Improved Findability of Energy Research Software by Introducing a Metadata-based Registry

Abstract for the 1. NFDI4Energy Conference 2024

Stephan Ferenz[1][2][https://orcid.org/0000-0001-9523-7227], and
Astrid Nieße[1][2][https://orcid.org/0000-0003-1881-9172]

[1]Department of Computer Science, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany
[2]OFFIS e.V. – Institute for Information Technology, Oldenburg, Germany
*Correspondence: Stephan Ferenz, stephan.ferenz@uol.de

Abstract:

Keywords: Interoperability, Digital Libraries, Energy Research, FAIR, Research Software, Metadata, Open Source Software, Software Reusability

1 Introduction and Motivation

Energy research software (ERS, as defined in [1]) is required for multiple purposes like visualization of processes and values, e.g., power quality [2], (co-)simulation of smart grids [3], or analysis of transition paths [4]. Within an exemplified research cycle, self-designed ERS is often fundamental for producing new research results while also presenting a result of performed research (see Figure 1).

New ERS is often developed without reusing existing ones. Therefore, a lot of time is spent on (re)developing software instead of doing research slowing down the progress in research.

Especially the findability of research software can be increased with useful metadata stored in a registry which enables the reuse of research software, as described in the FAIR criteria for research software [5]–[8]. Software registries only store metadata for software without the source code of the software while software repositories also contain source code [9]. Therefore, good metadata and a registry are a first step for improving the reuse in ERS and also the whole research process in energy research.

For this goal, we would like to present a concept for a metadata-based registry for ERS based on a good metadata scheme which we already proposed in [1][1]. In the following, we give a brief overview on the related work. Afterwards, we describe our concept and, finally, conclude the abstract.

---

[1]For this abstract, we reuse multiple texts and graphics from that submitted but yet not accepted paper.
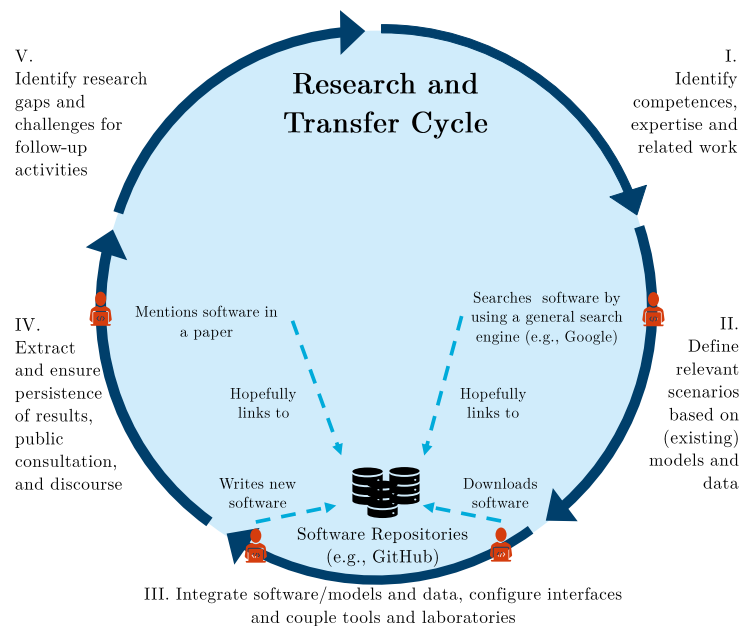
Figure 1. An exemplified research cycle of energy research with software usage

2 State-of-the-art

In the context of research software registries bio.tools[2] [10] is a good example for a well-functioning registry for research software. In the energy domain, the Open Energy Platform[3] provides kind of a registry without using a formalized metadata scheme and providing interoperable metadata.

While some metadata schemes for research software focus on some main properties, like *CodeMeta*[4], others try to include detailed domain knowledge based on domain ontologies like *biotoolsXSD* [10] (for Life Science) or the *Software Description Ontology* [11] (for Geoscience). In the energy domain, a formalized metadata scheme for ERS is still missing. However, the open energy platform[3], the openmod wiki[5], and the work of Schwarz and Lehnhoff [13] present good starting points for developing a formalized metadata scheme for ERS.

For good and interoperable metadata, domain ontologies are required. It seems promising that one or a combination of multiple existing energy domain ontologies, for example of the *OEO* [14] and *CIM* [15], can be used as value vocabulary for describing ERS.

The existing approaches for ERS, like the openmod wiki and the open energy platform, lack formalization, are not based on existing approaches and/or do not use value vocabularies. But, there exist good approaches for registries and metadata schemes for research software in other domains usable as orientation for developing a domain specific solution for the energy domain, including a metadata scheme for ERS. For this metadata scheme, it seems promising that the existing domain ontologies can be used as value vocabularies.

