



Melusina Press
Humanities

From Global to Local?

**Digitale Methoden in den Geisteswissenschaften im
deutschsprachigen Raum: Ein Triptychon**

*Ulrike Wuttke, Christopher Nunn, Christian Schröter,
Melanie Seltmann & Christian Wachter*

**Lexikographische Daten als Linked Open Data
Historischer Pandemiewortschatz in PandeLexBase und
Pandemicictionary**

Susanne Kabatnik
Anne Klee
Maria Hinzmann
Claudia Bamberg
Frank Queens
Thomas Burch

**From Global to Local? Digitale Methoden in den
Geisteswissenschaften im deutschsprachigen Raum: Ein Triptychon**
Herausgegeben von Ulrike Wuttke, Christopher Nunn,
Christian Schröter, Melanie Seltmann, Christian Wachter

Publiziert in Melusina Press (Universität Luxemburg), 2026
11, Porte des Sciences
L-4366 Esch-sur-Alzette
<https://www.melusinapress.lu>

Verlagsleitung: Niels-Oliver Walkowski, Johannes Pause
Lektorat: Carolyn Knaup, Niels-Oliver Walkowski, die Herausgeber*innen
Gestaltung: Valentin Henning, Erik Seitz
Umschlagsbild: Sara-Lee Nußbaum

Die digitale Version dieser Publikation steht unter <https://www.melusinapress.lu> frei zur Verfügung.

Bibliografische Information der Nationalbibliothek Luxemburg:
Die Nationalbibliothek Luxemburg verzeichnet diese Publikation in der Luxemburgischen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über bnl.public.lu abrufbar.

Druck: Libri Plureos GmbH, Friedensallee 273, 22763 Hamburg.

DOI (Publication): 10.26298/1981-5722-ldal
DOI (Container): 10.26298/1981-5722
ISBN (Web): 978-2-919815-72-2
ISBN (PDF): 978-2-919815-73-9
ISBN (Print): 978-2-919815-74-6

Das vorliegende Werk steht unter einer CC BY-SA 4.0 Lizenz. Informationen zu dieser Lizenz finden Sie unter <https://creativecommons.org/licenses/by-sa/4.0/deed.de>. Die in diesem Werk enthaltenen Bilder und Ressourcen unterliegen der selben Lizenz, sofern sie keiner anderen Quelle entnommen wurden oder mit einer anderen Lizenz versehen sind.





2026

Lexikographische Daten als Linked Open Data. Historischer Pandemiewortschatz in PandeLexBase und Pandemicictionary

Susanne Kabatnik, Anne Klee, Maria Hinzmann,
Claudia Bamberg, Frank Queens & Thomas Burch

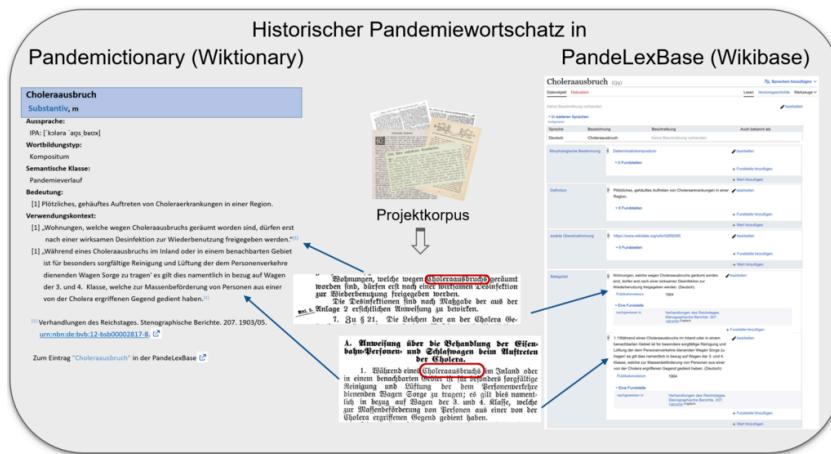


ABBILDUNG 1: Anne Klee, Artefakt „in progress“. Historischer Pandemie-
wortschatz in zwei Wiki-Instanzen.

