NFDI4Energy Conference

1. NFDI4Energy Conference

DOI: https://zenodo.org/doi/10.5281/zenodo.10560077

Published: 24.01.2024

Organizing Scientific Knowledge From Energy System Research Using the Open Research Knowledge Graph

Oliver Karras[1][https://orcid.org/0000-0001-5336-6899], Jan Göpfert[23][https://orcid.org/0000-0003-4722-4012], Patrick Kuckertz[2][https://orcid.org/0000-0002-2314-7107], Tristan Pelser[23][https://orcid.org/0009-0006-0424-4784], and Sören Auer[14][https://orcid.org/0000-0002-0698-2864]

[1]TIB - Leibniz Information Centre for Science and Technology, Hanover, Germany
[2]Forschungszentrum Jülich GmbH, Institute of Energy and Climate Research – Techno-economic Systems Analysis (IEK-3), Jülich, Germany
[3]RWTH Aachen University, Chair for Fuel Cells, Faculty of Mechanical Engineering, Aachen, Germany
[4]Leibniz University Hannover, L3S Research Center, Hannover, Germany

Keywords: Energy System Research, FAIR, Open Science, Scenario, Scientific Knowledge, Scientific Communication, Infrastructure, Knowledge Graph, Ontology

1 Introduction

Engineering sciences, such as energy system research, play an important role in developing solutions to technical, environmental, economic, and social challenges of our modern society [1], [2]. In this context, the transformation of energy systems into climate-neutral systems is one of the key strategies for mitigating climate change. For the transformation of energy systems, engineers model, simulate and analyze scenarios and transformation pathways to initiate debates about possible transformation strategies. For these debates and research in general, all steps of the research process must be traceable to guarantee the trustworthiness of published results, avoid redundancies, and ensure their social acceptance [1], [3]. However, the analysis of energy systems is an interdisciplinary field as the investigations of large, complex energy systems often require the use of different software applications and large amounts of heterogeneous data [2]. Engineers must therefore communicate, understand, and (re)use heterogeneous scientific knowledge and data [2], [4]. Although the importance of FAIR [5] scientific knowledge and data in the engineering sciences and energy system research is increasing, little research has been conducted on this topic [1], [2]. When it comes to publishing scientific knowledge and data from publications, software, and datasets (such as models, scenarios, and simulations) openly available and transparent, energy system research lags behind other research domains [3]. According to Schmitt et al. [1] and Nieße et al. [2], engineers need technical support in the form of infrastructures, services, and terminologies to improve communication, understanding, and (re)use of scientific knowledge and data.

2 Background

In 2020, the consortium National Research Data Infrastructure for Engineering Sciences[1] (NFDI4Ing) began developing and deploying infrastructures, i.a., the Open Research Knowledge Graph[2] (ORKG), and services, i.a., the Terminology Service[3] (TS), for engineers to organize FAIR scientific knowledge and data [1]. At the same time, first valuable building blocks for FAIR energy system research have been developed, such as the Open Energy Platform[4] (OEP) and the Open Energy Ontology[5] (OEO) [4]. The ORKG is a cross-domain research knowledge graph combining manual crowd-sourcing and (semi-)automated approaches for the production, curation, and (re)use of FAIR scientific knowledge from publications, software, and datasets [6]–[9]. The TS is a cross-domain service that supports the discovery, provision, design, and curation of ontologies [10]. Based on the work of the task area Ellen[6] in NFDI4Ing, the ORKG integrated the TS that in turn curates the OEO, so that engineers can describe and organize scientific knowledge and data as so-called ORKG *contributions*.

3 Contribution

In this proposal for a presentation, we outline how the consortium National Research Data Infrastructure for the Interdisciplinary Energy System Research[7] (NFDI4Energy) and thus the energy system research community can benefit from the ORKG by organizing scientific knowledge and data to improve its communication, understanding, and (re)use. In this way, we want to illustrate how NFDI4Energy can build on the results of NFDI4Ing to strengthen the collaboration of both consortia towards a joint NFDI. For this purpose, we organized scientific knowledge and data from scenarios of two exemplary use cases in the ORKG. In the first use case, we organized 14 scenario factsheets[8] of the OEP regarding the assumptions, the data used, the associated study, and the results. In the second use case, we organized 25 scenarios from publications on greenhouse gas (GHG) reduction studies for Germany, which have been collected and compared regarding reported electricity supply and installed capacities for different energy sources and the respective scenario goal by Robinius et al. [11].

