

Open Research Knowledge Graph: A System Walkthrough

Mohamad Yaser Jaradeh^{1,2}, Allard Oelen^{1,2}, Manuel Prinz², Markus Stocker²,
and Sören Auer^{2,1}

¹ L3S Research Center, Leibniz University of Hannover, Germany
{jaradeh,oelen}@l3s.de

² TIB Leibniz Information Centre for Science and Technology, Germany
{manuel.prinz,markus.stocker,auer}@tib.eu

Abstract. Despite improved digital access to scholarly literature in the last decades, the fundamental principles of scholarly communication remain unchanged and continue to be largely document-based. Scholarly knowledge remains locked in representations that are inadequate for machine processing. The Open Research Knowledge Graph (ORKG) is an infrastructure for representing, curating and exploring scholarly knowledge in a machine actionable manner. We demonstrate the core functionality of ORKG for representing research contributions published in scholarly articles. A video of the demonstration [7] and the system³ are available online.

Keywords: Digital Libraries, Information Science, Knowledge Graph, Research Infrastructure, Scholarly Communication

1 Introduction

Documents are central to scholarly communication. Virtually all research findings are nowadays communicated by means of electronic scholarly articles. Scholarly knowledge communicated in such form is hardly accessible to computers and the primary machine-supported tasks are largely limited to traditional full-text search. As such, the current scholarly infrastructure does not exploit modern information systems and technologies to their full potential [6].

We argue that there is an urgent need for a more flexible, fine-grained, context sensitive representation of scholarly knowledge and thus corresponding infrastructure for knowledge curation, publishing and processing. Furthermore, we suggest that representing scholarly knowledge as structured, interlinked, and semantically rich knowledge graphs is a key element of a technical infrastructure [3].

While some important conceptual foundations have been developed over several decades [1,6], knowledge graph infrastructure for science has recently gained momentum in the literature and community. The Research Graph [2] is a prominent example of an effort that aims to link publications, datasets, and

³ <https://orkg.org/orkg/>

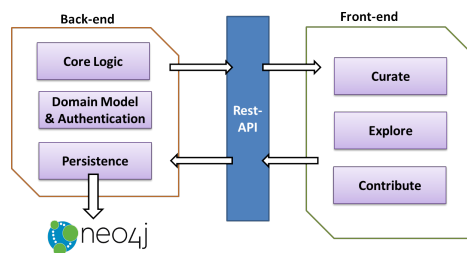


Fig. 1. The ORKG architecture showing the main infrastructure components.

researchers. The Scholix project [4] standardized the information about the links between scholarly literature and data exchanged among (primarily) publishers and data repositories. More recently, the FREYA H2020 project⁴ has released information on their work towards a PID Graph [5]. The key distinguishing factor between these systems and the ORKG is the granularity of captured scholarly knowledge (article bibliographic metadata vs. materials, methods, and results communicated in articles).

2 Architecture and Features

The ORKG leverages knowledge graph technologies to represent, store, link, and process scholarly knowledge. It has two main components: The back end, which contains the logic to handle requests by client applications and the front end through which users create, curate or explore scholarly knowledge.

The concept of **ResearchContribution** is central to the ORKG as it represents key aspects of scholarly knowledge in structured, machine actionable form. A **ResearchContribution** is an information object which relates the **ResearchProblem** addressed by the contribution with a **ResearchMethod** and at least one **ResearchResult**.

The ORKG back end represents descriptions by means of a graph data model. Similarly to the Research Description Framework⁵ (RDF), the data model is centered around the concept of a statement, a triple consisting of two nodes (resources) connected by a directed edge. In contrast to RDF, it allows annotating edges and statements. As metadata of statements, provenance information, e.g. when and by whom a statement was created, is a concrete and relevant application of such annotations.

ORKG users interact with the front end (UI), which guides users through the process of creating research contribution descriptions in a step by step manner. More advanced features of the infrastructure include the ability to directly find similar contributions (and related papers), thus enabling efficient state-of-the-art comparison and literature review. Figure 1 depicts the ORKG system architecture.

