FAIR Data for Energy System Research

An Overview of NFDI4Energy Task Area 4

Amanda Wein[1][https://orcid.org/0009-0009-2960-3474], Jan Reinkensmeier[1], Anke Weidlich[2][https://orcid.org/0000-0003-2361-0912], Johan Lilliestam[3][https://orcid.org/0000-0001-6913-5956], Veit Hagenmeyer[4][https://orcid.org/0000-0002-3572-9083], Mascha Richter[5], Sören Auer[6][https://orcid.org/0000-0002-0698-2864], Astrid Nieße[7][https://orcid.org/0000-0003-1881-9172], and Sebastian Lehnhoff[1][https://orcid.org/0000-0003-2340-6807]

[1] OFFIS e.V., Germany

[2] Albert-Ludwigs-Universität Freiburg, Germany

[3] Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

[4] Karlsruher Institute of Technology – Institute for Automation and Applied Informatics, Germany

[5] Reiner Lemoine Institut gGmbH, Germany

[6] Technische Informationsbibliothek, Germany

[7] Carl von Ossietzky Universität Oldenburg, Germany

Abstract. The NFDI4Energy consortium aims to establish new services filling a variety of needs for the energy system research community, from making FAIR research data easily accessible to promoting collaboration among community entities. Seven Task Areas (TAs) have been defined to achieve the consortium's objectives, each with a specific focus. Task Area 4 (TA4): FAIR Data for Energy System Research shall develop ontologies, metadata standards, and services to promote semantic consistency and improve interoperability of energy research projects, thereby supporting the harmonization of data management among various institutions and research fields.

Keywords: Energy System Research, FAIR Data, Ontologies, Metadata, Data Infrastructure

1. Motivation

Energy system research is a highly interdisciplinary field, relying heavily upon expertise and existing models from engineering, economics, and meteorology, among other areas. The sheer quantity of data required for energy system models, and the disparate sources of this data, challenge the efficiency of energy research projects. Standardization of research data and metadata according to the FAIR Principles [1] is expected to yield significant benefits, including improved cooperation among research entities and greater ease in accessing relevant data sets for projects [2].

To promote and enable the usage of FAIR data in energy system research, TA4 focuses on several objectives:

- creation of ontologies for energy system research,
- creation of metadata standards to cover relevant Digital Objects (DOs), and
- implementation of a metadata registry and persistent identifier (PID) service for DOs [2].

The services developed under TA4 will form a semantic layer for the overall NFDI4Energy platform. This TA is a central component of the consortium's goals, supported by and providing support to multiple other TAs. Input from TA1 (Building and Serving the Energy Research Community) and TA2 (Integrating Society and Policy in Energy Research) will be key to ensuring the relevance of the ontology and metadata standards produced by TA4. TA5 (Simulation in Interdisciplinary Energy Research) will develop an energy simulation software ontology and a simulation model registry, which must be capable of full integration with the domain ontology and metadata registry developed by TA4. In addition, TA6 (Use Cases for Community Services) will define use cases for the services developed by NFDI4Energy, based on TA4's methods and services. [2]

2. Methodology

2.1 Ontologies & Standards

Two related ontologies shall be worked on. The first, a domain ontology, shall establish a common vocabulary for energy system research and shall hierarchically store knowledge from this domain; it will indicate relationships between concepts and will be capable of linking to the ontologies of other research areas, enabling interdisciplinary applications. The second ontology will extend the domain ontology by establishing terminology for defining long-term energy system scenarios. This scenario ontology shall allow for better comparisons between scenarios, based on the vocabulary established by the domain ontology. [2]

Several well-known ontologies for the energy domain currently exist, such as the Open Energy Ontology [3]. Therefore, the approach taken by TA4 will start with assessing these ontologies to determine which one(s) may be optimal for extending into the new domain and scenario ontologies. This ontology selection, improvement, and extension will be based upon an examination of features needed in the new ontologies and a comparison to features available in existing ontologies. Continuous feedback from other TAs and from researchers throughout the development process will ensure that the new ontologies closely align with the needs of the scientific community.

