LODinG: Linked Open Data in the Humanities

Jacek Kudera, Claudia Bamberg, Thomas Burch, Folke Gernert, Maria Hinzmann, Susanne Kabatnik, Claudine Moulin, Benjamin Raue, Achim Rettinger, Jörg Röpke, Ralf Schenkel, Kristin Shi-Kupfer, Doris Schirra, Christof Schöch, Joëlle Weis

Trier University, Am Universitätsring 15, 54286 Trier, Germany
{kudera, bamberg, burch, gernert, hinzmannm, kabatnik, moulin, raue,
rettinger, roepke, schenkel, shikupfer, schirra, schoech, weis}@uni-trier.de

Abstract

We are presenting *LODinG – Linked Open Data in the Humanities* (abbreviated from *Linked Open Data in den Geisteswissenschaften*), a recently launched research initiative exploring the intersection of Linked Open Data (LOD) and a range of areas of work within the Humanities. We focus on effective methods of collecting, modeling, linking, releasing and analyzing machine-readable information relevant to (digital) humanities research in the form of LOD. LODinG combines the sources and methods of digital humanities, general and computational linguistics, digital lexicography, German and Romance philology, Sinology, translatology, cultural and literary studies, media studies, information science and law to explore and expand the potential of the LOD paradigm for such a diverse and multidisciplinary field. The project's primary objectives are to improve the methods of extracting, modeling and analyzing multilingual data in the LOD paradigm; to demonstrate the application of the linguistic LOD to various methods and domains within and beyond the humanities; and to develop a modular, cross-domain data model for the humanities.

Keywords: Linked Open Data, Humanities, Digital Humanities, Knowledge Graphs

1. Background

The potential of implementing the Linked Open Data (LOD) paradigm in the field of (Digital) Humanities is immense and has already been discovered by many scholars (Zhao, 2023). The interconnectedness of data from different, even seemingly unrelated disciplines has already allowed for a more comprehensive description of linguistic, cultural, sociological, and/or historical phenomena, often providing previously unnoticed contexts. A key finding of (Zhao, 2023) is that fields such as linguistics or actors like libraries have been producing new LOD resources for some time; however, projects emerging from other areas of the humanities primarily use existing LOD resources to uniquely identify and disambiguate entities relevant to their domain, but only rarely produce substantial new LOD resources themselves.

Despite the availability of conceptual reference models (such as CIDOC-CRM; (Faraj and Micsik, 2021)), ontology representation frameworks (e.g., OWL) and Semantic Web technologies (e.g., RDF; (Hitzler, 2021)), different disciplines in the humanities develop independent ways of categorizing entities. In the humanities, the domain-specific terminology often circulates within a particular area of research and rarely takes advantage of conceptual interlinking of uniquely-identified items. Descriptive studies that refrain from placing their entities within the structures of a formalized ontology significantly reduce their interdisciplinary potential and leave the opportunities that lie in knowledge networks undiscovered. Furthermore, the lack of presence of linguistic LOD (LLOD) across the disciplines canonically associated with the humanities limits the existing data to digitally-structured sources only and renounces the methods that can benefit from the exploration of robust knowledge graphs (KGs).

Relevant work specifically at the intersection of literary studies and LOD is emerging recently, such as the *GOLEM* project (Graphs and Ontologies for Literary Evolution Models) at Groningen University (Pianzola et al., 2023) or the *MEDIATE* project at Radboud University (Montoya, 2021). Similarly, initiatives such as the one aiming to 'lodify' the *European Literary Text Collection* (ELTeC) Odebrecht et al. (2021) are building bridges between linguistics and literary studies (Ikonić Nešić et al., 2022; Schöch et al., 2021). LODinG, however, takes up and expands on the questions opened by its preceding research initiative, i.e., *MiMoText – Mining and Modeling Text* (2019-2023) hosted by Trier University and coordinated by the Trier Center for Digital Humanities (TCDH). LODinG's predecessor has developed a KG for the domain of the French novel of the Enlightenment, the *MiMoTextBase*. The project team used computational methods to extract information from a wide range of sources – from bibliographic resources and primary texts from the 18th century to current research literature (Schöch et al., 2022). The information ranges from bibliographic data (such as places and dates of publication) to book formats, themes, narrative locations, protagonists and sentiment trajectories or stylistic similarities between texts. The LOD paradigm allows this heterogeneous information to be linked to form a common body of knowledge. Its contents are

LDL: Linked Data in Linguistics @ LREC-COLING 2024, pages 49–54

formally modeled and linked to each other. In addition, the extracted information is also linked to external knowledge resources such as Wikidata. The numerous query options that this allows – including federated queries originating from both *MiMoTextBase* and *Wikidata* – open up entirely new perspectives on both well-known and lesser-known literary-historical knowledge. With LODinG, we can now extend this paradigm to a much wider range of domains within and beyond the humanities.