⁴ <https://project-freya.eu>

⁵ <https://www.w3.org/RDF/>

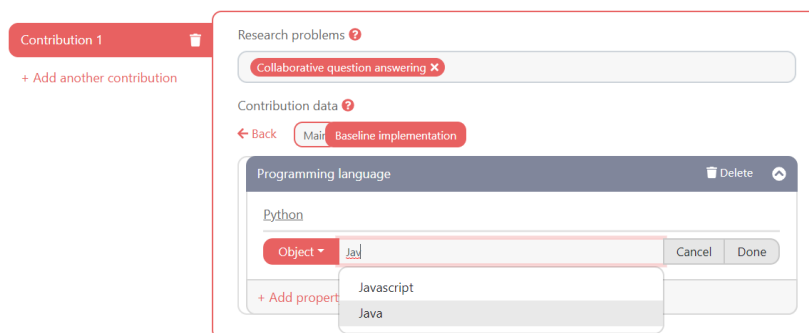


Fig. 2. ORKG UI curation wizard step (3) depicting the auto-completion feature that enables linking existing resources (here, Java).

3 Use Case

Consider the following research contribution: *FRANKENSTEIN [8] is a collaborative question answering (QA) framework written in Java and Python. It generates QA pipelines based on predictions for the best performing pipelines obtained via a supervised learning model. FRANKENSTEIN evaluates the results against QALD and LC-Quad datasets using the f1-score and accuracy@k metrics.*

We can identify the following instances of relevant concepts:

- **Problem:** Collaborative question answering
- **Programming Language:** Python, Java
- **Approach:** Generate optimal QA pipelines
- **Datasets:** QALD, LC-Quad
- **Evaluation Metrics:** f1-score, accuracy@k

Using the “Add paper” wizard (Figure 2), we can create structured descriptions that encode, in machine actionable manner, the key information of research contributions. This process is straightforward also for non-technical users. Firstly, bibliographic metadata is collected, either via DOI lookup using the Crossref API or manually. Secondly, users can classify their paper according to the research domain. Finally, the research contributions described in the paper are collected using a flexible and dynamic interface.

4 Conclusion and Future Work

We presented the Open Research Knowledge Graph, an infrastructure that makes the first steps of a larger research and development agenda that aims to transition document-based scholarly communication to a knowledge-based information representation. In future work, we will include additional techniques from machine support to content creation and curation (such as NLP tools to suggest/annotate relevant concepts on behalf of users). Furthermore, we will further develop novel features such as state-of-the-art comparisons (Figure 3).

Properties	Why Reinvent the Wheel (FRANKENSTEIN) Contribution 1	Capturing Knowledge in Semantically-typed Relational Patterns to Enhance Relation Linking Contribution 1
Research Problem	Collaborative question answering	Relation Linking
Dataset	LC-Quad QALD	QALD
Programming language	java python	python

Fig. 3. ORKG UI state-of-the-art comparison for research contributions, showing a subset of shared properties between two articles.

Such features will underscore the possibilities enabled by machine actionable scholarly knowledge and corresponding infrastructure.

Acknowledgment

This work has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (Grant agreement No. 819536).

References

1. Allen, R.: Model-oriented scientific research reports. *D-Lib Magazine* **17**(5/6) (May 2011). <https://doi.org/10.1045/may2011-allen>
2. Aryani, A., Wang, J.: Research Graph: Building a Distributed Graph of Scholarly Works using Research Data Switchboard. In: *Open Repositories CONFERENCE* (2017)
3. Auer, S., Kovtun, V., Prinz, M., Kasprzik, A., Stocker, M., Vidal, M.E.: Towards a knowledge graph for science. In: *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*. p. 1. ACM (2018)
4. Burton, A., Koers, H., Manghi, P., Stocker, M., Fenner, M., Aryani, A., Bruzzo, S.L., Diepenbroek, M., Schindler, U.: The Scholix Framework for Interoperability in Data-Literature Information Exchange. *D-Lib Magazine* **Volume 23**(1/2) (2017)
5. Fenner, M., Aryani, A.: Introducing the PID Graph (2019). <https://doi.org/10.5438/JWVF-8A66>
6. Hars, A.: Designing scientific knowledge infrastructures: The contribution of epistemology. *Information Systems Frontiers* **3**(1), 63–73 (2001)
7. Jaradeh, M.Y.: A demo of the open research knowledge graph. *Technische Informationsbibliothek TIB, Leibniz Universität Hannover LUH, L3S Research Center* (2019), <https://doi.org/10.5446/42537> *Last accessed* : 21Jun2019
8. Singh, K., Radhakrishna, A.S., Both, A., Shekarpour, S., Lytra, I., Usbeck, R., Vyas, A., Khikmatullaev, A., Punjani, D., Lange, C., Vidal, M.E., Lehmann, J., Auer, S.: Why reinvent the wheel: Let’s build question answering systems together. In: *Proceedings of the 2018 World Wide Web Conference. WWW ’18* (2018)