

Open Science und Open Education aus der Perspektive der Digital Humanities

Christof Schöch

Trier Center for Digital Humanities
Universität Trier

Open Education und Open Science im Dialog:
Synergien für die digitale Transformation der Hochschule
KNOER-Tagung 2025

17 Juni 2025

Einstieg

Ziele / Überblick

- 1 – Einstieg
- 2 – ‘Openness’ in den Digital Humanities
- 3 – Einheit von Open Science und Open Education in den DH
- 4 – Herausforderung der ‘Openness’ durch generative Sprachmodelle
- 5 – Fazit: Was tun?

‘Openness’ in den Digital Humanities

Was ist mit Digital Humanities gemeint?

- Grundidee
 - Forschungsfragen aus den Geisteswissenschaften
 - Methoden aus Informatik und Statistik
- Wechselwirkungen
 - Informatischer Blick auf Gegenstände
 - Geisteswiss. Blick auf Methoden
- Interdisziplinarität
 - Viele beteiligte Disziplinen
 - Unterschiedliche Mischungsverhältnisse



Digital Humanities als Fach

- 34 Fachgesellschaften weltweit
(ab 1973: Association for Literary and Linguistic Computing)
- 62 Fachzeitschriften
(ab 1966: Computers and the Humanities)
- 100+ DH-Zentren
(ab 1964: Literary and Linguistic Computing Centre, Cambridge)
- ~40 Professuren in Deutschland
(ab 2008: Fotis Jannidis in Würzburg)
- 200+ Studiengänge

