Editors (Hrsg.): INFORMATIK 2023 – Workshop on Research Data Infrastructures in DS and AI,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2023 1

Research Knowledge Graphs in NFDI4DS

Saurav Karmakar[1] Matthäus Zloch[1] Fidan Limani[2] Benjamin Zapilko[1] Sharmila Upadhyaya[1] Jennifer D'Souza[3] Leyla Jael Castro [4] Georg Rehm[5] Marcel R. Ackermann[6] Harald Sack[7] Zeyd Boukhers[8] Sonja Schimmler[9] Danilo Dessí[1] Peter Mutschke[1] Stefan Dietze[1 10]

Abstract: The ever-increasing amount of research output through scientific articles requires means to enable transparency and a better understanding of key entities of the research lifecycle, referred to as research artifacts, such as methods, software, datasets, etc. Research Knowledge Graphs (RKG) make research artifacts findable, accessible, interoperable, and reusable (FAIR) and facilitate their interpretability. In this article, we describe the role of RKGs, from their construction to the expected benefits, including an overview and a vision of their use within the German National Research Data Infrastructure (NFDI) consortium NFDI4DataScience (NFDI4DS). This paper includes insights about the existing RKGs, how to formally represent research artifacts, and how this supports better transparency and reproducibility in data science and artificial intelligence. We also discuss key challenges, such as RKG construction, and integration, and give an outlook on future work.

Keywords: NFDI; NFDI4DS; Data Science; Research Knowledge Graph; Scholarly Data; Knowledge Graph Integration; Knowledge Graph Federation

1 Introduction

Scientific results are typically disseminated through unstructured research publications that describe problems, outline methodologies, present contributions, and discuss experimental results. However, processing information from unstructured documents is costly and requires substantial efforts, making it challenging to actually understand the state-of-the-art in a respective field and the relations and dependencies between tasks, datasets, methods, machine learning models, or overall performance. This problem has been elevated by the shift towards increasingly data-driven and deep learning-based systems, which are known to suffer from dramatic reproducibility issues and often do not outperform simple baselines despite being considered state-of-the-art approaches [DCJ19]. Behind these significant challenges in the field of data science and AI is a lack of transparency about the inherent

[1] GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany saurav.karmakar@gesis.org
[2] ZBW – Leibniz Information Centre for Economics, Kiel, Germany
[5] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Berlin, Germany
[3] TIB – Leibniz Information Centre for Science and Technology, Hannover, Germany
[4] ZB MED Information Centre for Life Sciences, Cologne, Germany
[6] DBLP computer science bibliography, Schloss Dagstuhl - Leibniz Center for Informatics, Trier, Germany
[7] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Karlsruhe, Germany
[8] Fraunhofer Institute for Applied Information Technology -FIT-, Sankt Augustin, Germany
[9] Fraunhofer Institute for Open Communication Systems -FOKUS-, Berlin, Germany
[10] Heinrich Heine University Düsseldorf, Germany

dependencies of data, models and code as being used in increasingly complex data processing pipelines.

The current practices as part of the scientific process still involve substantial efforts in searching, browsing and parsing unstructured documents to find relevant resources and insights, where the increasing amount of scholarly literature only increases required efforts. This has further increased the need to represent, store, and document research outcomes, including intricate and complex relationships among research questions, models, data, results, and all-produced research artifacts. Making research outcomes available implies making the entire research lifecycle transparent and all relevant research artifacts findable, accessible, interoperable, and re-useable in the best sense of the FAIR Data Principles [Wi16, Ba22]. While some solutions suggest prompting authors to annotate certain research entities (e.g., contributions) during the creation of their articles [BBK23], the variable nature of manuscripts pre-publication leaves no assurance that authors will consistently and diligently annotate these entities. Therefore, novel and scalable means for extracting and representing digital research artifacts are crucial for the preservation, exploration, use, and reuse of research.

Knowledge graphs (KGs) are being widely used to represent data in a machine-actionable fashion following established principles, such as the FAIR principles and W3C standards, and are particularly focused on capturing relations between entities, usually expressed as RDF triples. KGs have been used already to represent and manage information in scientific contexts, where examples include Nanopublications[1] [GGV10], SoftwareKG [SZK20], Springer SciGraph[2], Computer Science KG [De22b] and the Open Research Knowledge Graph (ORKG) [Au20, Ja19]. We do refer to such applications of KGs for research data and information management as *Research Knowledge Graphs* (RKG).