Against this background, the presented project *LODinG* – initiated by a group of researchers at Trier University, Germany, and coordinated by the Trier Center for Digital Humanities — aims to explore the potential of the LOD paradigm at the intersection of qualitative and quantitative studies in the humanities. The project seeks to enrich the methods of annotation and information extraction applied to domain data relevant to a range of fields in the humanities. The initiative is currently exploring the potential of bridging multiple semi-structured datasets using formally-modeled, domain-adapted and modular ontologies and taxonomies pertinent to literary studies, linguistics, digital lexicography, scholarly editing, media studies, scientometrics and law. The LOD paradigm is a cornerstone of innovation in (digital) humanities. It enables the linking of multidisciplinary data using a coherent ontological classification and interoperable formats. The project aims to promote the interdisciplinary and transparent research supported by state-of-the-art data management infrastructure driven by knowledge networks. Overall it aims to bring a new quality of interdisciplinary reasoning to the area of data science and the humanities.

2. Objectives

The presented project emphasizes an interdisciplinary approach to building a modular ontology using LOD. It combines the methods commonly used to build, explore and query knowledge networks with the apparatus traditionally employed in Natural Language Processing and Information Retrieval. Furthermore, LODinG aims not only to explore the potential of existing digital resources in the context of LOD but also to generate, publish and integrate new resources. Finally, the project aims to demonstrate the potential of linguistic LOD for innovative research endeavors in the humanities. To further present the potential of LOD in interdisciplinary research, LODinG bridges multimodal and multilingual areas of study, including Romance studies, German studies, Sinology and law, and demonstrates the possibilities arising from computing multimodal KGs embedded in linguistic, literary, cultural and legal contexts.

The LODinG project conceptualizes and deploys the knowledge networks that enable querying, statistical analysis, data visualization, and linking to open datasets via formal modeling of entities and properties and the application of a modular, human-generated ontology. At the conceptual level, we use named entity recognition and other information retrieval tasks to provide entities (such as people, places, organizations, motifs, methods, works or themes, etc.) with unique identifiers and labels that allow us to build robust linked knowledge networks. Furthermore, with the application of LOD, the already existing criteria for classifying the entities and creating typologies can be easily inspected to reconsider their analytical value. Such an approach helps to strike a balance between typologies driven by too many categories, resulting in overly specific information, on the one hand, and too few categories, carrying the risk of deriving inaccurate generalizations, on the other.

3. Areas of activity

LODinG is organized in closely interlinked areas of activity, each focusing on different aspects of linguistic LOD (LLOD). In the scope of LODinG, we will explore the enormous potential of LOD for innovative research in the humanities with a focus on linguistics, law, as well as literary, cultural and media studies.

The first research area focuses on the lexical level of the language system and examines the neologisms coined as a result of the Covid pandemic crisis. This work bridges the fields of German lexicology and digital lexicography. In addition to studying recent lexical phenomena, this subproject introduces a diachronic perspective by taking into account historical lexicons rich in pandemic-related vocabulary (Zacherl, 2022). This sub-project applies LOD and LLOD methods based on the *semantics by reference* (McCrae et al., 2012; Cimiano et al., 2020) framework. The research agenda of this subproject also includes the exploration of the semantic domain related to infections and diseases through the prism of standard Semantic Web data (Wandl-Vogt and Declerck, 2014). The diachronic perspective on established synsets would enable tracing of lexical changes of the analyzed set of lexemes as well as multiword expressions semantically related to infectious diseases. Furthermore, this subproject of LODinG emphasizes the importance of the representation of dictionary entries in the linguistic LOD framework and provides support for the integration of lexicons into the Semantic Web (Passarotti et al., 2020; Khan et al., 2022; Lindemann et al., 2022) .