Typ: Knowledge Graph und Online-Wörterbuch. *Autor*innen:* Susanne Kabatnik, Anne Klee, Maria Hinzmann, Claudia Bamberg, Frank Queens, Thomas Burch. *Status:* in Entwicklung. *Software:* MediaWiki, Wikibase. *Medium:* Onlineressource. *Sprache:* Deutsch, Englisch.

Zur Artefaktbiographie

Im Kontext der Bestrebungen nach einer globalen Standardisierung für weltweite Vernetzung und Nachhaltigkeit im Sinne von Open Science bei gleichzeitigen, spezifischen Anforderungen von lokalen Einzelprojekten gewinnt das Linked Open Data (LOD)-Paradigma zunehmend an Bedeutung.¹ Es gibt jedoch bisher nur wenige Ansätze, die Wissensgebiete vollständig in LOD umwandeln, um Ergebnisse systematisch zu vernetzen.

Die Idee von LOD zielt darauf ab, unstrukturierte und unverbundene Dokumente und Daten zu vernetzen und als offene, maschinenlesbare und semantisch interpretierbare Daten nach den FAIR-Prinzipien der Auffindbarkeit, Zugänglichkeit, Interoperabilität und Wiederverwendbarkeit verfügbar zu machen.² Eine vielversprechende Domäne für die LODifizierung von Forschungsdaten ist die digitale Lexikographie, da sich zentrale lexikografische Praktiken – wie die Generierung von Glossaren und ihre Systematisierung, das Benennen und Beschreiben sprachlicher Einheiten – nahtlos in das Paradigma der LOD einfügen. Für die Publikation linguistischer und insbesondere lexikographischer Daten im Semantic Web hat sich das OntoLex-Lemon-Modell als De-facto-Standard etabliert.³ Es bil-

¹ Katherine Thornton et al., „Linking Women Editors of Periodicals to the Wikipedia Knowledge Graph,“ *Semantic Web* 14, Nr. 2 (15. Dezember 2022): 443–55, <https://doi.org/10.3233/SW-222845>; Daniel Alves, Hg., *IJHAC: A Journal of Digital Humanities* 16: Special Issue: Linked Open Data in the Arts and the Humanities, Nr. 1 (2022); Milica Ikonić Nešić et al., „From ELTeC Text Collection Metadata and Named Entities to Linked Data (and Back),“ in *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference* (Marseille: European Language Resources Association, 2022): 7–16, <https://aclanthology.org/2022.ldl-1.2>; Fudie Zhao, „A Systematic Review of Wikidata in Digital Humanities Projects,“ *Digital Scholarship in the Humanities* 38, Nr. 2 (28. Dezember 2022): 852–74, <https://doi.org/10.1093/lhc/fqac083>.

² Tim Berners-Lee, „Linked Data: Design Issues,“ W3C, zuletzt geändert am 18. Juni 2009, <https://www.w3.org/DesignIssues/LinkedData.html>; Mark D. Wilkinson et al., „The FAIR Guiding Principles for Scientific Data Management and Stewardship,“ *Scientific Data* 3, Nr. 1 (15. März 2016): 160018, <https://doi.org/10.1038/sdata.2016.18>.

³ Philipp Cimiano, et al., *Linguistic Linked Data* (Cham: Springer International Publishing, 2020); John P. McCrae und Thierry Declerck, „Linguistic Linked Open Data for All,“ *Zenodo* (5. Dezember 2019), <https://doi.org/10.5281/zenodo.3607272>; Anas Fahad Khan et al., „When Linguistics Meets Web Technologies: Recent Advances in Modelling Linguistic Linked Data,“ *Semantic Web* 13, Nr. 6 (26. September

det die Grundlage vieler Ressourcen in der Linguistic Linked Open Data Cloud, eine Visualisierung des Netzwerkes aller als LOD veröffentlichten linguistischen Datensätze.⁴ Eine aktuelle Entwicklung ist der Aufbau eines globalen Lexemnetzwerks in Wikidata, das ebenfalls durch das Onto-Lex-Lemon-Modell unterstützt wird. Wörterbuchressourcen können dabei entweder direkt in Wikidata importiert oder über Referenzen angebunden werden, was die Verknüpfung lokaler Daten in einem globalen Netzwerk ermöglicht.⁵

In einem Teilprojekt des Trierer Verbundprojekts LODinG – Linked Open Data in den Geisteswissenschaften⁶ wird am Trier Center for Digital Humanities (TCDH) Cholera-Wortschatz nach LOD-Prinzipien lexikographisch aufbereitet. Im Folgenden werden der Projektaufbau sowie insbesondere die beiden entstehenden Artefakte PandeLexBase und Pandemicictionary vorgestellt. Ausgangspunkt bildet dabei die Frage, wie sich globale und lokale Forschungsfragen verbinden und mit dem LOD-Paradigma in Einklang bringen lassen.