Building on these foundations, the NFDI4DataScience (NFDI4DS)[3] consortium, part of the German National Research Data Infrastructure (NFDI) initiative, aims at addressing the aforementioned challenges with respect to reproducibility and transparency in data science and AI through an integrated RKG infrastructure. The goal is to establish an integrated infrastructure that provides efficient access to DS and AI resources, most notable, data, code, and machine learning models, and to provide transparent information about their interdependencies. This will be achieved through both building on established RKGs and constructing novel RKGs, e.g. through scholarly information extraction models applied to large corpora of data science and AI literature. This work gives an overview of current work and future directions within this context.

[1] https://nanopub.org/

[2] https://www.springernature.com/gp/researchers/scigraph

[3] https://www.nfdi4datascience.de/

2 Research Knowledge Graphs in NFDI4DS

RKGs allow for structuring relations between research artifacts and other entities (e.g., authors, organizations, etc.), making them machine-understandable and searchable. They are also a means toward re-usability and further machine-actionability of the structured information. RKGs are typically generated from different kinds of (meta)data sources. They can serve various use cases, such as search across a specific knowledge domain, understanding of entities dependencies, use of data within computer-based approaches, etc. Examples include searching for research datasets, understanding research methods, their use, and dependencies on other research artifacts (code, used datasets, etc.), as well as understanding scientific discourse. Thus, RKGs enable researchers to explore a research field from various facets by exploiting their inherent linkage structures.

Category	**RKG**	**Scope**
quality-controlled ground truths datasets	SoMeSci [Sc21]	software mentions in scientific articles, currently consisting of about 4.4 million triples, describing metadata and context of 3,756 mentions in 1,367 articles
scholarly resource metadata	GESIS KG Infrastructure[4]	tools for constructing RKGs of research information, metadata and primary research data at GESIS, i.e., the GESIS Search KG [Hi19]
scholarly resource metadata	DBLP KG [db23]	metadata about 6.7 million computer science publication entities and 3.2 million author entities, 362 million RDF triples in total (Jun. 2023)
primary research data	TweetsKB [Fa18]	metadata about 1.5 billion tweets (collected Feb. 2013 - Mar. 2018)
primary research data	TweetsCOV19 [Di20]	subset of TweetsKB containing COVID-related tweets reflecting the societal discourse about COVID-19 on Twitter (Oct 2019 - Apr 2020)
primary research data	ClaimsKG [Tc19, Ga23]	claims and their evaluation from fact-checking websites, currently holding about 28 thousand claims from six English-language websites
automatically (NLP-)generated resource relations	SoftwareKG [Sc22]	information about software mention statements from more than 51,000 scientific articles from the social sciences
automatically (NLP-)generated resource relations	CS-KG [De22a, De22c]	large-scale automatically generated knowledge graph composed by over 350 million RDF triples describing 41 million statements from 6.7 million articles about 10 million entities (tasks, methods, materials, metrics, etc.) linked by 179 semantic relations
community expression of scholarly papers	The Open Research Knowledge Graph [Au20, Ja19]	structured semantic descriptions of articles, crowdsourced from authors and researchers, incl comparisons of contributions, visualizations, leaderboards

Tab. 1: Overview of RKGs in NFDI4DS

The NFDI4DS RKGs aim at providing knowledge about research artifacts and their relationships in a well-structured representation. Next to providing individual research knowledge graphs, NFDI4DS will focus on integrating individual research artifacts (e.g., datasets) within each respective category and into a joint research knowledge graph infrastructure that preserves their identity and provenance. A particular focus is on sharing data as well as methods and infrastructure for extracting and curating scholarly knowledge.

The RKG infrastructure of NFDI4DS will be based upon existing RKGs and resource collections targeting different types of entities and purposes, which typically can be classified into the following categories: (i) RKGs built on quality-controlled ground truth datasets, (ii) RKGs built on scholarly resource metadata, (iii) RKGs built on primary research data, (iv) RKGs built on automatically (NLP-)generated resource relations, and (v) RKGs built on community expression of scholarly papers. For an overview of the RKGs in NFDI4DS and their categorization, we refer you to Table 1. Additionally, efforts will also be made to establish and include new RKGs that will be developed, e.g., by the methods generated in the shared tasks (see Section 3).

