









NFDI4Energy Task Area 4: FAIR Data for Energy System Research


Amanda Wein ¹, Jan Reinkensmeier², Anke Weidlich ³, Johan Lilliestam ⁴,
Veit Hagenmeyer ⁵, Mascha Richter⁶, Sören Auer ⁷, Astrid Nieße ⁸ and
Sebastian Lehnhoff ⁹


Abstract: The NFDI4Energy consortium will create a research data infrastructure for energy system research, emphasizing the openness and FAIRness of data and models in this research domain. Within the consortium, Task Area 4 focuses on the development of resources and services that will provide a semantic layer for the overall platform built by NFDI4Energy. The team of this Task Area will produce artifacts including a domain ontology, metadata standards, a knowledge graph, a Persistent Identifier service, and integration infrastructure to join these artifacts to the NFDI4Energy platform..

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

¹ OFFIS e.V., R&D Division Energy, Escherweg 2, Oldenburg, 26121, amanda.wein@offis.de, 
<https://orcid.org/0009-0009-2960-3474>


² OFFIS e.V., R&D Division Energy, Escherweg 2, Oldenburg, 26121, jan.reinkensmeier@offis.de


³ Albert-Ludwigs-Universität Freiburg., Institut für Nachhaltige Technische Systeme (INATECH),
Friedrichstr. 39, Freiburg, 79110, anke.weidlich@inatech.uni-freiburg.de, 
<https://orcid.org/0000-0003-2361-0912>


⁴ Friedrich-Alexander-Universität Erlangen-Nürnberg, Lehrstuhl für Sustainability Transition
Policy, Lange Gasse 20, Nürnberg, 90403, Johan.Lilliestam@iass-potsdam.de, 
<https://orcid.org/0000-0001-6913-5956>

⁵ Karlsruhe Institute of Technology, Institute for Automation and Applied Informatics, Kaiserstr.
89, Karlsruhe, 76131, veit.hagenmeyer@kit.edu, 
<https://orcid.org/0000-0002-3572-9083>

⁶ Reiner Lemoine Institut gGmbH, Research Unit Transformation of Energy Systems, Rudower
Chaussee 12, Berlin, 12489, Mascha.Richter@rl-institut.de

⁷ Technische Informationsbibliothek, Data Science & Digital Libraries research group,
Welfengarten 1 B, Hannover, 30167, auer@tib.eu, 
<https://orcid.org/0000-0002-0698-2864>

⁸ Carl von Ossietzky Universität Oldenburg, Department für Informatik, Ammerländer Heerstr.
114-118, Oldenburg, 26129, astrid.niesse@uni-oldenburg.de, 
<https://orcid.org/0000-0003-1881-9172>

⁹ OFFIS e.V., R&D Division Energy, Escherweg 2, Oldenburg, 26121,
sebastian.lehnhoff@offis.de, 
<https://orcid.org/0000-0003-2340-6807>

1 Introduction

The National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, or NFDI) initiative in Germany aims to establish research data infrastructures for many scientific disciplines. Various consortia have been established to pursue this goal, each focusing on a specific research area. Among these consortia is the NFDI4Energy consortium, started in 2023, which is building research data infrastructure for energy system research. The envisioned final product will support researchers in obtaining, sharing, and understanding FAIR research data and models. [Ni22]

NFDI4Energy has divided its work into seven Task Areas (TAs) that together will create services for researchers focusing on the energy system domain (see Figure 1). TAs 1, 2, and 3 focus on the involvement of the energy research community, society, and the energy industry, respectively, to ensure that the services provided by the consortium are able to meet the needs of these diverse stakeholders. TAs 4, 5, and 6 will tackle the creation of services and infrastructure, while TA7 provides organizational support across the consortium. [Ni22]

This paper describes the objectives of TA4: FAIR Data for Energy System Research, as well as the methodology that the TA4 team will use to achieve these objectives. Section 2 provides background and motivation for the consortium's work, with an emphasis on TA4's role. Section 3 describes the artifacts that TA4 has been tasked with creating, as well as the proposed methods to create them. Finally, Section 4 concludes the paper with an overview of the TA4 stakeholders and participants.