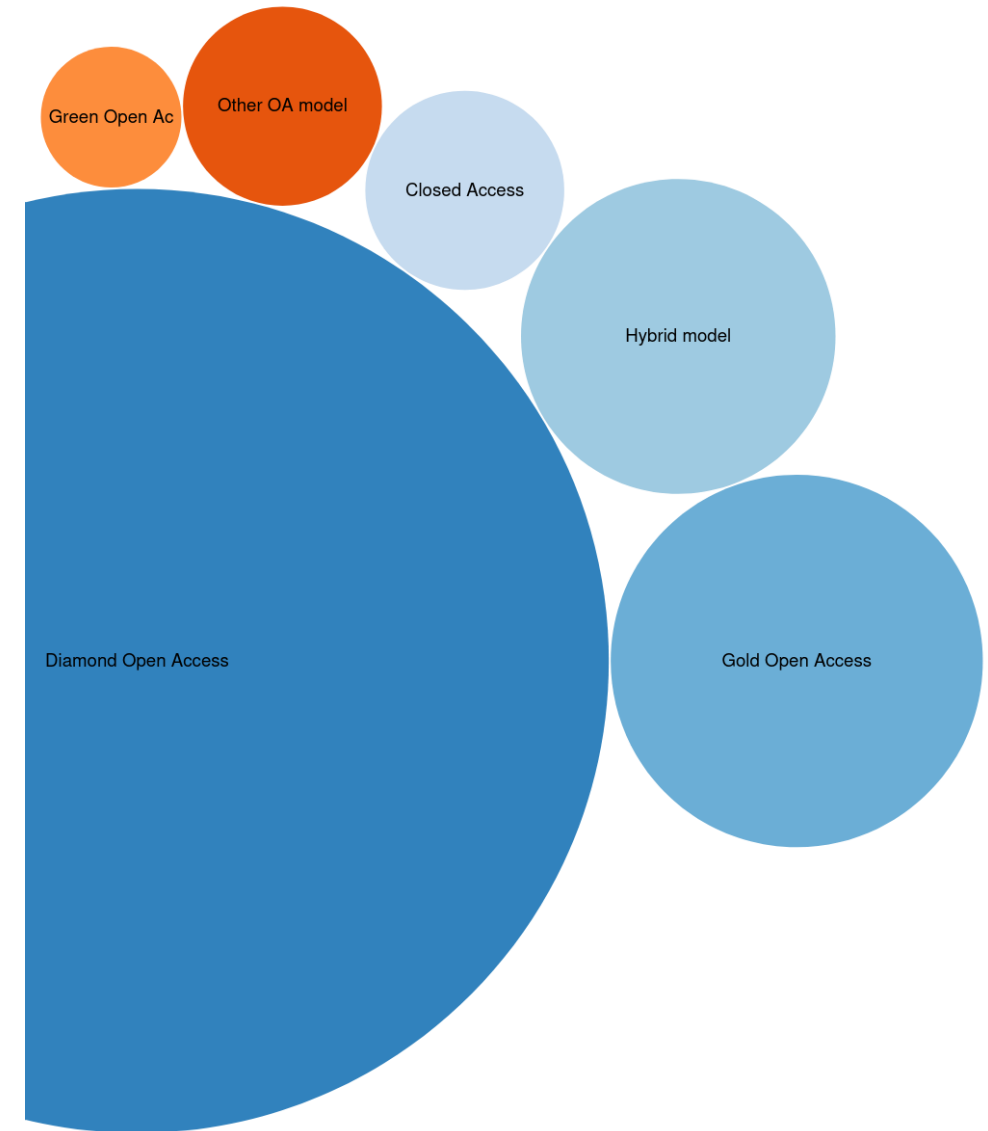


Digital Humanities und Openness

Grounded in humanistic values but catalyzed by Internet values, the digital humanities seeks to push the humanities into new territory by promoting collaboration, openness, and experimentation. [...] Openness operates on several levels in the digital humanities, describing a commitment to the open exchange of ideas, the development of open content and software, and transparency [...]. The digital humanities community embraces openness because of both self-interest and ethical aspirations. (Spiro 2012)

Digital Humanities und Openness

Grounded in humanistic values but catalyzed by Internet values, the digital humanities seeks to push the humanities into new territory by promoting collaboration, openness, and experimentation. [...] Openness operates on several levels in the digital humanities, describing a commitment to the open exchange of ideas, the development of open content and software, and transparency [...]. The digital humanities community embraces openness because of both self-interest and ethical aspirations. (Spiro 2012)



Digital Humanities und Openness

Grounded in humanistic values but catalyzed by Internet values, the digital humanities seeks to push the humanities into new territory by promoting collaboration, openness, and experimentation. [...] Openness operates on several levels in the digital humanities, describing a commitment to the open exchange of ideas, the development of open content and software, and transparency [...]. The digital humanities community embraces openness because of both self-interest and ethical aspirations. (Spiro 2012)

idle Dokument Text Edition BERGISCHE UNIVERSITÄT WUPPERTAL DFG

a catalog of **Digital Scholarly Editions**
v.4.045 2020ff, edited by Patrick Sahle et al., last change 2025-06-03+02:00

Editions catalog entries

Edenda

- ☐ single work ⁱⁱ 235
- ☐ collected works ⁱⁱ 96
- ☐ collection of texts ⁱⁱ 106
- ☐ single manuscript ⁱⁱ 120
- ☐ papers ⁱⁱ 71
- ☐ serial documents ⁱⁱ 24
- ☐ letters ⁱⁱ 141
- ☐ diaries ⁱⁱ 50
- ☐ charters ⁱⁱ 28
- ☐ inscriptions ⁱⁱ 22

Subject

- ☐ history 407
- ☐ literature 316
- ☐ theology & philosophy 61
- ☐ history of science 47
- ☐ music 27
- ☐ history of art 23

Period

- ☐ antiquity 46
- ☐ early ma 46
- ☐ high ma 85
- ☐ late ma 134
- ☐ early modern 220
- ☐ modern 390

Language ☐ Show all

- ☐ english 284
- ☐ german 236
- ☐ latin 161
- ☐ french 78
- ☐ italian 31

Sort
Publication Year

1 2 3 4 5 > Editions: 868

The Charette Project 2
Ed. by Karl D. Uitti, Gina Greco et al. Waco (TX): Baylor University, 1991-2006.
single work french high ma literature 1991

Kolb-Proust Archive for Research
Ed. by Caroline Szylowicz. Urbana (IL): University of Illinois at Urbana-Champaign, 1994-2006.
papers french modern literature 1994

Transforming Middelmarch: A Genetic Edition of Andrew Davies' 1994 BBC
Justin Smith; Gabriel Egan; Anna Blackwell. Leicester: De Montfort University, 2022.
single work english modern literature 1994

The Samuel Taylor Coleridge Archive
Created by Marjorie A. Tiefert. Charlottesville (VA): University of Virginia, 1994-1999.
collected works english modern literature 1994

Alciato's Book of Emblems: The Memorial Web Edition in Latin and English
Ed. by William Barker, Mark Feltham and Jean Guthrie. St. John's (NL): Memorial University of
single work latin early modern literature 1995

Eleventh-Century Anglo-Saxon Glossary From Ms. Brussels, Royal Library 16
David W. Porter. Baton Rouge (LA): Southern University, 1995/1996.
single manuscript english high ma literature 1995

Erste Zwischenbilanz



Woher kommt die tiefe Verbundenheit der DH mit 'Offenheit'?