3 Scholarly Information Extraction and RKG Construction

There are different ways RKG construction takes place depending on the category the RKG belongs to and the purpose it serves. For example, RKGs can be generated by lifting a structured database (e.g., DBLP), manually annotating high-quality-controlled corpora (e.g., SoMeSci), by information extraction (IE) using a supervised mechanism (e.g., SoftwareKG), and also by community efforts as the case of ORKG. Depending on the mechanism an RKG is generated, it bears pros and cons with respect to costs, quality, and scalability. An important effort in NFDI4DS is about scholarly IE methods. In fact, a shared task is in place within NFDI4DS for software mention, MLModel, and dataset mention detection which aims at (i) collecting insights and reproducible methodologies from the ML and DS community, and (ii) further promoting the automatic extraction of research artifacts. A major task in NFDI4DS remains the enrichment of existing RKGs with entities (representing research artifacts) and relations extracted from scholarly papers and the provision of ground truth datasets based on scholarly IE. Recent IE tasks on scholarly publications are:

- Classification of Scholarly Publications: This IE task aims at fostering discoverability by multi-class single-label research field classifiers trained on a manually built benchmark dataset based on the ORKG research field taxonomy. For the area of DS and AI research, a specific dataset will be provided to further extend the classifier to solve a multi-class multi-label classification on a more fine-grained taxonomy of research fields.
- Software Mention Detection: To advance transparency of the software used in research initial efforts are made to provide an RKG containing articles that link to the software cited, e.g., as done in SoftwareKG [Sc22]. However, as software mentions are usually

informal and incomplete we aim at advancing the field of software mention detection and disambiguation through novel methods.

- Leaderboard Extraction: *Leaderboards* aim to aggregate a few related research artifacts, such as tasks, models, datasets, evaluation metrics, and scores [Ho19]. The goal is to automatically mine leaderboards for empirical AI research [KDA21, KDA23a, KDA23b] and provide them in a machine-readable way to enable scientists to better keep track of research progress.
- Machine Learning and Dataset Mention Detection: Today, two of the most important entities in scientific articles are *Machine Learning (ML) models* and *datasets*. In an era where DS has an ever-increasing contribution of data to scholarly communication, data is processed and produced by various ML paradigms in different experimental scenarios. This paves the way for intricate relationships between these two entity types such as "related to", "derived from", "cited by", etc.
- Metadata Extraction: Linking methodologies, datasets, and results to their respective authors and research papers is crucial. However, the task is challenging due to the variety of paper templates [Bo21]. Hence, the objective is to leverage multimodal techniques for extracting author metadata and for parsing and extracting references [BAS19].

In addition to IE activities, there exist other initiatives for constructing RKGs from existing data. An example of this is the GESIS Data Search[5] platform, consisting of research and survey studies data provided to users through a search portal. Very recently, an effort has been made to thoroughly exploit these data via metadata mining and structuring them as an RKG. Finally, a further activity is lifting the GESIS Methods Hub into a RKG. The Methods Hub is being developed within the Digital Behavioral Data[6] project about reproducible and reusable methods, that aims to support social scientists to reuse state-of-the-art methods on newly collected data. The Methods Hub RKG will provide methods as first-class research output, and enable their search and citability.

4 NFDI4DS Registry and RKGs Integration

Given the variety of RKGs and their use cases, we envision an NFDI4DS Registry to make them searchable and available to the community. More precisely, we plan to launch an NFDI4DS Registry platform that will present the RKGs along different categories (e.g., scholarly resource metadata, primary research data, etc. as presented in Section 2.), including their documentation, schemas, example queries, etc. Additionally, the NFDI4DS Registry will provide links to the SPARQL endpoints that can be used to explore the RKGs. However, just providing the RKGs is not sufficient to meet the demand for using RKGs to address

[5] https://datasearch.gesis.org/start

[6] https://www.gesis.org/en/institute/digital-behavioral-data

cross-domain research questions; therefore, a particular focus is set on RKG integration and related topics such as schema integration, linking, federation, use of PIDs, among others. In this regard, we pursue the following activities:

Schema Integration. One of the main goals of the NFDI4DS initiative is to make research digital objects FAIR. To achieve this, we are investigating schemas and vocabularies to formally represent research entities, their semantics across disciplines (i.e., an entity can have different meanings in different disciplines), and RKGs' intended uses (e.g., find machine learning models applied on a dataset, describe research trends, etc.). In detail, this activity involves the alignment of current classes and properties among the schemas and vocabularies that are already in use in the existing RKGs (e.g., SoftwareKG, GESIS Search KG, etc.). As an outcome, we plan to develop a unified schema that describes existing RKGs that can be reused for other RKGs (e.g., new RKGs that are being built within the NFDI4DS activities). We will deliver mappings among the existing schemas and provide core schema elements that can be reused in a variety of scenarios. The schema alignment will enable a common representation and search possibilities (e.g., the same query can be used to explore more than one RKG), making the exploration and use of research digital objects easier. In addition, one more expected benefit is that services that are tailored to a specific RKG can be adopted with less or zero costs to other RKGs.