2022): 987–1050, <https://doi.org/10.3233/SW-222859>; Philipp Cimiano et al., „Lexicon Model for Ontologies: Community Report, 10 May 2016.“ W3C, zuletzt geändert am 10. Mai 2016, <https://www.w3.org/2016/05/ontolex/>; Sabine Tittel, *Historische Lexikalische Semantik und Linked Data: Modellierung von Ressourcen der mittelalterlichen Galloromania für das Semantic Web* (Berlin, Boston: De Gruyter, 2024), <https://doi.org/10.1515/9783111326481>.

4 Christian Chiarcos et al., „Building a Linked Open Data Cloud of Linguistic Resources: Motivations and Developments,“ in *The People’s Web Meets NLP*, hg. von Iryna Gurevych et al. (Berlin, Heidelberg: Springer, 2013), 315–48; John P. McCrae et al., „The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud,“ in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)* (Portorož: European Language Resources Association (ELRA), 2016): 2435–41, <https://aclanthology.org/L16-1386>.

5 David Lindemann et al., „When OntoLex Meets Wikibase: Remodeling Use Cases,“ in *The 4th Wikidata Workshop. Proceedings of the Wikidata Workshop 2023* (13. November 2023), <https://ceur-ws.org/Vol-3640/paper14.pdf>; Finn Årup Nielsen, „Lexemes in Wikidata: 2020 Status,“ in *Proceedings of the 7th Workshop on Linked Data in Linguistics within the 12th Language Resources and Evaluation Conference (LDL-2020)* (Marseille: European Language Resources Association (ELRA), 2020): 82–86, <https://aclanthology.org/2020.ldl-1.12.pdf>.

6 „LODinG – Linked Open Data in den Geisteswissenschaften,“ Universität Trier, aufgerufen am 30. Juli 2025, <https://www.uni-trier.de/forschung/nationale-forschungsfoerderung/forschungsinitiative-rheinland-pfalz-2024-2028/loding>.

Daten und Methode

Da sich die Rekonstruktion von historischem Wortschatz aufgrund unvollständiger oder unzugänglicher Quellen sowie der Veränderung der Schriftsprache schwierig gestaltet, wurden zur Erstellung eines Glossars zu historischen Pandemien, hier zunächst am Beispiel der Cholera, Daten aus mehreren Quellen zusammengetragen: (1) Wörterbücher des Trierer Wörterbuchnetzes und des Digitalen Wörterbuchs der deutschen Sprache (DWDS); (2) Sekundärliteratur zu Cholera und anderen Pandemien und darin zitierte Texte, etwa Berichte, Gedichte, Briefe und autobiographische Werke aus dem 19. und 20. Jahrhundert⁷ und (3) historische Korpora wie das des DWDS (2024; 1,53 Milliarden Tokens) und des Deutschen Referenzkorpus (DeReKo-2024-I, etwa 70 Millionen Tokens). Diese Korpora wurden mit Hilfe der Software Sketch Engine durchsucht, um spezifischen Cholera-Wortschatz zu identifizieren und die Frequenz von Lexemen, Keywords, Kollokationen und Komposita auszuwerten. Die daraus entstandenen Wortlisten wurden manuell bereinigt und es wurde ein Glossar mit aktuell 1.300 Einträgen zu historischem Pandemie-Wortschatz wie Krankheitsbezeichnungen, Metaphern und Cholera-Komposita erstellt.