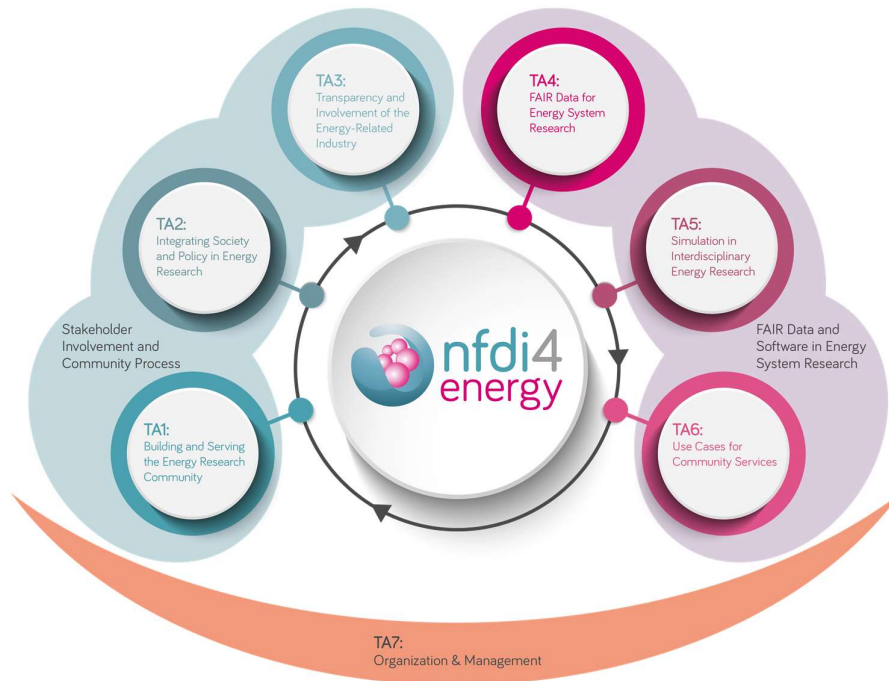


Fig. 1: The seven Task Areas (TAs) of NFDI4Energy. Figure from [Ni22].

2 Background

The FAIR Principles, first published in 2016 [Wi16], are a set of data management guidelines which the NFDI4Energy consortium aims to follow in its service development. To conform to these principles, research data and/or metadata must be:

- Findable (clearly indexed such that it can be searched for and located),
- Accessible (retrievable by standardized methods and usable in multiple contexts),
- Interoperable (using standardized terminology so that the data may be used as broadly as possible), and
- Reusable (described in detail, including provenance and usage license information, according to relevant standards).

In the domain of energy system research, FAIR data can be difficult to obtain and use. Energy systems impact and are impacted by many diverse factors, ranging from new technology adoption to governmental climate change policies, to global weather patterns, to financial structures of energy markets, and more. Studies in this domain therefore often require information from many disciplines, necessitating substantial time and effort on the part of researchers to not only track down relevant datasets, but to clean and map the data appropriately. [Ni22]

The NFDI4Energy consortium aims to ease this burden by building a data infrastructure that promotes and supports the use of FAIR data in energy system research. The consortium's platform will include services for researchers to discover best practices and methods to be used in their projects, obtain standardized datasets and models, find and collaborate with like-minded research teams, and share their work with the scientific community. [Ni22]

Within the consortium, TA4 will develop the semantic layer of this platform. Ontologies will define terminology and relationships for relevant concepts in energy system research, and metadata standards will promote the findability and reusability of research data. Various services, including a Knowledge Graph, a Terminology Service, and a Persistent Identifier (PID) service, will enable the storage and retrieval of research artifacts. These resources, when integrated into the NFDI4Energy platform, will be key components of the new data infrastructure, streamlining the work of researchers in the energy domain. [Ni22]

3 Objectives & methodology

The deliverables that TA4 will produce are summarized in Figure 2. As shown, the artifacts can be divided into two categories: semantic resources, and services & infrastructure. Each of the software services will be designed according to the terminologies and relationships defined by the semantic resources. Input from several other TAs will be incorporated into TA4's work, and TA4's work will in turn be incorporated into the tasks of other TAs, ensuring a cohesive NFDI4Energy platform.