In both cases, we used ORKG *templates* to facilitate the extraction of information and ensure consistent modeling. ORKG *templates* specify the structure of ORKG *contributions* similar to SHACL shapes [12] and can use ontologies (with the help of the integrated TS) so that the semantic descriptions of scientific knowledge are consistent and comparable across all considered publications. For example, we developed ORKG *templates* for the scenario goal[9] or the electricity supply[10] and used the OEO to ensure clear definitions and the logical interpretation of different types of sectors[11].

For data analysis, we published an ORKG *comparison* with a DOI for each use case to provide a referenceable, citable, and detailed overview of the scenario factsheets [13]

[1] https://nfdi4ing.de/
[2] https://orkg.org/
[3] https://terminology.tib.eu/ts
[4] https://openenergy-platform.org/
[5] https://openenergy-platform.org/ontology/
[6] https://nfdi4ing.de/archetypes/ellen/
[7] https://nfdi4energy.uol.de/
[8] https://openenergy-platform.org/factsheets/scenarios/
[9] https://orkg.org/template/R153118/
[10] https://orkg.org/template/R152170/
[11] https://terminology.tib.eu/ts/ontologies/oeo/terms?iri=http%3A%2F%2Fopenenergy-platform.org%2Fontology%2Foeo%2FOEO_00000367&subtab=graph

and GHG scenarios [14] (see Figure 1). In contrast to the traditional way of publishing an overview of scenarios within a publication, ORKG *comparisons* provide the benefit that they are versionable and can thus be continuously (re)used, updated, and expanded. When researchers publish new scenarios as factsheets or in publications, the ORKG *comparisons* can be easily extended by describing the new scenarios using the same ORKG *templates*, adding the new ORKG *contribution* to the respective ORKG *comparison*, and publishing the updated ORKG *comparison* as a new version. The ORKG also supports the supplementation of ORKG *comparisons* by creating visualizations based on the data contained therein either directly from the web frontend or via various access points, such as a REST API, a Python or R package, or a SPARQL endpoint, for example in combination with a Jupyter notebook.

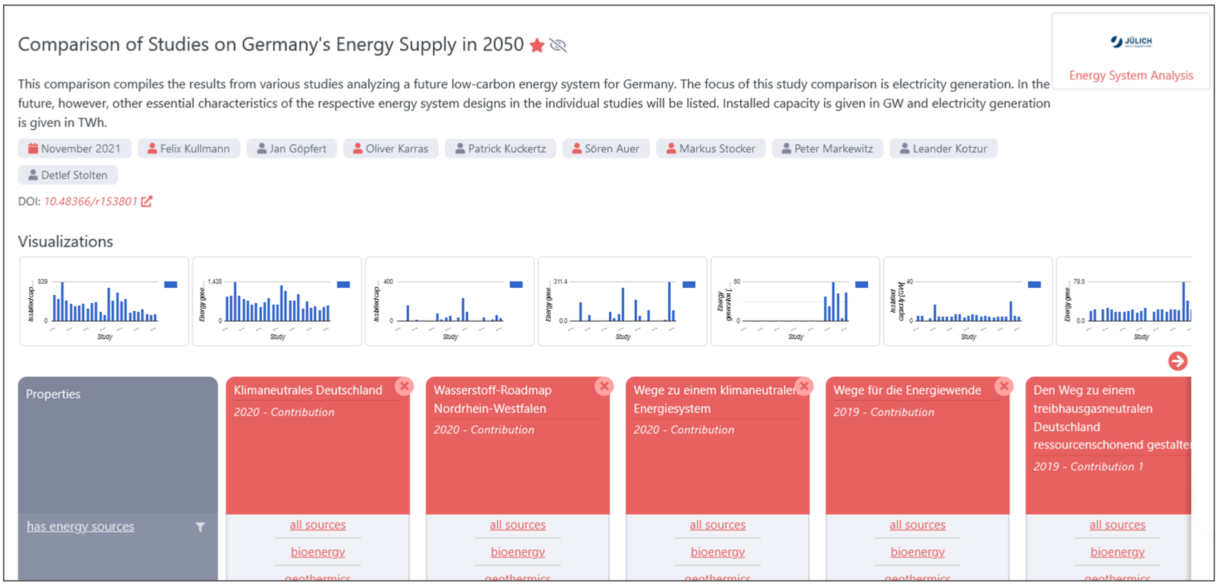

Figure 1. ORKG *comparison* of 25 scenarios from GHG studies for Germany [14].

In addition, we established an ORKG *observatory* on Energy System Research[12]. The ORKG *observatory* serves as a central access point to all related curated publications, comparisons, and visualizations so that other researchers can easily explore the content. For example, Auer et al. [15] already reused the curated scientific knowledge from our two ORKG *comparisons* by identifying and answering further natural language competency questions from domain experts beyond the previous consideration. For this purpose, they specified the competency question as SPARQL query (see Listing 1). We executed this query on the SPARQL endpoint and visualized the results in Figure 2. In particular, these results show that average energy supply from photovoltaics and onshore wind power increased approximately fourfold from the 2006 – 2010 interval to the 2016 – 2020 interval.