Interlinking and Enrichment. Different RKGs can describe the same entity from different perspectives. For instance, the conceptual entity *Random Forest* is described by both SofwareKG and CS-KG. In SoftwareKG, *Random Forest* is described based on its mentions in research papers and the kind of mention i.e., use, creation, deposition, and allusion; in CS-KG the same entity is described by its connection to other entities e.g., <cskg:random_forest, skos:broader, cskg:machine_learning_method>, <cskg:random_forest, cskg-onto:methodUsedBy, cskg:classification>, etc. Linking various RKGs is an essential activity within NFDI to take advantage of all the current RKGs developments. This will be done based on the type of content RKGs describe (e.g., by exploiting connections to Semantic Web hubs like DBpedia[7] and Wikidata[8], identifiers for authors e.g., ORCIDs, DOIs for papers, etc.). The linking methodology will be based on the content of the various RKGs and their intended use. The linking will enable a more complete representation of the entities, thus allowing RKG users to get a better understanding of the entities and their scope. To achieve this, we plan to explore automatic methods based on semantic similarity and relatedness in order to scale to the millions of entities that the various RKGs listed in Table 1 contain. As a deliverable, we plan to provide links among RKGs to enable federation and enrich the overall knowledge about the described entities.

Persistent Identifiers. When it comes to linking entities and resources from different RKGs, persistent identifiers (PIDs) play an important role. A PID inventory and assessment at GESIS revealed that different RKGs involve heterogeneous PID systems, namespaces, and URI schemas, ranging from third-party ensured DOIs to informal, self-hosted URIs. The

[7] https://www.dbpedia.org/
[8] https://www.wikidata.org/wiki/Wikidata:Main_Page

situation in NFDI4DS is similar, where formal PIDs and DOIs (e.g. used to refer to scientific literature and datasets as e.g. in DBLP KG [db23]) and informal URIs (like e.g. in the SoftwareKG [Sc22]) coexist. To overcome these inconsistencies, we will generate a PID strategy that serves as a guideline and gives recommendations on the assignment of PIDs to entities and resources, e.g. which PID system (externally hosted like DOIs or internally hosted like URIs) to use for which entity types/resource types, and how to organize the joint existence of PIDs like DOIs and URIs for the same entities.

5 Conclusion and Outlook

This paper has summarised current work and future directions for building up an integrated RKG infrastructure for scientific resources in data science and AI, most notably, data, code, and machine learning models. Building on a rich set of established RKGs, datasets, and corpora, involving key resources such as the DBLP bibliographic repository, NFDI4DS aims at enabling a more integrated and machine-interpretable view of research information and resources. The RKG infrastructure will form the foundation for searching and using resources, understanding the dependencies, and more generally, provide a transparent view on the very dynamic research landscape in the field of artificial intelligence.

Another challenge is the need for a community-based agreement on shared PIDs and vocabularies used to identify and describe the research artifacts within the RKG infrastructure. While there is a wealth of vocabularies currently in use, consensus-finding processes are required to establish jointly shared and understand vocabularies and standards. An activity to be taken into account here is the initiative on a common NFDI core ontology[9] aiming to facilitate FAIR access to research data across all scientific domains within the NFDI program [Ti23]. As an outcome for 2024, we plan to provide a web-based registry on which all NFDI4DS RKGs are presented along the different categories and are accessible, e.g., via their SPARQL endpoints. This registry will also provide the elaborated schema mappings between different RKGs, exemplary federated SPARQL queries, computed links between resources/entities of different RKGs as well as a joint NFDI4DS PID strategy to ensure share practices for naming and identifying resources within DS and AI.

Acknowledgements

This work has received funding through the German Research Foundation (DFG) project NFDI4DS (no. 460234259).