- Rahmenbedingungen
 - Affinität mit Internet- und Hacker-Kultur
 - Digitalität und Open Science (Jäckel)
- Überzeugungen
 - Open Science: just science done right
 - Öffentlich finanziert => öffentlich zugänglich



Shout out to [Melanie Immig](#) for the awesome sticker!

Einheit von Open Science und Open Education in den DH

Computational Literary Studies Infrastructure

- *DraCor*-Dramenkorporus
 - XML-TEI sowie LOD und API
 - unterstützt Reproduzierbarkeit
- “Repetitive Research”-Paper
 - Theoretisches Paper zu Reproduzierbarkeit in den DH
 - Open Access (aber: Gold OA)
- *Survey of Methods in CLS*
 - Grid: frei wählbare Lektüre-Ordnung
 - Open Access und offene Formate
 - Offene Technologien: Gitlab, Zotero, Quarto

The screenshot displays the DraCor website interface. At the top, there is a navigation bar with links: ABOUT, CORPORA, TOOLS, HOW TO, and MERCH. On the right, there are logos for GitHub and TEI, along with a '2022 RAATZ PRIZE' badge. The main content area is divided into two panels: 'Fre DraCor' (French Drama Corpus) and 'Ger DraCor' (German Drama Corpus).

Fre DraCor (French Drama Corpus)	Ger DraCor (German Drama Corpus)
1,940 Number of plays	731 Number of plays
17,831 (M: 7958, F: 4247) Number of characters	15,778 (M: 11147, F: 3296) Number of characters
17,701,440 Text tokens	11,570,935 Text tokens
509,958 (16,295,137) (Tokens)	459,589 (11,026,173) (Tokens)
48,755 (393,275) (Tokens)	218,784 (1,351,557) (Tokens)
Last update 10/4/2024, 4:20:44 PM	Last update 6/9/2025, 9:07:58 PM

Below the statistics, there is a section titled 'If you want to cite DraCor, please use the following reference:' with icons for BibTeX and RIS. The reference text is: Fischer, Frank, et al. (2019). Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama. In *Proceedings of DH2019: "Complexities"*, Utrecht University, [doi:10.5281/zenodo.4284002](https://doi.org/10.5281/zenodo.4284002).

Drama Corpora Project

Unless otherwise stated, all corpora and the web design are released under Creative Commons [0 1.0 CC BY](https://creativecommons.org/licenses/by/4.0/)

At the bottom, there are version numbers for the infrastructure components: DraCor Frontend 2.4.0-rc.1, DraCor API 1.0.2, and eXist-db 6.2.0.

Siehe: dracor.org, Schöch (2023), surveyofclsinfra.io, [zenodo](https://zenodo.org) | CC BY

Computational Literary Studies Infrastructure

- *DraCor*-Dramenkorporus
 - XML-TEI sowie LOD und API
 - unterstützt Reproduzierbarkeit
- “Repetitive Research”-Paper
 - Theoretisches Paper zu Reproduzierbarkeit in den DH
 - Open Access (aber: Gold OA)
- *Survey of Methods in CLS*
 - Grid: frei wählbare Lektüre-Ordnung
 - Open Access und offene Formate
 - Offene Technologien: Gitlab, Zotero, Quarto

[Home](#) > [International Journal of Digital Humanities](#) > [Article](#)

Repetitive research: a conceptual space and terminology of replication, reproduction, revision, reanalysis, reinvestigation and reuse in digital humanities

Original Paper | [Open access](#)
Published: 06 November 2023

Volume 5, pages 373–403, (2023) [Cite this article](#)

[Download PDF](#) ↓

✓ You have full access to this [open access](#) article



[International Journal of Digital Humanities](#)

[Aims and scope](#) →

[Submit manuscript](#) →

Computational Literary Studies Infrastructure

- *DraCor*-Dramenkorporus
 - XML-TEI sowie LOD und API
 - unterstützt Reproduzierbarkeit
- “Repetitive Research”-Paper
 - Theoretisches Paper zu Reproduzierbarkeit in den DH
 - Open Access (aber: Gold OA)
- *Survey of Methods in CLS*
 - Grid: frei wählbare Lektüre-Ordnung
 - Open Access und offene Formate
 - Offene Technologien: Gitlab, Zotero, Quarto