---

[2]http://bio.tools, accessed 20.10.2023

[3]https://openenergy-platform.org/factsheets/models/, accessed 20.10.2023

[4]https://codemeta.github.io/, accessed 20.10.2023

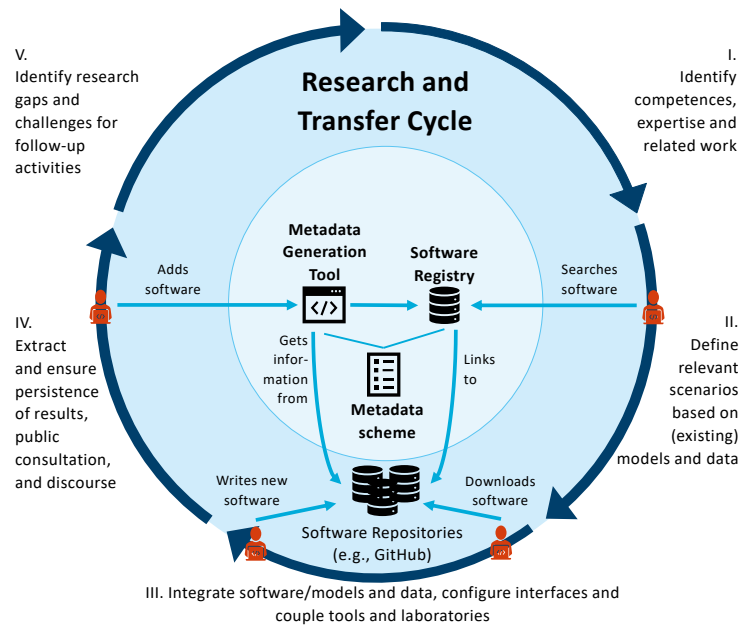[5]https://wiki.openmod-initiative.org/wiki/Open_Models, accessed 20.10.2023

Figure 2. A metadata-based registry support the exemplified research cycle of energy research with software usage [18]

3 A Metadata-Based Registry for ERS

Our concept for a Metadata-based Registry for ERS and its usage is shown in Figure 2. First in the research cycle, researchers should be able to use the registry to find relevant ERS. The registry should be based on a new developed metadata scheme for ERS and should reuse existing approaches from other domains like bio.tools[2].

After downloading ERS, researchers will use it, extend it, and/or write additional code for their research. The new or extended software should be published on any existing software repository. Afterwards, researchers should be able to add their software with good metadata to the registry by using the metadata generation tool. The tool should lower the entrance barrier for creating metadata for all researchers in the energy domain. Therefore, the tool consist of two main functionalities: First, the tool should extract as many metadata as possible from software repositories like GitHub and other sources. Different ways of extracting metadata have already described within SoMEF [16], HERMES[6], and Betty's Research Engine [17]. Second, the tool should allow researchers to curate all extracted metadata and to add additional non-extracted metadata. The tool should encourage the use of controlled value vocabularies and links to other semantic web sources.

Both the metadata generation tool and the registry are based on a new metadata scheme for ERS. This metadata scheme should be usable for all different types of ERS. It consists of elements describing the categories for the metadata, guidelines for creating according metadata, a syntax, and of constraints for the metadata, e.g., ontologies as value vocabularies [19]. The metadata scheme should be developed as an application profile according to the requirements of energy researchers and by reusing as many metadata elements as possible to foster a high interoperability.

---

[6]https://project.software-metadata.pub/, accessed 10.07.2023

4 Conclusion

There already exist a lot of ERS and even more will be needed in the future. ERS is often developed without reusing existing software because relevant ERS is difficult to find. The overall findability of ERS can be improved through register ERS in a registry with good metadata. Therefore, we want to present and discuss our concept of a **metadata-based registry for ERS** to enable this process on the NFDI4Energy conference. This registry is highly relevant in the context of research software management in the energy domain and we would like to discuss it with the NFDI4Energy community.

We currently perform a detailed requirement analysis based on multiple interviews with energy researchers. We would like to show and discuss first results of the analysis on the NFDI4Energy conference as well.

Author contributions

Stephan Ferenz contributed in conceptualization, methodology, writing - original draft. Astrid Nieße contributed in conceptualization, methodology, writing - review & editing.

Competing interests

The authors declare that they have no competing interests.

Funding

Acknowledgements

References

[1] S. Ferenz and A. Nieße, *Towards Improved Findability of Energy Research Software by Introducing a Metadata-based Registry*, en, Preprint, Mar. 2023. [Online]. Available: https://preprints.inggrid.org/repository/view/9/ (visited on 09/22/2023).

[2] A. J. Christe, S. I. Negrashov, P. M. Johnson, D. Nakahodo, D. Badke, and D. Aghalarpour, "OPQ Version 2: An Architecture for Distributed, Real-Time, High Performance Power Data Acquisition, Analysis, and Visualization," in *2017 IEEE 7th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, Jul. 2017, pp. 1415–1419. DOI: 10.1109/CYBER.2017.8446111.

[3] C. Steinbrink, M. Blank-Babazadeh, A. El-Ama, *et al.*, "CPES Testing with mosaik: Co-Simulation Planning, Execution and Analysis," *Applied Sciences*, vol. 9, no. 5, p. 923, Jan. 2019. DOI: 10.3390/app9050923.

