

Baustelle nach Baukastenprinzip: Kontrollierte Vokabulare im Projekt Linked Open Data in den Geisteswissenschaften (LODinG)

Bamberg, Claudia

bamberg@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0001-6453-8457

Bruchertseifer, Jens

jens.bruchertseifer@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0003-3536-3082

Burch, Thomas

burch@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-8050-1475

Gernert, Folke

gernertf@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-6424-0711

Hinzmann, Maria

hinzmannm@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0001-7199-1436

Kabatnik, Susanne

kabatnik@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-9022-9559

Klee, Anne

klee@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-1532-2649

Kudera, Jacek

kudera@uni-trier.de
Universität Trier, Deutschland

ORCID: 0000-0003-3678-1067

Moulin, Claudine

moulin@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0001-8091-3456

Rettinger, Achim

rettinger@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0003-4950-1167

Röpke, Jörg

roepke@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-1575-6105

Schenkel, Ralf

schenkel@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0001-5379-5191

Schirra, Doris

schirra@uni-trier.de
Universität Trier, Deutschland

Schlimpen, Hannah

schlimpen@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-8544-3374

Schöch, Christof

schoech@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-4557-2753

Shi-Kupfer, Kristin

shikupfer@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0003-2499-7047

Wassermayr, Veronica

wassermayr@uni-trier.de
Universität Trier, Deutschland
ORCID: 0009-0001-9638-2677

Weis, Joëlle

weis@uni-trier.de
Universität Trier, Deutschland
ORCID: 0000-0002-0080-4362

Wie zahlreiche Beiträge u.a. bei Konferenzen wie der DH sowie der DHd belegen, gewinnt das Linked Open Data (LOD) als Paradigma in den Digital Humanities in den letzten Jahren zunehmend an Interesse (u.a. Thornton et al. 2021; Alves 2022; Ikončić Nešić et al. 2022; siehe auch den Survey in Zhao 2022). Der Wunsch, Informationen in Projekten in maschinenlesbarer und standardisierter Form zu erfassen, zu modellieren, zu vernetzen, frei zugänglich zu machen und zu analysieren, ist weit verbreitet. Davon zeugen auch die Knowledge Graphen, die derzeit im Kontext von NFDI4Memory und NFDI4Culture aufgebaut werden (Steller et al. 2024, Liebing 2024). Die praktische Umsetzung birgt jedoch oft Herausforderungen, insbesondere im Hinblick auf die Infrastruktur, die tatsächliche Vernetzung von Ressourcen sowie die Nutzung formaler und standardisierter Datenmodelle. Ebenso ist die große Vielfalt von bereits vorhandenen Vokabularen vor allem in interdisziplinären Kontexten nicht leicht zu navigieren (Abb. 1). Sie alle können als Bausteine für eine projektspezifische Terminologie dienen, wobei einfache Nachnutzbarkeit und Kompatibilität nicht immer gegeben sind.

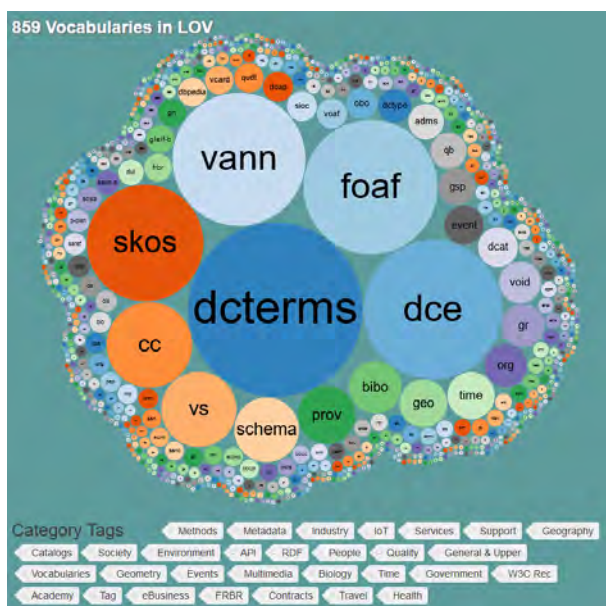


Abb.1: Übersicht der Vokabulare, die in Linked Open Vocabularies enthalten sind <https://lov.linkeddata.es/dataset/lov/>

Die Entwicklung eines kohärenten Forschungsfeldes sowie einer zunehmenden Etablierung des Paradigmas kann daher als “being under construction” beschrieben werden.