Survey of Methods in Computational Literary Studies

=CLS INFRA D3.2: Series of five short survey papers on methodological issues

EDITORS

Christof Schöch, Julia Dudar, Evgeniia Fileva

PUBLISHER

CLS INFRA

PUBLISHED

May 5, 2023 (v1.1.0)

DOI

10.5281/zenodo.7892112

Front Matter

Overview

	General Remarks	Authorship Attribution	Genre Analysis	Literary History	Gender	Canonicity / Prestige
General Remarks	General Introduction	What is Authorship Attribution?	What is Genre Analysis?	What is Literary History?	What is Gender Analysis?	What is Canonicity?
Corpus Building	Introduction to Corpus Building	Corpus Building for Authorship Attribution	Corpus Building for Genre Analysis	Corpus Building for Literary History	Corpus Building for Gender Analysis	Corpus Building for Canonicity
Preprocessing and Annotation	Introduction to Preprocessing and Annotation	Annotation for Authorship Attribution	Annotation for Genre Analysis	Annotation for Literary History	Annotation for Gender Analysis	Annotation for Canonicity
Data Analysis	Introduction to Data Analysis	Analysis in Authorship Attribution	Data Analysis for Genre	Analysis in Literary History	Analysis for Gender	Analysis of Canonicity
Evaluation	Introduction to Evaluation	Evaluation in Authorship Attribution	Evaluation in Genre Analysis	Evaluation in Literary History	Evaluation for Gender Analysis	Evaluation for Canonicity

Weitere Beispiele

- [The Programming Historian](#): OER als Diamond-OA Zeitschrift mit Open Peer Review
- [Mining and Modeling Text](#): Forschungsprojekt zu LOD mit offenen Tutorials
- [Journal of Computational Literary Studies](#): OA-Zeitschrift mit Open Data and Code policy
- Abgeleitete Textformate ([Schöch u. a. 2020](#)): Open Science-Strategie für urheberrechtlich geschützte Materialien
- [Fedihum](#): Offene Wissenschaftskommunikation über Mastodon-Instanz des DHd-Verbands
- Geisteswiss. NFDI-Konsortien: Weiter Infrastrukturbegriff / FAIR-Principles
- uvm.

Blinder Fleck: Lehrbücher / Monographien

- Lehrbücher
 - identitätsstiftend
 - oft nicht OA
- OA-Monographien als Herausforderung
 - Sichtbarkeit / Auffindbarkeit (BASE, DARIAH Teach)
 - Prestige
 - Kosten (bei OA mit Verlag)



Siehe: (Jannidis und Kohle 2017; Karsdorp, Kestemont, und Riddell 2021; Klinke 2025).

Zweite Zwischenbilanz

Woher kommt die enge Integration von OS und OE in den DH?

- 'Offenheit' als Wert und Überzeugung
 - OS und OE werden nicht separat übernommen
 - Sondern von Offenheit abgeleitet
 - Breite Anwendung: Lehre, Forschung, Infrastruktur, WissKomm, etc.
- Interdisziplinärer Arbeitskontext
 - Arbeitsteilung => Egalität => reduzierte Wertungsdifferenz Forschung / Lehre
 - Forschungsmethoden: importieren, adaptieren (OS) => vermitteln (OE)

Openness und generative Sprachmodelle: Herausforderungen

(1) Ersetzen generative LLMs unsere OERs?

- Vorteile
 - Interaktivität ('Nachfragen')
 - Individuelles Feedback (vgl. Stackoverflow)
 - Mehrsprachigkeit
- Probleme
 - Keine Referenzfunktion / Verlässlichkeit
 - Signifikante Lücken (bspw. SPARQL, Typst)
- Opportunity
 - OER auch zu neueren Themen erstellen
 - LLMs für mehrsprachige OER nutzen
 - OER zum kritischen Umgang mit genLLMs

(2) Offene Lizenzierung unter Druck

- Mögliche Entwicklung
 - Forschende und Lehrende verzichten lieber auf offene Lizenzen
 - Motivation: Kein LLM-Training ermöglichen
- Gegenargumente
 - Lizenzvereinbarungen der Verlage (Taylor & Francis und Microsoft)
 - Impact der eigenen Ressourcen durch Training
 - Ideal der Offenheit hochhalten!