Bibliography

[Au20] Auer, Sören; Oelen, Allard; Haris, Muhammad; Stocker, Markus; D'Souza, Jennifer; Farfar, Kheir Eddine; Vogt, Lars; Prinz, Manuel; Wiens, Vitalis; Jaradeh, Mohamad Yaser:

[9] NFDI core ontology: https://ise-fizkarlsruhe.github.io/nfdicore/

Improving access to scientific literature with knowledge graphs. Bibliothek Forschung und Praxis, 44(3):516–529, 2020.

[Ba22] Barker, Michelle; Chue Hong, Neil P; Katz, Daniel S; Lamprecht, Anna-Lena; Martinez-Ortiz, Carlos; Psomopoulos, Fotis; Harrow, Jennifer; Castro, Leyla Jael; Gruenpeter, Morane; Martinez, Paula Andrea et al.: Introducing the FAIR Principles for research software. Scientific Data, 9(1):622, 2022.

[BAS19] Boukhers, Zeyd; Ambhore, Shriharsh; Staab, Steffen: An end-to-end approach for extracting and segmenting high-variance references from pdf documents. In: 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE, pp. 186–195, 2019.

[BBK23] Bless, Christof; Baimuratov, Ildar; Karras, Oliver: SciKGTeX–A LaTeX Package to Semantically Annotate Contributions in Scientific Publications. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2023. 2023.

[Bo21] Boukhers, Zeyd; Beili, Nada; Hartmann, Timo; Goswami, Prantik; Zafar, Muhammad Arslan: MexPub: Deep Transfer Learning for Metadata Extraction from German Publications. In: 2021 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE, pp. 250–253, 2021.

[db23] dblp team: , dblp Knowledge Graph. Download, June 2023. Monthly RDF release of June 2023, `https://dblp.org/rdf/release/dblp-2023-06-01.nt.gz`.

[DCJ19] Dacrema, Maurizio Ferrari; Cremonesi, Paolo; Jannach, Dietmar: Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In (Bogers, Toine; Said, Alan; Brusilovsky, Peter; Tikk, Domonkos, eds): Proceedings of the 13th ACM Conference on Recommender Systems, RecSys 2019, Copenhagen, Denmark, September 16-20, 2019. ACM, pp. 101–109, 2019.

[De22a] Dessì, Danilo; Osborne, Francesco; Recupero, Diego Reforgiato; Buscaldi, Davide; Motta, Enrico: CS-KG: A Large-Scale Knowledge Graph of Research Entities and Claims in Computer Science. In: The Semantic Web - ISWC 2022 - 21st International Semantic Web Conference. volume 13489 of Lecture Notes in Computer Science. Springer, pp. 678–696, 2022.

[De22b] Dessí, Danilo; Osborne, Francesco; Recupero, Diego Reforgiato; Buscaldi, Davide; Motta, Enrico: SCICERO: A deep learning and NLP approach for generating scientific knowledge graphs in the computer science domain. Knowledge-Based Systems, 258:109945, 2022.

[De22c] Dessì, Danilo; Osborne, Francesco; Recupero, Diego Reforgiato; Buscaldi, Davide; Motta, Enrico: SCICERO: A deep learning and NLP approach for generating scientific knowledge graphs in the computer science domain. Knowl. Based Syst., 258:109945, 2022.

[Di20] Dimitrov, Dimitar; Baran, Erdal; Fafalios, Pavlos; Yu, Ran; Zhu, Xiaofei; Zloch, Matthäus; Dietze, Stefan: TweetsCOV19 - A Knowledge Base of Semantically Annotated Tweets about the COVID-19 Pandemic. In: Proceedings of the 29th ACM International Conference on Information and Knowledge Management. CIKM '20, Association for Computing Machinery, New York, NY, USA, p. 2991–2998, 2020.

[Fa18] Fafalios, Pavlos; Iosifidis, Vasileios; Ntoutsi, Eirini; Dietze, Stefan: TweetsKB: A Public and Large-Scale RDF Corpus of Annotated Tweets. In (Gangemi, Aldo; Navigli, Roberto; Vidal, Maria-Esther; Hitzler, Pascal; Troncy, Raphaël; Hollink, Laura; Tordai, Anna; Alam, Mehwish, eds): The Semantic Web. Springer International Publishing, Cham, pp. 177–190, 2018.

[Ga23] Gangopadhyay, Susmita; Boland, Katarina; Dessí, Danilo; Dietze, Stefan; Fafalios, Pavlos; Tchechmedjiev, Andon; Todorov, Konstantin; Jabeen, Hajira: Truth or Dare: Investigating Claims Truthfulness with ClaimsKG. In: Linked Data-driven Resilience Research 2023. CEUR Workshop Proceedings. volume 3401, 2023.