Das an der Universität Trier angesiedelte und von der Forschungsinitiative Rheinland-Pfalz geförderte Verbundvorhaben *LODinG. Linked Open Data in den Geisteswissenschaften* leistet einen Beitrag zu dieser Baustelle. Auf den stärker thematisch fokussierten Erfahrungen im Verbundvorhaben *Mining and Modeling Text* (2019-2023, siehe Hinzmann et al. 2024) aufbauend, verfolgt LODinG das Ziel, das enorme Potential von LOD für innovative Forschung in einem breiten Spektrum an Einsatzgebieten in den Geisteswissenschaften mit Fokus auf den Sprach-, Li-

teratur-, Kultur- und Medienwissenschaften zu erschließen und praktisch zu erproben.

Durch den interdisziplinären Zugang sollen Grenzen zwischen den verschiedenen Einsatzgebieten von LOD aufgebrochen sowie Ansätze nachgenutzt und Datenbestände miteinander verbunden werden. Darüber hinaus können durch die vielfältigen Einsatzgebiete sowohl qualitative als auch quantitative, automatisierte und manuelle Verfahren der Generierung, Verknüpfung und Auswertung der Daten ausgelotet werden.

In der Realisation des LOD-Paradigmas spielen kontrollierte Vokabulare eine zentrale Rolle, nicht zuletzt, da sie eine Verknüpfung auf der konkreten Entitäten-Ebene voraussetzen bzw. umgekehrt ermöglichen (Harpring 2010, Zaytseva & Đurčo 2020, Brown 2022). Die Relevanz kontrollierter Vokabulare zeigt sich in verschiedenen rezenten Initiativen, die sich der Standardisierung, besseren Sichtbarkeit und Verknüpfung widmen (Borek et al. 2021, Franc et al. 2022, Goldhahn 2024, Santini et al. 2024, Steiner et al. 2024).

Der Posterbeitrag verschränkt eine Vorstellung der Teilprojekte von LODinG mit einer Präsentation der jeweils relevanten Vokabulare. Dabei zeigt sich die Vielfalt der Teilprojekte auf der Ebene der Untersuchungsgegenstände, der Forschungsgebiete wie auch der eingesetzten Verfahren, welche wiederum in den verknüpften Vokabularen repräsentiert ist: Gegenstand sind multilinguale und multimediale Ressourcen, so werden beispielsweise Daten zum Pandemiewortschatz aus digitalen Wörterbüchern, zu Botanik und Pharmakologie aus frühneuzeitlicher medizinischer Literatur oder auch zu modernen und historischen Weinetiketten annotiert oder extrahiert, verknüpft und analysiert. Zwei andere Teilprojekte befassen sich sukzessive mit der Extraktion von semantischen Statements zum Inhalt geisteswissenschaftlicher Fachliteratur, unter anderem für den chinesischen Sprachraum

Geisteswissenschaftliches Wissen zeichnet sich dadurch aus, dass es historisch und kulturell situiert ist (Hara-way 1988), sodass unterschiedliche Perspektiven komplementär und sogar im Widerspruch nebeneinander existieren können und entsprechend zu modellieren sind. Die damit gegebene, programmatische Heterogenität repräsentiert exemplarisch Besonderheiten geisteswissenschaftlichen Wissens, aus der sich sowohl spezifische Herausforderungen als auch Potentiale für die Arbeit im LOD-Paradigma ergeben. Dabei kann an bisherige Beiträge in Bereichen wie digitaler Lexikographie und Linguistic LOD (Chiaros et al. 2020, Khan et al. 2022, Lindemann et al. 2022, Mambrini & Passarotti 2023, Tittel 2024), semantischer Edition (Spadini et al. 2021) und multimodaler Wissensgraphen (Pianzola et al. 2023, Thoma et al. 2017, Zhang et al. 2016) angeknüpft werden. Eine vergleichbare Vielfalt von Daten wurde jedoch bisher nicht für die geisteswissenschaftliche Domäne erhoben, untersucht oder verknüpft, wohingegen es im Umfeld von Cultural Heritage bzw. GLAM bereits seit längerem Bemühungen gibt, das LOD-Potential zu nutzen (Hyvönen 2012, Hooland & Verborg 2015, Faraj & Micsik 2021, Liu et al. 2023). Zwei in-

