

FedORKG: Accessing Federations of Open Research Knowledge Graphs

Abstract for the 1. NFDI4Energy Conference

Philipp D. Rohde¹²³[\[https://orcid.org/0000-0002-9835-4354\]](https://orcid.org/0000-0002-9835-4354),
Enrique Iglesias²³[\[https://orcid.org/0000-0002-8734-3123\]](https://orcid.org/0000-0002-8734-3123), and
Maria-Esther Vidal¹²³[\[https://orcid.org/0000-0003-1160-8727\]](https://orcid.org/0000-0003-1160-8727)

¹TIB Leibniz Information Centre for Science and Technology, Hannover, Germany

²L3S Research Center, Hannover, Germany

³Leibniz University of Hannover, Hannover, Germany

Keywords: Federated Access, Federated Query Processing, Open Research

Scholarly communication has not changed much in the last centuries. However, the number of scientific articles published per year is rapidly increasing. While the growing number of published articles reflects numerous scientific advancements, it is humanly impossible to carefully read all the new papers in one research field, such as physics or computer science. Additionally to the vast number of articles, many research contributions are not freely accessible, i.e., they cannot be accessed without paying for them or they might not even be accessible in certain countries. This is where *Open Research Knowledge Graphs* like the *Open Research Knowledge Graph* (ORKG) [1] come into play. They provide a knowledge-based view of the document-centric scholarly communication [2]. This knowledge-based view does not only help human researchers in searching and filtering interesting articles but also makes the research contributions machine-readable. All this is possible by encoding more information in the ORKG than just the metadata of the papers but also their research field, contribution, experimental setup, conclusions, etc. Open Research Knowledge Graphs also contribute towards a FAIR – following the FAIR principles [3] (Findable, Accessible, Interoperable, Reusable) – communication in science [4].

Providing the contributions of a paper in an open manner is only one aspect of FAIR scholarly communication, particularly in light of the reproducibility crisis [5]. Sharing data and potentially software in a FAIR way is also essential. Therefore, *Research Data Management* has become more crucial than ever. The *Leibniz Data Manager* (LDM) [6] serves as a data repository, enabling researchers to share their data in accordance with the FAIR principles, complete with a citeable DOI. Such data repositories, including the LDM, empower researchers to search for, find, and reuse data in their own investigations. Because the data is citeable, the original authors receive credit when others reuse it. Unlike other research data managers, the LDM offers a knowledge graph containing metadata for all its public data sets—the *LDM KG*. Given that open research demands open access to the data sets used, the LDM KG can be considered an open research knowledge graph.

So far, open research knowledge graphs like the ORKG and LDM KG contribute to FAIR and open research. However, they provide isolated views, i.e., the ORKG focuses on the articles while the LDM KG concentrates on the data sets. It is imperative that both views are combined, i.e., links between articles and data sets are established across the different open research knowledge graphs. Assuming the links are established, the federation of open research knowledge graphs can be used to gain further insights, like which papers are using the same data sets. FedORKG provides the means to query said federation of open research knowledge graphs using SPARQL [7] queries, the W3C recommendation language to query knowledge graphs. Querying this federation is possible using a federated query engine. Federated query engines are capable of deciding which parts of a query can be answered by which member of the federation. These parts, called *sub-query*, are then sent to the identified sources. The intermediate results are transformed into the final result at the query engine. The source selection is guided by the ontologies, i.e., the schemas, of the knowledge graphs in the federation. Hence, the query engine is aware of the different classes, the properties belonging to them, and which sources are serving them.

Accessing the federation of open research knowledge graphs in the above-mentioned manner requires knowledge about the ontologies and experience with SPARQL. While this might be true for a small subset of researchers in the *Semantic Web* community, most researchers, independent of their research field, do not have the necessary skills to exploit the potential of FedORKG when relying solely on writing SPARQL queries. Hence, a natural next step is to add a question answering (QA) system to FedORKG. The QA system allows researchers to post their questions in natural language. Even though SPARQL queries are still used in the backend, the user is not required to have any knowledge about SPARQL or the used ontologies. With the recent rise of AI, deploying such QA systems as required for FedORKG might be close [8].