[12] https://orkg.org/observatory/Energy_System_Research

Listing 1. SPARQL query for the natural language competency question: "*What is the average energy generation for each energy source considered in 5-year intervals in Greenhouse Gas Reduction Scenarios for Germany?*" by Auer et al. [15]

```
SELECT ?range ?srcLabel AVG(?val) AS ?avgVal
WHERE {
  r:R153801  p:compareContribution  ?contrib.
  ?paper     p:hasContribution      ?contrib;
             p:hasPublicationYear   ?year.
  BIND(xsd:int(?year) AS ?y).
  VALUES(?range ?min ?max) {
    ("2001-2005" 2001 2005)
    ("2006-2010" 2006 2010)
    ("2011-2015" 2011 2015)
    ("2016-2020" 2016 2020)
  } FILTER(?min <= ?y && ?y <= ?max).
  ?contrib    p:hasEnergySources  ?energySrc.
  ?energySrc  rdfs:label          ?srcLabel;
              p:hasGeneration     ?energyGen.
  ?energyGen  p:hasValue          ?genVal.
  BIND(xsd:float(?genVal) AS ?val).
} ORDER BY ASC(?range)
```

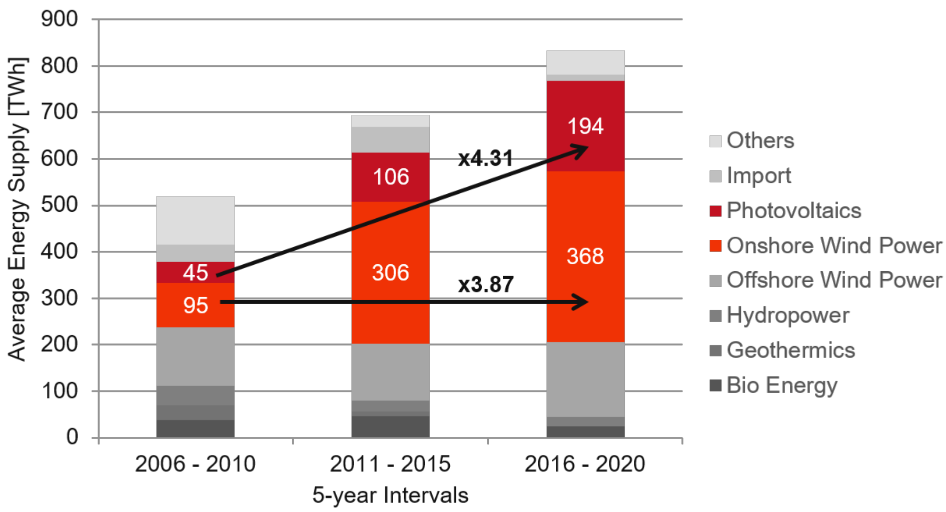

Figure 2. Visualized results from the SPARQL query by Auer et al. [15]

4 Conclusion

Overall, this proposal addresses multiple themes of the 1st NFDI4Energy conference[13]. In particular, we demonstrate how engineers can use the ORKG infrastructure in innovative ways to organize scientific knowledge and data from publications, software, and datasets by using ontologies, such as the OEO, for FAIR data management and thus open science. We hope to encourage and motivate engineers, especially in energy system research, to use the ORKG and its integrated services to communicate, understand, and (re)use scientific knowledge, thus promoting collaboration between NFDI4Energy and NFDI4Ing.

[13] https://nfdi4energy.uol.de/sites/conference/

Data availability statement

All data used are openly available in the Open Research Knowledge Graph: https://orkg.org/ and in particular in the ORKG *observatory* on Energy System Research: https://orkg.org/observatory/Energy_System_Research.

Author contributions

Oliver Karras: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Visualization, Supervision, Project administration.
Jan Göpfert: Conceptualization, Methodology, Validation, Investigation, Resources, Data Curation, Writing - Original Draft.
Patrick Kuckertz: Conceptualization, Writing - Review & Editing, Project administration.
Tristan Pelser: Writing - Review & Editing.
Sören Auer: Conceptualization, Supervision, Funding acquisition.

Competing interests

The authors declare that they have no competing interests.

Funding

The authors thank the Federal Government, the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the NFDI4Ing, NFDI4DataScience, and NFDI4Energy consortia. This work was funded by the German Research Foundation (DFG) -project numbers 442146713, 460234259, 501865131, by the European Research Council for the project ScienceGRAPH (Grant agreement ID: 819536), and by the TIB - Leibniz Information Centre for Science and Technology.