[4] K. Löffler, K. Hainsch, T. Burandt, P.-Y. Oei, C. Kemfert, and C. Von Hirschhausen, "Designing a Model for the Global Energy System—GENeSYS-MOD: An Application of the Open-Source Energy Modeling System (OSeMOSYS)," *Energies*, vol. 10, no. 10, p. 1468, Oct. 2017. DOI: 10.3390/en10101468.

[5] W. Hasselbring, L. Carr, S. Hettrick, H. Packer, and T. Tiropanis, "From FAIR research data toward FAIR and open research software," *it - Information Technology*, vol. 62, no. 1, pp. 39–47, Feb. 2020, ISSN: 1611-2776, 2196-7032. DOI: 10.1515/itit-2019-0040.

[6] A.-L. Lamprecht, L. Garcia, M. Kuzak, *et al.*, "Towards FAIR principles for research software," *Data Science*, vol. 3, no. 1, pp. 37–59, Jan. 2020, ISSN: 2451-8484. DOI: 10.3233/DS-190026.

[7] D. S. Katz, M. Gruenpeter, and T. Honeyman, "Taking a fresh look at FAIR for research software," *Patterns*, vol. 2, no. 3, Mar. 2021, ISSN: 2666-3899. DOI: 10.1016/j.patter.2021.100222.

[8] S. Druskat, O. Bertuch, G. Juckeland, O. Knodel, and T. Schlauch, *Software publications with rich metadata: State of the art, automated workflows and HERMES concept*, Preprint, Jan. 2022. DOI: 10.48550/arXiv.2201.09015.

[9] D. Garijo, H. Ménager, L. Hwang, *et al.*, "Nine best practices for research software registries and repositories," en, *PeerJ Computer Science*, vol. 8, e1023, Aug. 2022, Publisher: PeerJ Inc., ISSN: 2376-5992. DOI: 10.7717/peerj-cs.1023.

[10] J. Ison, H. Ienasescu, P. Chmura, *et al.*, "The bio.tools registry of software tools and data resources for the life sciences," *Genome Biology*, vol. 20, no. 1, p. 164, Aug. 2019, ISSN: 1474-760X. DOI: 10.1186/s13059-019-1772-6.

[11] D. Garijo, M. Osorio, D. Khider, V. Ratnakar, and Y. Gil, "OKG-Soft: An Open Knowledge Graph with Machine Readable Scientific Software Metadata," in *2019 15th International Conference on eScience (eScience)*, Sep. 2019, pp. 349–358. DOI: 10.1109/eScience.2019.00046.

[12] B. Schembera and D. Iglezakis, "EngMeta – Metadata for Computational Engineering," *International Journal of Metadata, Semantics and Ontologies*, vol. 14, no. 1, p. 26, 2020, ISSN: 1744-2621, 1744-263X. DOI: 10.1504/IJMSO.2020.107792.

[13] J. Schwarz and S. Lehnhoff, "Ontological Integration of Semantics and Domain Knowledge in Energy Scenario Co-simulation," in *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, Vienna, Austria: SCITEPRESS - Science and Technology Publications, 2019, pp. 127–136, ISBN: 978-989-758-382-7. DOI: 10.5220/0008069801270136.

[14] 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, p. 100 074, Sep. 2021, ISSN: 2666-5468. DOI: 10.1016/j.egyai.2021.100074.

[15] M. Uslar, M. Specht, S. Rohjans, J. Trefke, and J. M. Vasquez Gonzalez, *The Common Information Model CIM* (Power Systems). Berlin, Germany and Heidelberg, Germany: Springer Berlin Heidelberg, 2012, vol. 66, ISBN: 978-3-642-25215-0. DOI: 10.1007/978-3-642-25215-0.

[16] A. Mao, D. Garijo, and S. Fakhraei, "SoMEF: A Framework for Capturing Scientific Software Metadata from its Documentation," in *2019 IEEE International Conference on Big Data (Big Data)*, Dec. 2019, pp. 3032–3037. DOI: 10.1109/BigData47090.2019.9006447.

[17] V. Seibert, A. Rausch, and S. Wittek, *Betty's (Re)Search Engine: A client-based search engine for research software stored in repositories.* en, Preprint, Mar. 2023. [Online]. Available: https://preprints.inggrid.org/repository/view/16/ (visited on 07/17/2023).

[18] S. Ferenz, "Towards More Findable Energy Research Software by Introducing a Metadata-based Registry," in *Abstracts of the 11th DACH+ Conference on Energy Informatics*, Anke Weidlich, Gunther Gust and Mirko Schäfer, Ed., Springer, 2022, ISBN: 2520-8942. DOI: 10.1186/s42162-022-00215-6.

[19] M. L. Zeng and J. Qin, *Metadata*, Third edition. London, Great Britain: Facet Publishing, 2022, ISBN: 978-1-78330-588-9.