tegrative Querschnittsprojekte befassen sich schließlich mit der Normdatenverknüpfung und den relevanten Infrastrukturlösungen (Diefenbach et al. 2021, Fischer 2022, Rossanova et al. 2023) sowie der domänenübergreifenden Modellierung geisteswissenschaftlichen Fachwissens und der Entwicklung von Strategien für die Vernetzung aller Daten verbunden mit dem Aufbau einer modularen Ontologie (Shimizu et al. 2023). Die einzelnen Module dokumentieren wiederkehrende Problemstellungen und strukturieren zentrale Klassen und Properties, die für alle Teilprojekte Relevanz haben, aber auch solche, die spezifisch für einzelne Subdomänen sind (wie bspw. die Text-Bild-Relationen in multimodalen Daten). Infrastruktur- und Datenmodellierungsentscheidungen hängen eng zusammen: Der Schwerpunkt für die Integration der Teilprojekte liegt auf Wikibase als Infrastruktur. Zugleich sind übergreifende Standards (z.B. SKOS) und damit die Interoperabilität außerhalb des ‚Wikiversums‘ maßgeblich. Ebenso wird eine Verschränkung mit Initiativen der NFDI (wie dem 4Culture-Knowledge Graphen) angestrebt.

Die gemeinsame Arbeit und der enge Austausch im Kontext des LODinG-Verbunds sind unabdingbar, um den beschriebenen Paradigmenwechsel zu bewirken. Sie erlauben uns, fachspezifische Expertise technischer, methodischer und inhaltlicher Art zusammenzuführen, in den unterschiedlichen Teilbereichen voneinander zu lernen und den Mehrwert des LOD-Paradigmas in den verschiedenen Forschungskontexten zu konkretisieren. Bestimmte Teilaspekte von LOD (wie Normdaten, Ontologie-Design, Abfragetechniken, Nutzungsszenarien, Verknüpfung mit anderen Wissensressourcen etc.) sind in verschiedenen Arbeitsbereichen unterschiedlich gut etabliert und können im Verbund enger aufeinander bezogen, kohärenter eingesetzt und miteinander vernetzt werden.