The second area of research in LODinG focuses on the terminology frequently used in historical medical and botanical sources from the early modern

period of Romance and Germanic texts. In addition to linguistic LOD, this subproject is based on the methodological apparatus of translatology and scholarly editing in a LOD context (Spadini et al., 2021). This work package aims to compare historical sources containing standardized botanical *nomina propria* from different languages. The comparison will be done through direct source-target translation and interlanguage interference. The use of the latter approach may reveal the influence of an intermediate language that is typologically and phylogenetically distant the from source and target languages on equivalent matching in translation. Several historical sources dating back to the 16th century have already been digitized in the preparatory stage for this analysis (Moulin, 2018). The use of supervised OCR and LLOD methods will enable the triangulation of quantitative and qualitative analyses planned for this project.

The third area of work combines the methods of digital humanities and computer science. This subproject focuses on extracting statements related e.g. to datasets, methods and results, utilizing, in part, the OpenAlex resource. Currently, the production of scientific works far exceeds the reading capacity of researchers and research teams. Traditional indexing solutions, such as keywords, abridged abstracts and reviews often fail to address central questions of publications. Algorithmic, scalable methods of information extraction and synthesis are becoming increasingly important, supporting *semantic publishing* (Shotton, 2009; Schöch, 2021; Verma et al., 2023). To address the question of the lack of so-called *semantic statements*, this subproject aims to employ the LLOD methods in combination with manual tagging and machine content retrieval to semantically annotate a collection of works spanning across the field of the humanities. The subproject focuses on transforming abstracts and keywords into a limited number of machine-readable LOD statements (Metzger et al., 2011). This subproject partly departs from the OpenAlex platform, that contains metadata and LOD statements, but so far is lacking a systematic domain-dependent content modelling and a solid architecture of a formal ontology. The identified area for improvement and LODinG's involvement is to supplement the available data with content analyses and provide additional domain-specific context.

The fourth area of work builds upon the previous subproject and combines sinology with computer science. It focuses on scientific literature published in Modern Standard Chinese and converts its findings into machine-readable synthetic LOD statements. This work will compare the performance of entity extraction methods excerpted from the source language on available non-Chinese sources. The analyses of information extraction will be conducted from a cross-linguistic perspective. The project aims to develop language-specific tools for information extraction and synthesis. The goal is to provide non-Chinese readers with a toolkit to discover Chinese-language scholarly literature based on linguistic LOD. This approach provides an alternative to common machine translation solutions, which often lack high-quality training data that involves matching specialized terminology across multiple languages.

The fifth subproject introduces the synergy between cultural studies and Natural Language Processing. The material of focus, namely wine labels, consists of items that are constrained to a specific domain and often combine text and image. The coherence between the text and image on wine labels varies, sometimes resulting in the juxtaposition of opaque concepts that do not clearly correspond with one another. This is a good starting point for investigating the potential of combining text and image analyses to build more robust KGs. The subproject employs textual content and images from wine labels to develop robust multimodal knowledge representation networks. Today, some Large Language Models (LLMs) are trained using multimodal data that combine lexical input and images (Wu et al., 2023; Zhang et al., 2024). This means they support multimodal indexing processes, where text and image recognition can benefit from each other. The challenge is to harness the power of generative LLMs to make them sufficiently robust and predictable to create standards-based LOD. Thus, the objective of this subproject is to investigate the potential of this method for indexing collections of wine labels scraped from the web. The subproject aims to create a more generalizable process by starting with in-domain source material that could then be extended to other types of sources such as postcards, geographical maps, advertisements, or book illustrations.

The sixth area of work focuses on the conceptual indexing of multilingual European texts. The primary goal of this subproject is to develop a multilingual parallel corpus of European legal texts thematically related to digitization processes, datasets and digital data processing (such as Digital Services Act, Regulation EU 2022/2065). This work package aims to identify differences in legally-binding terminology among all official EU languages. Additionally, using LOD, the subproject will develop methods for transparent equivalent matching, supported by conceptual indexing. To accomplish these objectives, this work package includes the following consecutive steps: automatic sentence-level alignment of legal texts; identification of key jurisprudential concepts and their integration into the LOD framework; multilingual annotation (both manual and automatic) of the identified concepts; and concept-

driven search to identify mismatches across selected languages. This workflow is based on a set of similar tasks involving a multilingual corpus that comprises translations of an 18th-century legal text (Bretschneider et al., 2020). Such an approach enables the identification of contextual or conceptual discrepancies across multiple languages, which is particularly important in legal texts.