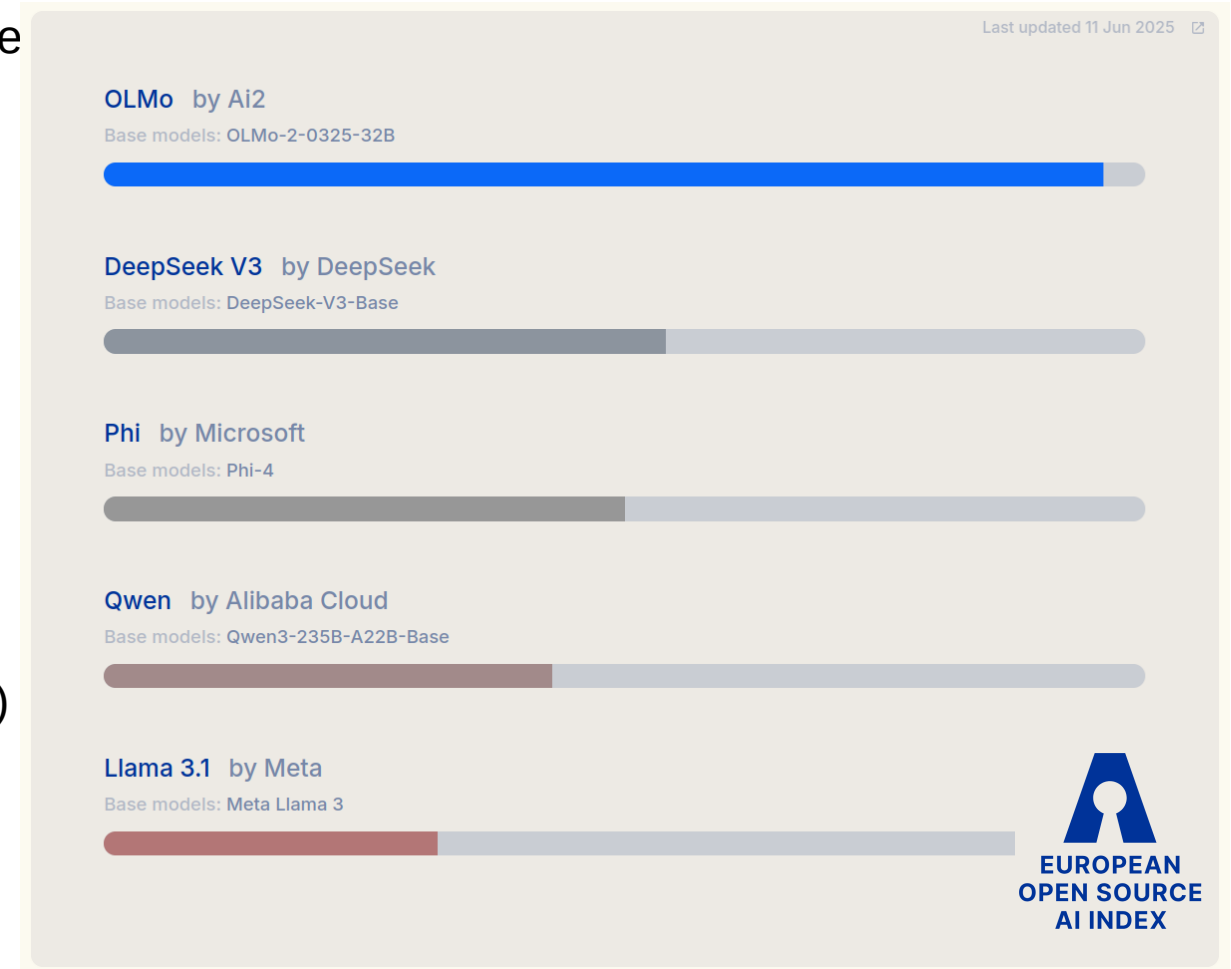
(3) Viele genLLMs sind nicht 'offen'!

- Herausforderungen

- European Open Source AI Index zeigt: bekannte LLMs sind nicht offen
- 'Open Washing': Offenheit wird redefiniert / untergraben
- Gedanke: Offene LLMs haben weniger gute Performance

- Vorteile offener und lokaler Modelle

- transparenter und reproduzierbarer (Versionierung)
- bieten Rechtssicherheit und Privatsphäre (lokal)
- keine Kosten für die API-Nutzung (aber: Hardware!)