Bibliographie

- Alves, Daniel, ed.** 2022. „IJHAC: A Journal of Digital Humanities. Special Issue: Linked Open Data in the Arts and the Humanities” 16 (1). <https://www.eupublishing.com/doi/epdf/10.3366/ijhac.2022.0271>.
- Borek, Luise, Canan Hastik, Vera Khramova, Klaus Illmayer, and Jonathan D. Geiger.** 2021. „Information Organization and Access in Digital Humanities: TaDiRAH Revised, Formalized and FAIR.” In *Information between Data and Knowledge. Information Science and Its Neighbors from Data Science to Digital Humanities. Proceedings of the 16th International Symposium of Information Science (ISI 2021)*. Universität Regensburg. <https://doi.org/10.5283/EPUB.44951>.
- Brown, Susan.** 2022. „Same Difference: Identity and Diversity in Linked Open Cultural Data.” *International Journal of Humanities and Arts Computing* 16 (1): 1–16. <https://doi.org/10.3366/ijhac.2022.0273>.
- Chiarcos, Christian, Christian Fäth, and Maxim Ionov.** 2020. „The ACoLi Dictionary Graph.” In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, edited by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, et al., 3281–90. Marseille, France: European Language Resources Association. <https://aclanthology.org/2020.lrec-1.401>.
- Diefenbach, Dennis, Max De Wilde, and Samantha Alipio.** 2021. „Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph.” In *The Semantic Web – ISWC 2021*, edited by Andreas Hotho, Eva Blomqvist, Stefan Dietze, Achille Fokoue, Ying Ding, Payam Barnaghi, Armin Haller, Mauro Dragoni, and Harith Alani, 631–47. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-88361-4_37.
- Faraj, Ghazal, and András Micsik.** 2021. „Representing and Validating Cultural Heritage Knowledge Graphs in CIDOC-CRM Ontology.” *Future Internet* 13 (11): 277. <https://doi.org/10.3390/fi13110277>.
- Fischer, Barbara.** 2022. „Towards an Open and Collaborative Authority Control.” *JLIS.It* 13 (1): 283–90. <https://doi.org/10.4403/jlis.it-12767>.
- Franc, Yann Le, Luiz Bonino, Hanna Koivula, Jessica Parland-Von Essen, and Robert Pergl.** 2022. „D2.8 FAIR Semantics Recommendations Thrd Iteration,” February. <https://doi.org/10.5281/zenodo.6675295>.
- Goldhahn, Dirk, Uwe Kretschmer, Peter Muehleder, Franziska Naether, Anja Becker, and Cecilia Graiff.** 2024. „PUDEL: Paving the Way for Pawsome Data Models and Vocabularies in the Academic Community,” February. <https://doi.org/10.5281/zenodo.10698403>.
- Haraway, Donna.** 1988. „Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective.” *Feminist Studies* 14 (3): 575. <https://doi.org/10.2307/3178066>.
- Harpring, Patricia.** 2010. *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*. First edition. Los Angeles, California: The Getty Research Institute.
- Hinzmann, Maria, Matthias Bremm, Tinghui Duan, Anne Klee, Johanna Konstanciak, Julia Röttgermann, Christof Schöch, and Joëlle Weis.** 2024. „Patterns in Modeling and Querying a Knowledge Graph for Literary History [Preprint].” Zenodo. <https://doi.org/10.5281/zenodo.12080340>.
- Hooland, Seth van, and Ruben Verborgh.** 2015. *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. 1st ed. Facet. <https://doi.org/10.29085/9781783300389>.
- Hyvönen, Eero.** 2012. *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*. Synthesis Lectures on the Semantic Web: Theory and Technology 3. Palo Alto, Calif.: Morgan & Claypool.
- Ikonić Nešić, Milica, Ranka Stanković, Christof Schöch, and Mihailo Skoric.** 2022. „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, France: European Language Resources Association. <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. 2022. “When Linguistics Meets Web Technologies. Recent Advances in Modelling Linguistic Linked Data.” Edited by Philipp Cimiano, Julia Bosque-Gil, Philipp Cimiano, and Milan Dojchinovski. *Semantic Web* 13 (6): 987–1050. <https://doi.org/10.3233/SW-222859>.

Liebing, Katja. 2024. “Register Historischer Und Objektbezogener Vokabulare Und Normdaten Mit Workshop Erfolgreich Gestartet.” *Data Connectivity – NFDI4Memory* (blog). April 26, 2024. <https://blogs.urz.uni-halle.de/nfdi4memory/2024/04/register-historischer-und-objektbezogener-vokabulare-und-normdaten-mit-workshop-erfolgreich-gestartet/>.

Lindemann, David, Penny Labropoulou, and Christiane Klaes. 2022. “Introducing LexMeta: A Metadata Model for Lexical Resources.” Application/pdf. In *Dictionaries and Society. Proceedings of the XX EURALEX International Congress*, 12–16 July 2022, Mannheim, Germany, 310–20. IDS-Verlag. <https://doi.org/10.14618/IDS-PUB-11187>.

Liu, F., J. Hindmarch, and M. Hess. 2023. “A Review of the Cultural Heritage Linked Open Data Ontologies and Models.” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLVIII-M-2–2023* (June):943–50. <https://doi.org/10.5194/isprs-archives-XLVIII-M-2-2023-943-2023>.

Mambrini, Francesco, and Marco Carlo Passarotti. 2023. “The LiLa Lemma Bank: A Knowledge Base of Latin Canonical Forms.” *Journal of Open Humanities Data* 9 (November):28. <https://doi.org/10.5334/johd.145>.