[GGV10] Groth, Paul; Gibson, Andrew; Velterop, Jan: The anatomy of a nanopublication. Information Services & Use, 30(1-2):51–56, 2010.

[Hi19] Hienert, D.; Kern, D.; Boland, K.; Zapilko, B.; Mutschke, P.: A digital library for research data and related information in the social sciences. In: Proceedings of 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL). pp. 148–157, 2019.

[Ho19] Hou, Yufang; Jochim, Charles; Gleize, Martin; Bonin, Francesca; Ganguly, Debasis: Identification of Tasks, Datasets, Evaluation Metrics, and Numeric Scores for Scientific Leaderboards Construction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Florence, Italy, pp. 5203–5213, July 2019.

[Ja19] Jaradeh, Mohamad Yaser; Oelen, Allard; Farfar, Kheir Eddine; Prinz, Manuel; D'Souza, Jennifer; Kismihók, Gábor; Stocker, Markus; Auer, Sören: Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. In: Proceedings of the 10th International Conference on Knowledge Capture. K-CAP '19, Association for Computing Machinery, New York, NY, USA, p. 243–246, 2019.

[KDA21] Kabongo, Salomon; D'Souza, Jennifer; Auer, Sören: Automated Mining of Leaderboards for Empirical AI Research. In: Towards Open and Trustworthy Digital Societies: 23rd International Conference on Asia-Pacific Digital Libraries, ICADL 2021, Virtual Event, December 1–3, 2021, Proceedings 23. Springer, pp. 453–470, 2021.

[KDA23a] Kabongo, Salomon; D'Souza, Jennifer; Auer, Sören: ORKG-Leaderboards: A Systematic Workflow for Mining Leaderboards as a Knowledge Graph. arXiv preprint arXiv:2305.11068, 2023.

[KDA23b] Kabongo, Salomon; D'Souza, Jennifer; Auer, Sören: Zero-shot Entailment of Leaderboards for Empirical AI Research. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2023. 2023.

[Sc21] Schindler, David; Bensmann, Felix; Dietze, Stefan; Krüger, Frank: SoMeSci- A 5 Star Open Data Gold Standard Knowledge Graph of Software Mentions in Scientific Articles. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management. CIKM '21, Association for Computing Machinery, New York, NY, USA, p. 4574–4583, 2021.

[Sc22] Schindler, David; Bensmann, Felix; Dietze, Stefan; Krüger, Frank: The role of software in science: a knowledge graph-based analysis of software mentions in PubMed Central. PeerJ Computer Science, 8, 2022.

[SZK20] Schindler, David; Zapilko, Benjamin; Krüger, Frank: Investigating Software Usage in the Social Sciences: A Knowledge Graph Approach. In (Harth, Andreas; Kirrane, Sabrina; Ngonga Ngomo, Axel-Cyrille; Paulheim, Heiko; Rula, Anisa; Gentile, Anna Lisa; Haase, Peter; Cochez, Michael, eds): The Semantic Web. Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 271–286, 2020.

[Tc19] Tchechmedjiev, A.; Fafalios, P.; Boland, K.; Gasquet, M.; Zloch, M.; Zapilko, B.; Dietze, S.; Todorov, K.: ClaimsKG: A Live Knowledge Graph of Fact-Checked Claims. In: 18th International Semantic Web Conference (ISWC19). 2019.

[Ti23] Tietz, Tabea; Bruns, Oleksandra; Söhn, Linnaea; Tolksdorf, Julia; Posthumus, Etienne; Steller, Jonatan Jalle; Fliegl, Heike; Norouzi, Ebrahim; Waitelonis, Jörg; Schrade, Torsten; Sack, Harald: From Floppy Disks to 5-Star LOD: FAIR Research Infrastructure for NFDI4Culture. In: Proc. of the 3rd Workshop on Metadata and Research (objects) Management for Linked Open Science, DaMaLOS 2023. PUBLISSO - Fachrepositorium Lebenswissenschaften (FRL), May 2023.

[Wi16] Wilkinson, Mark D; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; Axton, Myles; Baak, Arie; Blomberg, Niklas; Boiten, Jan-Willem; da Silva Santos, Luiz Bonino; Bourne, Philip E et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1):1–9, 2016.