Komplementäre Nutzung von Pandemicionary und PandeLexBase

Im Projekt werden zwei komplementäre Ressourcen, die PandeLexBase und das Pandemicionary, miteinander verknüpft, um unterschiedliche Vorteile zu kombinieren. Beide Ressourcen basieren auf MediaWiki-Instanzen, bieten jedoch unterschiedliche Datenumfänge und Formate. Das Pandemicionary folgt einer vertrauten Wiktionary-Struktur mit standardisierten Einträgen, die eine sorgfältige Kuratierung und authentische

⁷ Norman Aselmeyer, „Cholera und Tod: Epidemieerfahrungen und Todesanschauungen in autobiografischen Texten von Arbeiterinnen und Arbeitern“, *Archiv für Sozialgeschichte* 55 (2015): 77–106, https://library.fes.de/pdf-files/afs/bd55/afs55_08_aselmeyer.pdf; Barbara Dettke, *Die asiatische Hydra: Die Cholera von 1830/31 in Berlin und den preußischen Provinzen Posen, Preußen und Schlesien* (Berlin, New York: de Gruyter, 1995).

Beispiele erfordert. Im Gegensatz dazu stellt die Wikibase-Instanz als Graphdatenbank (*Triplestore*) eine umfassendere und weniger kuratierte Datenbasis zur Verfügung, einschließlich Einträgen mit nur einem Beleg und ohne Selektion der Belege. Das Projekt verfolgt das Ziel, ‚lokale‘ Daten – sowohl räumlich als auch zeitlich fokussierte – so zu repräsentieren, dass sie global verknüpfbar sind. Während die Daten auf der lokalen Ebene im spezifischen Anwendungskontext abhängig vom Forschungsinteresse genutzt werden können – entweder aufbereitet als Wörterbucheinträge im Pandemionary oder durch gezielte Abfrage der PandeLexBase – sind sie auf globaler Ebene im Semantic Web eingebettet und so mit Daten anderer Domänen und Formate vernetzt. Das Wikibase-Ökosystem fungiert als Vermittlungsebene zwischen diesen beiden Polen.

PandeLexBase: Wikibase als Forschungsplattform

In der Wikibase-Instanz des Projekts, der PandeLexBase, sind die Pandemiebegriffe die zentralen Einheiten, die als *Items* mit einer eindeutigen Q-ID identifiziert werden (vgl. Abbildung rechts). Diese Begriffe werden durch *Properties* in Form von *Triplestatements* charakterisiert, wobei die Objektposition entweder durch Elemente eines kontrollierten Vokabulars, Strings wie Definitionen oder Belegzitate oder durch URLs besetzt ist. Die Informationen umfassen morphologische Eigenschaften, semantische Merkmale (z.B. Pandemiezugehörigkeit, semantische Klasse, Definition) und die ungefilterte Menge aller Textbelege aus dem Projektkorpus.

Die PandeLexBase nutzt kontrollierte Vokabulare wie LexInfo und OLiA für linguistische Datenkategorien und verknüpft zu Normdatensätzen wie Wikidata und dem grammatischen Informationssystem grammis. Dies stellt die Anbindung an das Semantic Web und die konzeptuelle Interoperabilität sicher. Nutzende können durch komplexe Abfragen in SPARQL spezifischen Forschungsinteressen nachgehen. So können die Daten in der Linguistik zur Untersuchung des Sprachwandels und zu Wortschatzanalysen dienen. Beispielsweise können Abfragen pandemieübergreifend nach allen Begriffen desselben Wortbildungsmusters (X-freien/X-maßnahme), derselben semantischen Klasse (Emotion, Maßnahme, Pandemieverlauf), nach Synonymen (z.B. für Cholera wie Koller, Pestruhr und Würgeengel) sowie in Bezug auf die Klassifizierung von Texten nach Genre und Zeit mit direktem Zugang zur Quelle gestellt wer-

den, sodass der Zugriff auch für geschichtswissenschaftliche oder soziologische Untersuchungen zu gesellschaftlichen Veränderungen in Pandemien nützlich sein kann.