Together with the ontologies, a set of metadata standards will be created using the domain ontology. TA1, TA2, and TA6 will assist with defining requirements for these standards; in addition, TA4 will collaborate with the other NFDI consortia to discuss best practices for designing metadata standards that encourage interoperability with other standards. [2]

Working groups shall be established to encourage the scientific community's participation in this ontology and standards development effort [2].

2.2 Infrastructure & Services

Several software services are planned to support TA4's ontologies and standards, as well as the FAIRification of DOs.

A Terminology Service (TS) based on W3C standards will function as the main access point to search for and retrieve terminology information. It will include a registry tool to enable the registering in the TS of not only the TA4 domain and scenario ontologies, but also additional ontologies which are relevant for energy systems research. [2]

An Open Research Knowledge Graph (ORKG) [4] for energy research scenarios is also planned, as a service to simplify scenario comparison. The ORKG information structure is designed to semantically link knowledge from various research publications, making this knowledge both human- and machine-readable. In this context, the graph shall contain data on scenarios defined according to the scenario ontology. [2]

A metadata registry shall be developed based on the Leibniz Data Manager (LDM) [5]. This tool will link digital repositories of research data, allowing for visualizations and metadata maintenance of data sets. It shall assist researchers in selecting appropriate data sets for their projects, and improve the interoperability of data sets through standardized metadata. [2]

PIDs are a key element of FAIR data [6], and therefore a key element of TA4's work. A PID service will be created to interface with the metadata registry, using either an existing PID concept or, if necessary to meet user needs, a new domain-specific concept. All PIDs will be connected in a PID graph [7] to improve DO findability. [2]

Finally, several services shall facilitate the integration of TA4's artifacts with the overall NFDI4Energy platform: federated access and search services, REST APIs, and a submission service for publishing DOs. [2]
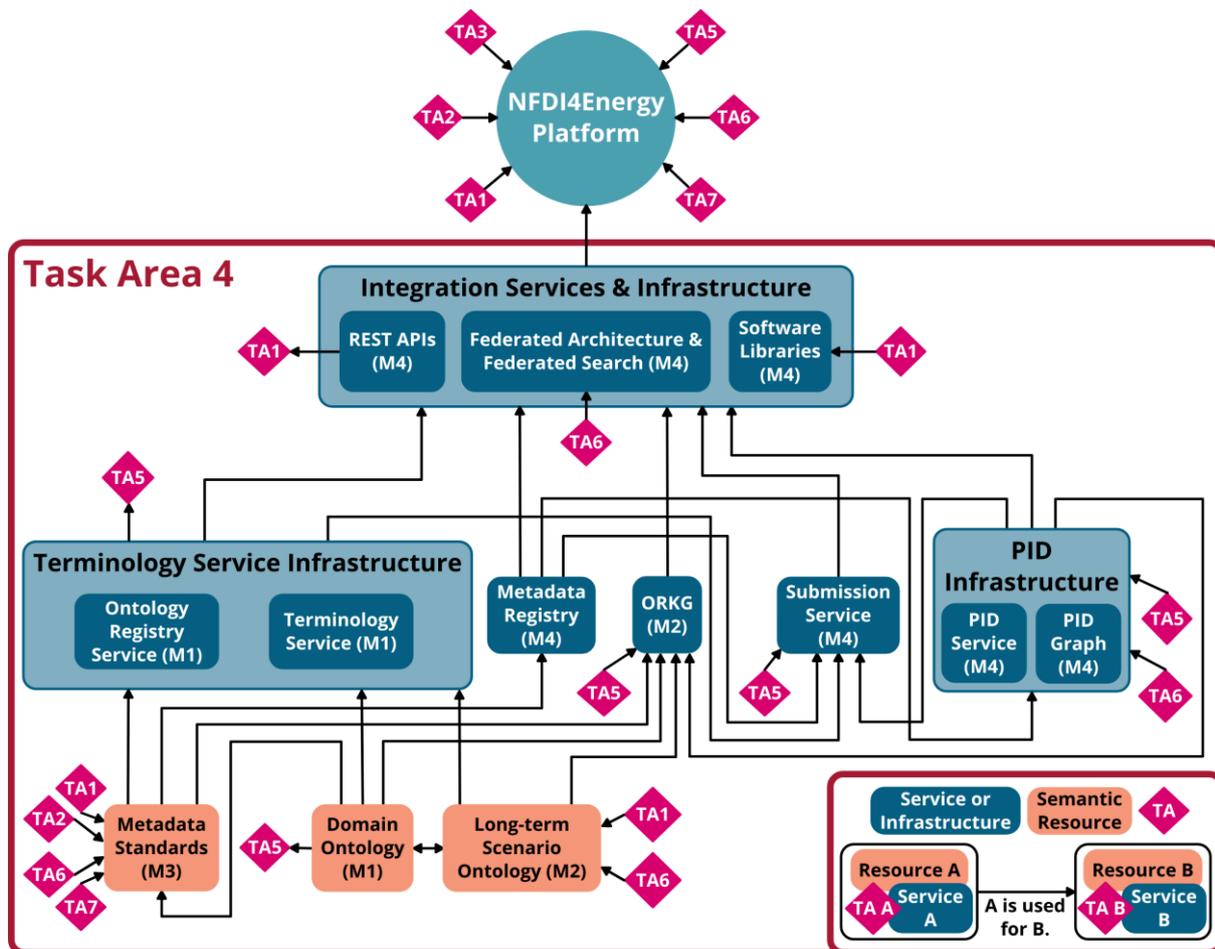