Acknowledgements

The ORKG is a team effort comprising overall more than 50 people distributed worldwide working on the development of the infrastructure and its services, curating content, and researching novel features.

References

[1] R. H. Schmitt, V. Anthofer, S. Auer, *et al.*, *NFDI4Ing - The National Research Data Infrastructure for Engineering Sciences*, 2020. DOI: 10.5281/zenodo.4015201.

[2] A. Nieße, S. Ferenz, S. Auer, *et al.*, *NFDI4Energy – National Research Data Infrastructure for the Interdisciplinary Energy System Research*, 2022. DOI: 10.5281/zenodo.6772013.

[3] S. Pfenninger, J. DeCarolis, L. Hirth, S. Quoilin, and I. Staffell, "The Importance of Open Data and Software: Is Energy Research Lagging Behind?" *Energy Policy*, vol. 101, pp. 211–215, 2017. DOI: 10.1016/j.enpol.2016.11.046.

[4] M. Booshehri, L. Emele, S. Flügel, *et al.*, "Introducing the Open Energy Ontology: Enhancing Data Interpretation and Interfacing in Energy Systems Analysis," *Energy and AI*, vol. 5, 2021. DOI: 10.1016/j.egyai.2021.100074.

[5] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, *et al.*, "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data*, vol. 3, no. 1, 2016. DOI: 10.1038/sdata.2016.18.

[6] S. Auer, A. Oelen, M. Haris, *et al.*, "Improving Access to Scientific Literature with Knowledge Graphs," *Bibliothek Forschung und Praxis*, vol. 44, no. 3, 2020. DOI: 10.1515/bfp-2020-2042.

[7] O. Karras, E. C. Groen, J. A. Khan, and S. Auer, "Researcher or Crowd Member? Why not both! The Open Research Knowledge Graph for Applying and Communicating CrowdRE Research," in *2021 IEEE 29th International Requirements Engineering Conference Workshops (REW)*, IEEE, 2021, pp. 320–327. DOI: 10.1109/REW53955.2021.00056.

[8] S. Auer, M. Stocker, O. Karras, A. Oelen, J. D'Souza, and A.-L. Lorenz, "Organizing Scholarly Knowledge in the Open Research Knowledge Graph: An Open-Science Platform for FAIR Scholarly Knowledge," in *Proceedings of the Conference on Research Data Infrastructure*, vol. 1, 2023. DOI: 10.52825/cordi.v1i.272.

[9] M. Stocker, A. Oelen, M. Y. Jaradeh, M. Haris, O. A. Oghli, *et al.*, "FAIR Scientific Information with the Open Research Knowledge Graph," *FAIR Connect*, vol. 1, no. 1, 2023. DOI: 10.3233/FC-221513.

[10] P. Strömert, V. Limbachia, P. Oladazimi, J. Hunold, and O. Koepler, "Towards a Versatile Terminology Service for Empowering FAIR Research Data: Enabling Ontology Discovery, Design, Curation, and Utilization Across Scientific Communities," in *Knowledge Graphs: Semantics, Machine Learning, and Languages*, IOS Press, 2023, pp. 53–69. DOI: 10.3233/SSW230005.

[11] M. Robinius, P. Markewitz, P. Lopion, *et al.*, *WEGE FÜR DIE ENERGIEWENDE Kosteneffiziente und klimagerechte Transformationsstrategien für das deutsche Energiesystem bis zum Jahr 2050* (Schriften des Forschungszentrums Jülich Reihe Energie & Umwelt / Energy & Environment). Forschungszentrum Jülich GmbH Zentralbibliothek, Verlag, 2020, vol. 499, VIII, 141 S. ISBN: 978-3-95806-483-6. [Online]. Available: https://juser.fz-juelich.de/record/877960.

[12] H. Knublauch and D. Kontokostas, "Shapes Constraint Language (SHACL)," W3C, W3C Recommendation, 2017. [Online]. Available: https://www.w3.org/TR/2017/REC-shacl-20170720/.

[13] J. Göpfert and O. Karras, *Comparison of Scenario Factsheets from the Open Energy Platform*, Open Research Knowledge Graph, 2022. DOI: 10.48366/R150337.

[14] F. Kullmann, J. Göpfert, O. Karras, *et al.*, *Comparison of Studies on Germany's Energy Supply in 2050*, Open Research Knowledge Graph, 2021. DOI: 10.48366/R153801.

[15] S. Auer, D. A. C. Barone, C. Bartz, *et al.*, "The SciQA Scientific Question Answering Benchmark for Scholarly Knowledge," *Nature Scientific Reports*, vol. 13, no. 7240, 2023. DOI: 10.1038/s41598-023-33607-z.