The seventh area of work in LODinG exhibits a cross-sectional character. Its aim is to integrate the standardized entries of all above-mentioned work areas into a modular ontology that supports federated queries (Shimizu et al., 2023, 2022; Cimiano et al., 2020). This will be accomplished by using established techniques for constructing ontology and metadata structures, glossing, cataloging, employing semantic networks, and comparing taxonomies (Borek et al., 2021). We plan to contribute, wherever possible, to recent and emerging initiatives that strengthen the alignment of vocabularies and KGs and the potential of federation (Steller et al., 2024). The subproject aims to integrate both domain-specific and cross-domain general elements. Currently, only a limited set of predicates are being used across domains, such as person, place, publication, discipline, century, country, or continent. To bridge domain- and discipline-specific entities, a larger set of predicates will be implemented including methods, procedures, epochs, subdomains, phenomena etc. (Bodard, 2021; Burrows and Nurmikko-Fuller, 2020). We aim to employ Wikidata identifiers along with other available authority file data allowing for the enhancement of KGs. Striving for a balance between a project-specific micro-perspective and an overarching macro-perspective, the objective is twofold: (1) to utilize and develop domain-specific resources to generate new research perspectives, and (2) to support and promote overarching integration in the LOD paradigm enabling cross-disciplinary and cross-domain linking and reuse of information (Brown, 2022; Santini et al., 2024).

The final work area focuses on developing technical solutions necessary to achieve the project's goals. This cross-cutting area aims to create an environment that promotes interoperability of data curated in the above-mentioned subprojects. This work area also aims to ensure the quality of research data management strategies used by LODinG. Furthermore, this infrastructural project aims to create interfaces for unstructured data, enabling non-standard formats to adapt to the LOD framework. The tools developed within the scope of this work package will allow for data annotation using standardized classifiers and facilitating interlinking between disciplines traditionally associated with the research area of the humanities.

A further goal of this work area is to examine the intersections between information extraction, ontology design, KG engineering and artificial intelligence. We plan to host several local Wikibase instances and to anchor LODinG in the *Wikibase ecosystem* (Diefenbach et al., 2021; Faulhaber, 2022; Simons, 2023; Rossenova, Lozana et al., 2023). Through our membership in the Wikibase Stakeholder Group, we aim to further contribute to this environment and develop its workflows and tool chains, for example, to semi-automatically convert exports from the semantic annotation tool INCEpTION or manual annotations from the virtual research environment FuD into LOD statements (Klie et al., 2018; Bamberg et al., 2023).

4. Anticipated Results

The LODinG initiative is still in its early stages, having been launched only at the beginning of 2024. However, we can outline several areas in which we anticipate results and outcomes. These areas pertain to domain-specific results from various work areas, capacity building and networking. Additionally, a modular cross-domain ontology for the humanities is proposed. LODinG aims to research, design, and publish interconnections between various standards, practices, knowledge domains, tools, and models with different focal points. The research will introduce and exemplify new applications of LOD within and across various disciplines in the humanities, while also rethinking LOD as a research paradigm. The LOD framework is expected to be suitable for both qualitative (hermeneutic and contextualizing) and quantitative (algorithmic and statistical) research, making it attractive to scholars from various areas of the humanities.

LODinG focuses on using LOD for multilingual and multimodal resources. The project aims to enrich a set of indexed entities involving several languages, addressing the dominance of English in models and tools for information extraction. Furthermore, LODinG has potential in multimodal applications of LOD, such as combining lexical information with images.

The experiences from the undertaken tasks will be documented extensively to provide guidance for other projects or published as best practices guidelines. The presented project offers an interdisciplinary platform to explore the effectiveness and necessity of domain-specific solutions, as well as the potential for holistic linguistic LOD infrastructures.

The novelty of LODinG lies in combining quantitative and qualitative research methods from various fields in the humanities with the LOD paradigm. We believe that such a multidisciplinary orientation of our research project has a potential to open a new chapter in digital humanities.