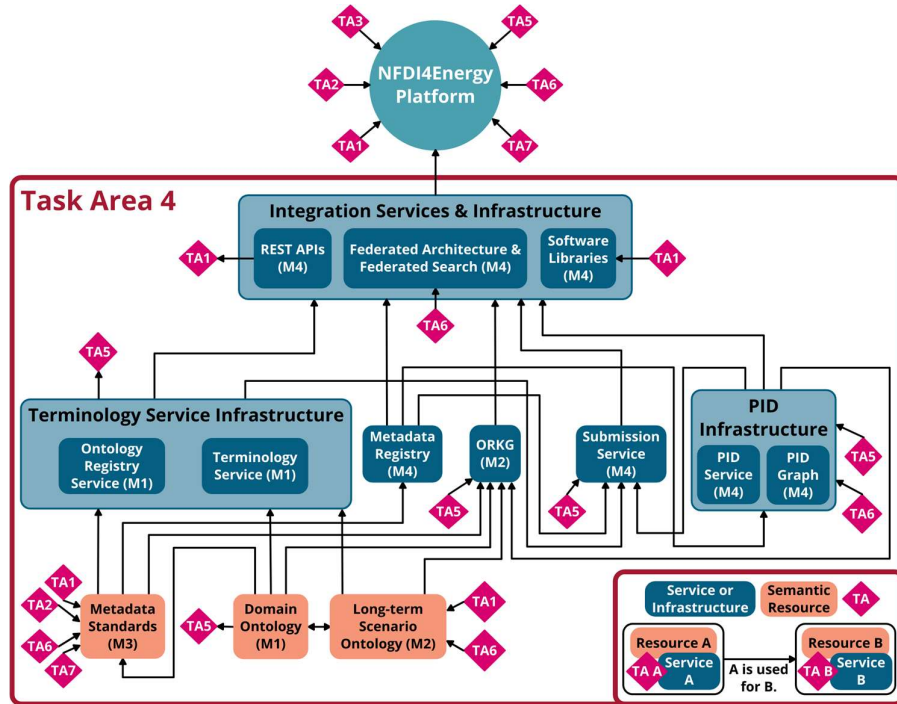


Fig. 2: The resources and services to be created by TA4. Arrows indicate how the artifacts are linked together, and also how the work of other TAs will influence the work of TA4. The tasks of TA4 are referred to internally as Measures, abbreviated here as M1 – M4.

3.1 Semantic resources

Two related ontologies will be fashioned by TA4.

A domain ontology will provide the semantic backbone for NFDI4Energy’s research data infrastructure, defining a standardized vocabulary for energy system research. This ontology will lay out concept definitions and relationships in a hierarchical arrangement. It will form a basis for research data interoperability not just within the energy domain, but also in interdisciplinary applications. [Ni22]

Building from the domain ontology, a scenario ontology will provide a vocabulary for specification of energy system scenarios. Various long-term scenarios are often used in studies examining what the power grids of the future may look like and how entities can best meet their carbon reduction goals; however, there is generally wide variation in scenario criteria (such as time frames, energy demand predictions, energy price changes,

greenhouse gas emission forecasts, etc.) from study to study. The scenario ontology aims to normalize the definitions of these criteria and simplify comparisons between scenarios, making the resulting research data more accessible to policy-makers. [Ni22]

Multiple ontologies already exist for the energy domain, such as the Open Energy Ontology [Bo21] and the Common Information Model [Mc07]. Rather than reinvent the wheel, one or more of these existing ontologies will be used as a starting point for TA4's work. The starting ontologies will be selected based on a review of their features and a comparison to features requested by energy system researchers. The most appropriate ontologies will then be combined and extended to create the NFDI4Energy domain and scenario ontologies.