Another challenge for FedORKG is the alignment between the open research knowledge graphs, i.e., the links between entities across the knowledge graphs, e.g., a scientific paper in the ORKG and the corresponding data set in the LDM KG. A possible solution would be to ask for the DOI of the corresponding paper when adding a data set to LDM. During the creation of the entry in LDM and the LDM KG, it is checked whether the paper already exists in the ORKG. If the paper is not yet present in the ORKG, it should be added to it. However, the DOI lookup service only provides metadata about the article. Ideally, basic information about the research field, problem, contribution, solution, and conclusions would also be added to the ORKG. Due to an expected high quality and correctness of the data, AI-based solutions are not yet used for extracting this kind of information.

FedORKG is the vision of virtually integrating open research knowledge graphs like the ORKG and LDM KG towards truly FAIR and open research. While FedORKG is capable of providing many new insights and can be of assistance in finding relevant articles and data sets, there are many open challenges that need to be addressed for the vision to become true. These challenges include establishing links between the different open research knowledge graphs as well as developing user interfaces that do not require knowledge of any query language.

Author contributions

Philipp D. Rohde: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision

Enrique Iglesias: Software, Validation, Investigation, Resources, Writing - Review & Editing

Maria-Esther Vidal: Conceptualization, Methodology, Investigation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

Competing interests

The authors declare that they have no competing interests.

Funding

The project is funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the Nationale Forschungsdateninfrastruktur (National Research Data Infrastructure) project *NFDI4Energy* (grant no. 501865131). Federal Ministry for Economic Affairs and Energy of Germany (BMWK) in the project CoyPu (project number 01MK21007[A-L]). Leibniz Association, program "Leibniz Best Minds: Programme for Women Professors", project TrustKG-Transforming Data in Trustable Insights; Grant P99/2020.

References

- [1] S. Auer, A. Oelen, M. Haris, *et al.*, "Improving Access to Scientific Literature with Knowledge Graphs," *Bibliothek Forschung und Praxis*, vol. 44, no. 3, pp. 516–529, 2020. DOI: [10.1515/bfp-2020-2042](https://doi.org/10.1515/bfp-2020-2042).
- [2] S. Auer, V. Kovtun, M. Prinz, A. Kasprzik, M. Stocker, and M.-E. Vidal, "Towards a Knowledge Graph for Science," in *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*, New York, NY, USA: ACM, 2018. DOI: [10.1145/3227609.3227689](https://doi.org/10.1145/3227609.3227689).
- [3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, 2016. DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).
- [4] M. Stocker, A. Oelen, M. Y. Jaradeh, *et al.*, "FAIR scientific information with the Open Research Knowledge Graph," *FAIR Connect*, vol. 1, no. 1, pp. 19–21, 2023. DOI: [10.3233/FC-221513](https://doi.org/10.3233/FC-221513).
- [5] J. P. A. Ioannidis, "Why Most Published Research Findings Are False," *PLOS Medicine*, vol. 2, no. 8, Aug. 2005. DOI: [10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124).
- [6] A. Beer, M. Brunet, V. Srivastava, and M.-E. Vidal, "Leibniz Data Manager – A Research Data Management System," in *The Semantic Web: ESWC 2022 Satellite Events*, Cham: Springer International Publishing, 2022, pp. 73–77. DOI: [10.1007/978-3-031-11609-4_14](https://doi.org/10.1007/978-3-031-11609-4_14).
- [7] E. Prud'hommeaux and A. Seaborne, *Shapes Constraint Language (SHACL)*, W3C Recommendation, 2008. [Online]. Available: <https://www.w3.org/TR/rdf-sparql-query/>.
- [8] D. Diefenbach, V. López, K. D. Singh, and P. Maret, "Core techniques of question answering systems over knowledge bases: A survey," *Knowl. Inf. Syst.*, vol. 55, no. 3, pp. 529–569, 2018. DOI: [10.1007/S10115-017-1100-Y](https://doi.org/10.1007/S10115-017-1100-Y). [Online]. Available: <https://doi.org/10.1007/s10115-017-1100-y>.