5. Bibliographical References

Claudia Bamberg, Cornelia Bögel, Thomas Bürger, Thomas Burch, Ruth Golyschkin, Bianca Müller, Radoslav Petkov, Thomas Stern, Jochen Strobel, and Olivia Varwig. 2023. Digitale Edition der Korrespondenz August Wilhelm Schlegels – Ergebnisse, Erfahrungen, Entwicklungen. *Zeitschrift für digitale Geisteswissenschaften (ZfdG)*.

Gabriel Bodard. 2021. Linked Open Data for Ancient Names and People. *ISAW Papers*, 20(4).

Luise Borek, 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 (ISI 2021)*. Universität Regensburg.

Falk Bretschneider, Rainer Maria Kiesow, Claudine Moulin, and Christof Schöch. 2020. Les mots de Beccaria. Métalexicographie des langues du droit à partir de Dei delitti e delle pene (1764) et ses traductions en Europe (Metalex). *Beccaria. Revue d'histoire du droit de punir*, 5:117–125.

Susan Brown. 2022. Same Difference: Identity and Diversity in Linked Open Cultural Data. *International Journal of Humanities and Arts Computing*, 16(1):1–16.

Simon Burrows and Terhi Nurmikko-Fuller. 2020. Charting Cultural History through Historical Bibliometric Research. Methods, concepts, challenges, results. In *Routledge International Handbook of Research Methods in Digital Humanities*. Routledge.

Philipp Cimiano, Christian Chiarcos, John P. McCrae, and Jorge Gracia. 2020. Linguistic Linked Open Data Cloud. In Philipp Cimiano, Christian Chiarcos, John P. McCrae, and Jorge Gracia, editors, *Linguistic Linked Data: Representation, Generation and Applications*, pages 29–41. Springer, Cham.

Dennis Diefenbach, Max De Wilde, and Samantha Alipio. 2021. Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph. In *The Semantic Web – ISWC 2021*, Lecture Notes in Computer Science, pages 631–647, Cham. Springer.

Ghazal Faraj and András Micsik. 2021. Representing and Validating Cultural Heritage Knowledge Graphs in CIDOC-CRM Ontology. *Future Internet*, 13(11):277.

Charles B. Faulhaber. 2022. PhiloBiblon y el mundo wiki. *Magnificat Cultura i Literatura Medievals*, 9:203.

Pascal Hitzler. 2021. A review of the semantic web field. *Communications of the ACM*, 64(2):76–83.

Milica Ikonić Nešić, 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*, pages 7–16, Marseille, France. ELRA.

Anas Fahad Khan, Christian Chiarcos, Thierry Declerck, Daniela Gifu, Elena González-Blanco García, Jorge Gracia, Maxim Ionov, Penny Labropoulou, Francesco Mambrini, John P. McCrae, Émilie Pagé-Perron, Marco Passarotti, Salvador Ros Muñoz, and Ciprian-Octavian Truică. 2022. When linguistics meets web technologies. Recent advances in modelling linguistic linked data. *Semantic Web*, 13(6):987–1050.

Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. ACL.

David Lindemann, Penny Labropoulou, and Christiane Klaes. 2022. Introducing LexMeta: a metadata model for lexical resources. In *Dictionaries and Society. (EURALEX 2022)*. IDS-Verlag.

John McCrae, Guadalupe Aguado-de Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, et al. 2012. Interchanging lexical resources on the semantic web. *Language Resources and Evaluation*, 46:701–719.

Steffen Metzger, Shady Elbassuoni, Katja Hose, and Ralf Schenkel. 2011. S3K: seeking statement-supporting top-k witnesses. In *Proceedings of CIKM 2011*, pages 37–46. ACM.

Alicia C. Montoya. 2021. The Shark in the Library: Books and Non-book Artifacts in Private Library Auction Catalogues, 1665–1830. In *Figurations animalières à travers les textes et l'image en Europe*, pages 249–265. Brill.

Claudine Moulin. 2018. Textwandlungen – Eucharius Rösslin, Der Swangern Frauwen und hebammen Rosegarten als sprachhistorische Quelle. In Christina Waldvogel Luise Czajkowski,

Sabrina Ulbrich-Bösch, editor, *Sprachwandel im Deutschen*, pages 319–336. De Gruyter, Berlin/Boston.