Metadata standards will be developed alongside the ontologies to promote data FAIRness. These standards will utilize the terminology of the domain ontology, but will also be designed for interoperability with other research domains, as far as is possible. To that end, TA4 will collaborate closely with other TAs and other NFDI consortia in the creation of the metadata standards. As with the ontologies, existing metadata standards will be investigated and those that are suitable will be integrated into the new NFDI4Energy standards. [Ni22]

Two working groups shall be established as part of the ontology and standards development process. The Metadata & Ontologies working group will provide a forum for energy system researchers and other NFDI consortia to participate in the development of NFDI4Energy's semantic resources, aligning the ontologies and metadata standards and providing training to support the adoption of these resources. The Scenario Description Standardization working group will focus on harmonizing descriptions of energy system scenarios for interdisciplinary research, ensuring alignment with the NFDI4Energy ontologies and metadata standards. Both working groups will allow the TA4 team to receive continuous feedback from stakeholders throughout the ontology and standards development, thus making certain that the new artifacts will be well-suited to serve the needs of the scientific community. [Ni22]

3.2 Services & infrastructure

In support of FAIR research data and to enable the use of the semantic resources, TA4 will develop several software services that will be integrated into the NFDI4Energy platform. Various registries and submission services will promote data findability and accessibility by providing researchers with infrastructure for sharing data and models.

The primary means of accessing the ontologies and standards will be through a Terminology Service. This service will also support the registry of ontologies from other domains that are considered relevant for research in the energy domain. It will follow W3C standards and allow users to search for and retrieve any terminology information that is stored in the platform. [Ni22]

TA4 will also prepare an Open Research Knowledge Graph (ORKG) [Au20] for energy research. ORKGs are platforms designed to link knowledge from multiple research sources and make this knowledge machine-readable, thus simplifying the location and comparison of data. TA4's ORKG will be focused specifically on energy system scenarios, easing comparisons between scenarios to give users a clearer picture of the potential future pathways for energy systems that have been studied. [Ni22]

Two additional services will focus specifically on improving the findability and reusability of research data. A PID service, using either existing concepts such as DOI or using a new concept if needed, will assign PIDs to digital objects and link the identified objects through a PID graph. This service will be tied to a metadata registry, based on the Leibniz Data Manager [Le23], which will register digital repositories of research data from across the internet. The metadata registry will allow users to visualize and compare data sets, thus helping researchers to locate the ideal data for their work. [Ni22]

4 Conclusion

TA4's work will be closely tied to the work of several other TAs. [Ni22]

- TA1 (Building and Serving the Energy Research Community) will provide feedback from researchers to inform the development of TA4's semantic resources. Infrastructure assembled by each of these TAs will be linked together.
- TA2 (Integrating Society and Policy in Energy Research), similar to TA1, will provide stakeholder input for the semantic resources.
- TA5 (Simulation in Interdisciplinary Research) will produce artifacts for energy system simulations, which must be compatible with TA4's resources and services.
- TA6 (Use Cases for Community Services) will identify use cases for the NFDI4Energy platform that TA4 will take into account during resource development.
- TA7 (Organization & Management) will support TA4 in connecting with other NFDI consortia, to pursue the creation of resources with interdisciplinary uses.

Several organizations will be participating in the efforts of TA4, which will be lead by OFFIS e.V. Additional contributors include Albert-Ludwigs-Universität Freiburg, Friedrich-Alexander-Universität Erlangen-Nürnberg, Karlsruhe Institute of Technology – Institute for Automation and Applied Informatics, Reiner Lemoine Institut, Technische Informationsbibliothek, and Carl von Ossietzky Universität Oldenburg. [Ni22] Taken together with those who choose to participate in the working groups, TA4 will benefit from a wealth of diverse experience, viewpoints, and suggestions throughout the crafting of its artifacts.

Bibliography

- [Au20] Auer, S. et.al.: Improving Access to Scientific Literature with Knowledge Graphs. *Bibliothek Forschung und Praxis*, 3/20, pp. 516-529, 2020
- [Bo21] Booshehri, M. et.al.: Introducing the Open Energy Ontology: Enhancing data interpretation and interfacing in energy systems analysis. *Energy and AI* 5/21, 2021.
- [Le23] Leibniz Data Manager, <https://labs.tib.eu/info/en/project/leibniz-data-manager/>, accessed: 11/05/2023.
- [Mc07] McMorran, A.W.: An Introduction to IEC 61970-301 & 61968-11: The Common Information Model. University of Strathclyde 124/07, 2007.
- [Ni22] Nieße, A. et.al.: nfdi4energy – National Research Data Infrastructure for the Interdisciplinary Energy System Research, <https://doi.org/10.5281/zenodo.6772013>, accessed: 03/05/2023.
- [Wi16] Wilkinson, M. D. et.al.: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3/16, 2016.