Pianzola, Federico, Xiaoyan Yang, Noa Visser, Michiel van der Ree, and Andreas van Cranenburgh. 2023. “Constructing the GOLEM: Graphs and Ontologies for Literary Evolution Models.” In *A DHO Digital Humanities 2023 Conference* (DH2023). Zenodo. <https://doi.org/10.5281/ZENODO.8206543>.

Rossenova, Lozana, Paul Duchesne, and Ina Blümel. 2023. “Wikidata and Wikibase as Complementary Research Data Management Services for Cultural Heritage Data.” *Proceedings of the 3rd Wikidata Workshop 2022 Co-Located with the 21st International Semantic Web Conference* (ISWC2022), 15. <https://doi.org/10.25968/OPUS-2573>.

Santini, Cristian, Nele Garay, Etienne Posthumus, and Harald Sack. 2024. “The Art of Relations.” In *DHd 2024 Quo Vadis DH* (DHd2024), Passau, Deutschland. Zenodo. <https://doi.org/10.5281/ZENODO.10698245>.

Shimizu, Cogan, Karl Hammar, and Pascal Hitzler. 2023. “Modular Ontology Modeling.” Edited by Sabrina Kirrane, Axel-Cyrille Ngonga Ngomo, Sabrina Kirrane,

and Axel-Cyrille Ngonga Ngomo. *Semantic Web* 14 (3): 459–89. <https://doi.org/10.3233/SW-222886>.

Spadini, Elena, Francesca Tomasi, and Georg Vogeler, eds. 2021. *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing. Schriften Des Instituts Für Dokumentologie Und Editorik*, Band 15. Norderstedt: BoD – Books on Demand.

Steller, Jonatan Jalle, Linnaea Charlotte Söhn, Julia Tolksdorf, Oleksandra Bruns, Tabea Tietz, Etienne Posthumus, Heike Fliegl, Sarah Pittroff, Harald Sack, and Torsten Schrade. 2024. “Communities, Harvesting, and CGIF: Building the Research Data Graph at NFDI4Culture.” In *DHd 2024 Quo Vadis DH* (DHd2024), Passau, Deutschland. Zenodo. <https://doi.org/10.5281/ZENODO.10698300>.

Thoma, Steffen, Achim Rettinger, and Fabian Both. 2017. “Towards Holistic Concept Representations: Embedding Relational Knowledge, Visual Attributes, and Distributional Word Semantics.” In *The Semantic Web – ISWC 2017*, edited by Claudia d’Amato, Miriam Fernandez, Valentina Tamma, Freddy Lecue, Philippe Cudré-Mauroux, Juan Sequeda, Christoph Lange, and Jeff Heflin, 10587:694–710. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-68288-4_41.

Thornton, Katherine, Kenneth Seals-Nutt, Marianne Van Remoortel, Julie M. Birkholz, and Pieterjan De Potter. 2022. “Linking Women Editors of Periodicals to the Wikidata Knowledge Graph.” Edited by Mehwish Alam, Victor de Boer, Enrico Daga, Marieke van Erp, Lydia Pintscher, Eero Hyvönen, Albert Meroño Peñuela, et al. *Semantic Web* 14 (2): 443–55. <https://doi.org/10.3233/SW-222845>.

Tittel, Sabine. 2024. *Historische Lexikalische Semantik Und Linked Data: Modellierung von Ressourcen Der Mittelalterlichen Galloromania Für Das Semantic Web*. De Gruyter. <https://doi.org/10.1515/978311326481>.

Zaytseva, Ksenia, and Matej Ďurčo. 2020. “Controlled Vocabularies and SKOS. Version 1.1.0.” Edited by Matej Ďurčo and Tanja Wissik. DARIAH-Campus. [Training module]. <https://campus.dariah.eu/id/D8d6OrLdpLIGRqBSQDVN0>.

Zhang, Lei, Achim Rettinger, and Ji Zhang. 2016. “A Knowledge Base Approach to Cross-Lingual Keyword Query Interpretation.” In *The Semantic Web – ISWC 2016*, edited by Paul Groth, Elena Simperl, Alasdair Gray, Marta Sabou, Markus Krötzsch, Freddy Lecue, Fabian Flöck, and Yolanda Gil, 9981:615–31. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-46523-4_37.

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