Marco Carlo Passarotti, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. 2020. Interlinking through Lemmas. The Lexical Collection of the LiLa Knowledge Base of Linguistic Resources for Latin. *Studi e Saggi Linguistici*, 58(1):177–212.

Federico Pianzola, Xiaoyan Yang, Noa Visser, Michiel van der Ree, and Andreas van Cranenburgh. 2023. Constructing the GOLEM: Graphs and Ontologies for Literary Evolution Models. In *Digital Humanities Conference (DH2023)*.

Rossenova, Lozana, Duchesne, Paul, and Blümel, Ina. 2023. Wikidata and Wikibase as complementary research data management services for cultural heritage data. In *Proceedings of the 3rd Wikidata Workshop 2022 co-located with ISWC2022*.

Cristian Santini, Nele Garay, Etienne Posthumus, and Harald Sack. 2024. The Art of Relations. In *DHd 2024*.

Christof Schöch. 2021. Open Access für die Maschinen. In Maria Effinger and Hubertus Kohle, editors, *Die Zukunft des kunsthistorischen Publizierens*. arthistoricum.net-ART-Books.

Christof Schöch, Maria Hinzmann, Julia Röttgermann, Katharina Dietz, and Anne Klee. 2022. Smart Modelling for Literary History. *International Journal of Humanities and Arts Computing*, 16(1):78–93.

Christof Schöch, Roxana Patras, Tomaž Erjavec, and Diana Santos. 2021. Creating the european literary text collection (eltec): Challenges and perspectives. *Modern Languages Open*, 1.

Cogan Shimizu, Andrew Eells, Seila Gonzalez, Lu Zhou, Pascal Hitzler, Alicia Sheill, Catherine Foley, and Dean Rehberger. 2022. Ontology Design Facilitating Wikibase Integration – and a Worked Example for Historical Data.

Cogan Shimizu, Karl Hammar, and Pascal Hitzler. 2023. Modular ontology modeling. *Semantic Web*, 14(3):459–489.

David Shotton. 2009. Semantic publishing: the coming revolution in scientific journal publishing. *Learned Publishing*, 22(2):85–94.

Olaf Simons. 2023. Keine Selbstverständlichkeit: Citizen Science auf der FactGrid Wikibase-Plattform. In René Smolarski, Hendrikje Carius, and Martin Prell, editors, *Citizen Science in den Geschichtswissenschaften*. V&R unipress, Göttingen.

Elena Spadini, Francesca Tomasi, and Georg Vogeler, editors. 2021. *Graph data-models and semantic web technologies in scholarly digital editing*. Number Band 15 in Schriften des Instituts für Dokumentologie und Editorik. BoD – Books on Demand, Norderstedt.

Jonatan Jalle Steller, 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*.

Shilpa Verma, Rajesh Bhatia, Sandeep Harit, and Sanjay Batish. 2023. Scholarly knowledge graphs through structuring scholarly communication: a review. *Complex & Intelligent Systems*, 9(1):1059–1095.

Eveline Wandl-Vogt and Thierry Declerck. 2014. Cross-linking Austrian dialectal dictionaries through formalized meanings. In *Proceedings of EURALEX XVI*.

Jiayang Wu, Wensheng Gan, Zefeng Chen, Shicheng Wan, and Philip S. Yu. 2023. Multimodal large language models: A survey. In *2023 IEEE International Conference on Big Data (BigData)*.

Florian Zacherl. 2022. Digitale Tiefenerschließung traditioneller Lexikographie – am Beispiel des Romanischen Etymologischen Wörterbuchs.

Duzhen Zhang, Yahan Yu, Chenxing Li, Jiahua Dong, Dan Su, Chenhui Chu, and Dong Yu. 2024. Mm-llms: Recent advances in multimodal large language models.

Fudie Zhao. 2023. A systematic review of Wikidata in Digital Humanities projects. *Digital Scholarship in the Humanities*, 38(2):852–874.

6. Language Resource References

Odebrecht, Carolin and Burnard, Lou and Schöch, Christof. 2021. *European Literary Text Collection (ELTeC)*. COST Action 16204, 1.1.0.