
...

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 1845
1793 (2)
1796 (1)
1797 (18)
1798 (44)
1799 (39)

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)

```

Facet / Filter Undo / Redo 270 / 270

Refresh Reset All Remove All

187 rows

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

Facet / Filter

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
- Faoug 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

	All	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 filter
Edit cells
Edit column
Transpose
Sort...
View
Reconcile

Text facet
Numeric facet
Timeline facet
Scatterplot facet
Custom text facet...
Custom Numeric Facet...
Customized facets

www.geonames.org/2867714
www.geonames.org/2867714
www.geonames.org/2867714
www.geonames.org/2867714
www.geonames.org/2867714



Eure Erfahrung

Datenbereinigung

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](#).



QUADRIGA

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

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



Die vorliegende *Open Educational Resource (OER)*^[1] wurde in Form eines *Jupyter Books*² auf der Grundlage einer Fallstudie entwickelt. Sie befasst sich mit einer filmwissenschaftlichen Fragestellung und erkundet, wie für deren Bearbeitung digitale Verfahren sinnvoll eingesetzt werden können. Im Mittelpunkt der exemplarischen Untersuchungen steht das Filmarchiv der Hochschule für Film und Fernsehen (HFF) "Konrad Wolf", das das

🔗 Fragen oder Feedback

[Stellen Sie eine Frage ↗](#)

[Geben Sie uns Feedback](#)

Mit Ihren Rückmeldungen können wir unser interaktives Lehrbuch gezielt an Ihre Bedürfnisse anpassen.

66 Zitierhinweis

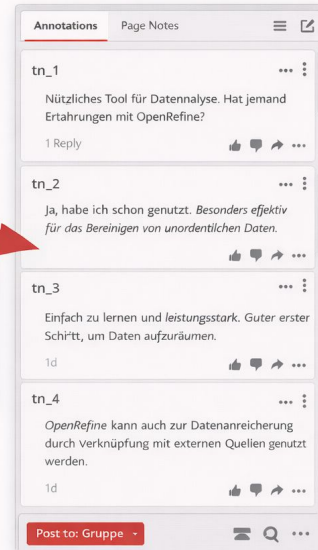
```
@book{Schick_Studentische_2025,
  title = {Studentische ...},
  author = {Schick, Thomas},
  year = {2025},
  version = {1.0.0-beta},
  doi = {10.5281/zenodo.1488888},
  url = {https://quadrant.berlin/},
  copyright = {CC-BY-SA-4.0},
  language = {deu},
  keywords = {Open Education}
```

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

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

Annotation direkt
im Text

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



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

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

Wie hat euch das **Annotieren** gefallen?



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

Potsdam | Fr. 23.1.2020

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

☐ Promotion

☐ Postdoc

☐ Sonstiges, nämlich:

3. Was war Ihre Hauptmotivation am Workshop teilzunehmen?

4. Wie haben Sie den heutigen Tag erlebt?

a) Ich habe viel gelernt.

☐ 1 (stimme überhaupt nicht zu) ☐ 2 ☐ 3 ☐ 4 ☐ 5 (stimme voll und ganz zu)

b) Ich habe mich beim Lernen in dieser Workshop-Umgebung wohlfühlt.

☐ 1 (stimme überhaupt nicht zu) ☐ 2 ☐ 3 ☐ 4 ☐ 5 (stimme voll und ganz zu)

c) Ich kann das in diesem Workshop Gelernte unmittelbar anwenden.

☐ 1 (stimme überhaupt nicht zu) ☐ 2 ☐ 3 ☐ 4 ☐ 5 (stimme voll und ganz zu)

Vielen Dank für Ihre Teilnahme!

Siehe die Karten in die Box am Ausgang einwerfen.

Herzlichen Dank!

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



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

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

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

Bearbeitet das Kapitel 5.2.3 “Anzahl der Gattungen pro Jahr”

(Wiederholung Pivot-Tabelle/-Grafik)

Bearbeitungszeit: ca. 15 Minuten

Optional: Bearbeitet das Kapitel 5.2.4 “Projektarten pro Jahr”

(Übung möglichst ohne zu spicken)

Bearbeitungszeit: ca. 15 Minuten

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

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

Wie hat euch das **Annotieren** gefallen?



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

Potsdam | Fr. 23.1.2020

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
☐ Promotion
☐ Postdoc
☐ Sonstiges, nämlich:
3. Was war Ihre Hauptmotivation am Workshop teilzunehmen?
4. Wie haben Sie den heutigen Tag erlebt?
a) Ich habe viel gelernt.
☐ 1 (stimme überhaupt nicht zu) ☐ 2 ☐ 3 ☐ 4 ☐ 5 (stimme voll und ganz zu)
b) Ich habe mich beim Lernen in dieser Workshop-Umgebung wohlfühlt.
☐ 1 (stimme überhaupt nicht zu) ☐ 2 ☐ 3 ☐ 4 ☐ 5 (stimme voll und ganz zu)
c) Ich kann das in diesem Workshop Gelernte unmittelbar anwenden.
☐ 1 (stimme überhaupt nicht zu) ☐ 2 ☐ 3 ☐ 4 ☐ 5 (stimme voll und ganz zu)

Vielen Dank für Ihre Teilnahme!

Siehe die Karten in die Box am Ausgang einwerfen.
Herzlichen Dank!



Ideen für datenbasierte Forschung