Figure 1. An overview of the resources to be created by TA4, with arrows indicating links between resources. The goals of TA4 are internally called Measures and are noted here as M1-M4.

3. Participants

TA4 will be led by OFFIS, supported by TA Lead Sebastian Lehnhoff and TA Coordinator Amanda Wein. Organizations contributing to TA4 are Albert-Ludwigs-Universität Freiburg, Friedrich-Alexander-Universität Erlangen-Nürnberg, Karlsruher Institute of Technology – Institute for Automation and Applied Informatics, Reiner Lemoine Institut, Technische Informationsbibliothek, and Carl von Ossietzky Universität Oldenburg. [2]

Competing interests

The authors declare that they have no competing interests.

Funding

References

1. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, et al., "The FAIR Guiding Principles for scientific data management and stewardship," *Sci Data*, vol. 3, article 160018, Mar. 2016, doi: https://doi.org/10.1038/sdata.2016.18.
2. A. Nieße, S. Ferenz, S. Auer, S. Dähling, S. Decker, J. Dorfner, et al. "nfdi4energy – National Research Data Infrastructure for the Interdisciplinary Energy System Research." Zenodo. https://zenodo.org/record/6772013#.ZEYuontBybg (accessed Apr. 24, 2023). doi: https://doi.org/10.5281/zenodo.6772013.
3. M. Booshehri, L. Emele, S. Flügel, H. Förster, J. Frey, U. Frey, et al., "Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis," *Energy and AI*, vol. 5, article 100074, Sept. 2021, doi: https://doi.org/10.1016/j.egyai.2021.100074.
4. S. Auer, A. Oelen, M. Haris, M. Stocker, J. D'Souza, K. E. Farfar, et al., "Improving Access to Scientific Literature with Knowledge Graphs," *Bibliothek Forschung und Praxis*, vol. 44, no. 3, pp. 516-529, 2020, doi: https://doi.org/10.1515/bfp-2020-2042.
5. "Leibniz Data Manager." TIB Labs. https://labs.tib.eu/info/en/project/leibniz-data-manager/ (accessed Apr. 18, 2023).
6. "Persistent Identifiers (PIDs)." PID Kompetenzzentrum: Ein Service der TIB. https://projects.tib.eu/pid-service/en/persistent-identifiers/persistent-identifiers-pids/ (accessed Apr. 18, 2023).
7. M. Fenner and A. Aryani. "Introducing the PID Graph." DataCite Blog. https://blog.datacite.org/introducing-the-pid-graph/ (accessed Apr. 18, 2023). doi: https://doi.org/10.5438/jwvf-8a66.