Pandemictionary: Ein digitales Wörterbuch zu historischen Pandemien

Die zweite Instanz des Projekts ist ein Wörterbuch zu historischen Pandemien, das nach dem Vorbild des Wiktionary in Form eines Wikis aufgebaut ist und sich in Darstellung und Zielgruppe von der Wikibase-Instanz unterscheidet (vgl. Abbildung 1, links). Die Informationen zu den Stichwörtern sind ähnlich wie in einem klassischen Wörterbuch angeordnet und farblich sowie typografisch voneinander abgegrenzt, und folgen einer festgelegten Reihenfolge: Stichwort, Wortart, Genus, Aussprache, Wortbildungstyp, semantische Klasse und Bedeutung, wobei grammatische Kategorien mit Artikeln im Grammatischen Informationssystem des IDS verlinkt sind.⁸ Die Stichwörter können alphabetisch oder über eine manuelle Suche ermittelt werden. Die Zielgruppe sind Nutzer*innen wie beispielsweise Sprach- oder Pandemieinteressierte (auch Lai*innen), die mit der technischen Komplexität der Wikibase-Instanz Schwierigkeiten haben könnten. Für sie werden Einträge nach historischer Pandemie, Wortbildungsmustern oder semantischen Klassen vorsortiert und als Liste abrufbar gemacht. Zudem sind die Artikel mit Quellen für Belegbeispiele und der PandeLexBase verlinkt, sodass Nutzer*innen je nach Bedarf entscheiden können, welche Instanz sie nutzen möchten.

Lokale und globale Vernetzung des historischen Pandemiewortschatzes

Das Pandemictionary-Projekt hat das Ziel, historische Pandemiebegriffe sowohl lokal als auch global zugänglich zu machen, indem es zwei komplementäre Instanzen kombiniert: die PandeLexBase und das Pandemicictionary. Erstere fungiert als semantische Graphdatenbank, die Wissen-

⁸ „Startseite,“ IDS grammis – Grammatisches Informationssystem, aufgerufen am 30. Juli 2025, <https://grammis.ids-mannheim.de/>.

schaftler*innen aus verschiedenen Disziplinen differenzierte Forschungsabfragen ermöglicht. Sie bietet eine breite, weniger kuratierte Datenbasis. Im Gegensatz dazu ist das Pandemictionary ein benutzerfreundliches, klassisches Wörterbuch, das für eine breite Zielgruppe optimiert ist. Diese Dualität erlaubt eine flexible und umfassende Erforschung des Pandemiewortschatzes. Das Projekt vereinigt dabei sowohl globale als auch lokale Ansätze. Global ermöglicht es den Zugang zu vernetzten Daten aus verschiedenen Institutionen, die Mehrsprachigkeit und internationale Kooperationen in den Vordergrund stellen. Lokal ist das Projekt in Trier angesiedelt und fokussiert sich auf die Erfassung des pandemischen Wortschatzes in einer bestimmten Sprache und deren Varianten. Obwohl die Daten lokal situiert sind, tragen sie zur globalen Verbreitung von spezifischem Wissen und somit zum historischen Pandemiewortschatz einer konkreten Sprache bei und werden in internationale Initiativen eingebunden. Die Verbindung zwischen lokal und global verläuft dabei in beide Richtungen: Lokale Daten werden global zugänglich gemacht, während globale Standards und Vokabulare für die lokale Arbeit genutzt werden. Während auf diese Weise diachrone und transregionale Vergleiche möglich sind, können gleichzeitig Begriffe in spezifischen lokalen Kontexten detailliert erforscht werden.

Bibliographie

Alves, Daniel, Hg. *IJHAC: A Journal of Digital Humanities* 16: Special Issue: Linked Open Data in the Arts and the Humanities, Nr. 1 (2022).

Aselmeyer, Norman. „Cholera und Tod: Epidemieerfahrungen und Todesanschauungen in autobiografischen Texten von Arbeiterinnen und Arbeitern.“, *Archiv für Sozialgeschichte* 55 (2015): 77–106. https://library.fes.de/pdf-files/afs/bd55/afs55_08_aselmeyer.pdf.

Berners-Lee, Tim. „Linked Data: Design Issues.“ W3C. Zuletzt geändert am 18. Juni 2009. <https://www.w3.org/DesignIssues/LinkedData.html>.

Chiarcos, Christian, Steven Moran, Pablo N. Mendes, Sebastian Nordhoff und Richard Littauer. „Building a Linked Open Data Cloud of Linguistic Resources: Motivations and Developments.“ In *The People's Web Meets NLP, Theory and Applications of Natural Language Processing*, herausgegeben von Iryna Gurevych, Jungi Kim und Nicoletta Calzolari, 315–48. Berlin und Heidelberg: Springer, 2013.

Cimiano, Philipp, Christian Chiarcos, John P. McCrae und Jorge Gracia. *Linguistic Linked Data*. Cham: Springer International Publishing, 2020. <https://doi.org/10.1007/978-3-030-30225-2>.