Beispiel: Ollama

- Frei verfügbare Software zur Nutzung von offenen LLMs
- Unterstützt Download und Interaktion mit LLMs
- Zahlreiche Modelle für Text, Bilder, Video sind verfügbar
- Ermöglicht interaktives Prompting und Skripting



Get up and running with large language models.

Run DeepSeek-R1, Qwen 3, Llama 3.3, Qwen 2.5-VL, Gemma 3, and other models, locally.

Download ↓

Available for macOS,
Linux, and Windows

Siehe: ollama.com.

dhtrier.quarto.pub/knoer | [Zenodo](https://zenodo.org) | CC BY



Beispiel: Ollama

- Frei verfügbare Software zur Nutzung von offenen LLMs
- Unterstützt Download und Interaktion mit LLMs
- Zahlreiche Modelle für Text, Bilder, Video sind verfügbar
- Ermöglicht interaktives Prompting und Skripting

```

28
29 # === Wine label description ===
30
31 def describe_label(model, image):
32     res = ollama.chat(
33         model=model,          # llava|minicpm-v|llama3.2-vision|llama4
34         messages=[
35             {
36                 'role': 'user',
37                 'content': 'You are an expert in the formal analysis of historical wine
                                labels. Proceed in step by step when analysing the image provided to you.
                                First, make a list with the name and description of each object shown
                                visually on this wine label. Second, make a list with any text that is
                                printed on the label.',
38                 'images': [image]
39             }
40         ]
41     )
42     print("\n===\n", model, "\n", image, "\n", res['message']['content'])
43
44     with open("responses.md", "a", encoding="utf8") as outfile:
45         outfile.write("\n\n=====")
46         outfile.write("\n" + str(datetime.now())[0:19])
47         outfile.write("\n" + model)
48         outfile.write("\n" + image)
49         outfile.write("\n\n" + res['message']['content'])
50
51
52 for image in images:
53     image = "./img/" + image
54     describe_label(model, image)

```

Dritte Zwischenbilanz



Was ist die Antwort auf die generativen Sprachmodelle?

- OER zu Themen erstellen, bei denen LLMs schwach sind
- OER zur Nutzung von offenen LLMs erstellen und verbreiten
- Die Forderung nach mehr offenen LLMs mittragen



Fazit: Was tun?

Abgeleitete Handlungsoptionen

- Offenheit an sich als Prinzip / Wert erklären und fördern
- Integration von OS und OE: dort unterstützen, wo schon digital geforscht wird
- Mit OER in strategische Lücken gehen
 - unterrepräsentierte Themen
 - offene generative LLMs
- Sich für diese Themen mit den NFDI-Konsortien solidarisieren

Vielen Dank für Ihre Aufmerksamkeit!



References

- Helregel, Nicole. 2025. „AI and Open Science: Implications and Library Practice Recommendations“. *Library Trends* 73 (3): 177–96. <https://doi.org/10.1353/lib.2025.a961191>.
- Jannidis, Fotis, und Malte Kohle Hubertus und Rehbein, Hrsg. 2017. *Digital Humanities: Eine Einführung*.
- Karsdorp, Folgert, Mike Kestemont, und Allen Riddell. 2021. *Humanities Data Analysis. Case Studies with Python*.
- Klinke, Harald. 2025. *Cultural Data Science. An Introduction to R*.
- Sahle, Patrick. 2015. „Digital Humanities? Gibt's doch gar nicht!“ *Zeitschrift für digitale Geisteswissenschaften*. https://doi.org/10.17175/sb001_004.
- Schöch, Christof. 2023. „Repetitive research: a conceptual space and terminology of replication, reproduction, revision, reanalysis, reinvestigation and reuse in digital humanities“. *International Journal of Digital Humanities* 5:373–403. <https://doi.org/https://doi.org/10.1007/s42803-023-00073-y>.
- Schöch, Christof, Frédéric Döhl, Achim Rettinger, Evelyn Gius, Peer Trilcke, Peter Leinen, Fotis Jannidis, Maria Hinzmann, und Jörg Röpke. 2020. „Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen“. *Zeitschrift für digitale Geisteswissenschaften*. https://www.zfdg.de/2020_006.
- Spiro, Lisa. 2012. „'This Is Why We Fight': Defining the Values of the Digital Humanities“. In *Debates in the Digital Humanities*, herausgegeben von Matt Gold. Univ. of Minnesota Press. <https://dhdebates.gc.cuny.edu/projects/debates-in-the-digital-humanities>.