Cimiano, Philipp, John P. McCrae und Paul Buitelaar, Hg. „Lexicon Model for Ontologies: Community Report, 10 May 2016.“ W3C. Zuletzt geändert am 10. Mai 2016. <https://www.w3.org/2016/05/ontolex/>.

Dettke, Barbara. *Die asiatische Hydra: Die Cholera von 1830/31 in Berlin und den preußischen Provinzen Posen, Preußen und Schlesien*. Veröffentlichungen der Historischen Kommission zu Berlin, Bd. 89. Berlin, New York: De Gruyter, 1995.

IDS grammis – Grammatisches Informationssystem. „Startseite.“ Aufgerufen am 30. Juli 2025. <https://grammis.ids-mannheim.de/>.

Ikonić Nešić, Milica, Ranka Stanković, Christof Schöch und Mihailo Skoric. „From ELTeC Text Collection Metadata and Named Entities to Linked Data (and Back).“ In *Proceedings of the 8th Workshop on Linked Data in Linguistics within the 13th Language Resources and Evaluation Conference*, 7–16. Marseille: European Language Resources Association, 2022. <https://aclanthology.org/2022.ldl-1.2>.

Khan, Anas Fahad, Christian Chiarcos, Thierry Declerck, Daniela Gifu, Elena González-Blanco García, Jorge Gracia, Maxim Ionov, et al. „When Linguistics Meets Web Technologies: Recent Advances in Modelling Linguistic Linked Data.“ *Semantic Web* 13, Nr. 6 (26. September 2022): 987–1050. <https://doi.org/10.3233/SW-222859>.

- Lindemann, David, Sina Ahmadi, Fahad Khan und Francesco Mambrini. „When Ontolex Meets Wikibase: Remodeling Use Cases.“ In *The 4th Wikidata Workshop. Proceedings of the Wikidata Workshop 2023* (13. November 2023). <https://ceur-ws.org/Vol-3640/paper14.pdf>.
- McCrae, John P., Christian Chiarcos, Francis Bond, Philipp Cimiano, Thierry Declerck, Gerard de Melo, Jorge Gracia, et al. „The Open Linguistics Working Group: Developing the Linguistic Linked Open Data Cloud.“ In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2435–41. Porto-čo: European Language Resources Association (ELRA), 2016. <https://aclanthology.org/L16-1386/>.
- McCrae, John P. und Thierry Declerck. „Linguistic Linked Open Data for All.“ *Zenodo* (5. Dezember 2019). <https://doi.org/10.5281/zenodo.3607272>.
- Nielsen, Finn Årup. „Lexemes in Wikidata: 2020 Status.“ In *Proceedings of the 7th Workshop on Linked Data in Linguistics within the 12th Language Resources and Evaluation Conference (LDL-2020)*, 82–86. Marseille: European Language Resources Association (ELRA), 2020. <https://aclanthology.org/2020.ldl-1.12.pdf>.
- Thornton, Katherine, Kenneth Seals-Nutt, Marianne Van Remoortel, Julie M. Birkholz und Pieterjan De Potter. „Linking Women Editors of Periodicals to the Wikidata Knowledge Graph.“ *Semantic Web* 14, Nr. 2 (15. Dezember 2022): 443–55. <https://doi.org/10.3233/SW-222845>.
- Tittel, Sabine. *Historische Lexikalische Semantik und Linked Data: Modellierung von Ressourcen der mittelalterlichen Galloromania für das Semantic Web*. Berlin, Boston: De Gruyter, 2024. <https://doi.org/10.1515/978311326481>.
- Universität Trier. „LODinG – Linked Open Data in den Geisteswissenschaften.“ Aufgerufen am 30. Juli 2025. <https://www.uni-trier.de/forschung/nationale-forschungsfoerderung/forschungsinitiative-rheinland-pfalz-2024-2028/loding>.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. „The FAIR Guiding Principles for Scientific Data Management and Stewardship.“ *Scientific Data* 3, Nr. 1 (15. März 2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.

Zhao, Fudie. „A Systematic Review of Wikidata in Digital Humanities Projects.“ *Digital Scholarship in the Humanities* 38, Nr. 2 (28. Dezember 2022): 852–74. <https://doi.org/10.1093/llc/fqac083>.