NFDI4Energy Conference

1. NFDI4Energy Conference

10.5281/zenodo.10614814

Published: 04.02.2024

Towards a Conception and the Implementation of FAIR Data Management Tools for the Energy Domain

Karl-Uwe Stucky[https://orcid.org/0000-0002-0065-0762],
Mohamed Anis Koubaa[https://orcid.org/0000-0001-8552-2008],
Nan Liu[https://orcid.org/0009-0005-8768-7072],
Andreas Schmidt[https://orcid.org/0000-0002-9911-5881],
Wolfgang Süß[https://orcid.org/0000-0003-2785-7736], and
Veit Hagenmeyer[https://orcid.org/0000-0002-3572-9083]

Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Germany
*Correspondance: Karl-Uwe Stucky, uwe.stucky@kit.edu

Keywords: FAIR Digital Object, Ontology, Conceptual Modeling, Knowledge Graph

1 Introduction

For establishing a research data infrastructure, it is essential to enrich the data with meta information belonging to three metadata types: descriptive, provenance, and structural metadata. It is envisaged to base this metadata on ontologies and integrate them into Knowledge Graphs (KGs). Data management should adhere to the FAIR principles [1]. Researchers should be provided with software tools for easy-to-use search capabilities and FAIR data access. Tools should also support the integration of metadata and ontology standards and in general, the creation of new metadata types has to be facilitated. We have identified two main constituents of an appropriate software development, which will be outlined here. In the end, a fully integrated FAIR data management system will be desirable.

Firstly, we focus on Conceptual Modeling and an object-oriented approach, inspired by literature like [2], [3], and [4] that compare ontologies and Conceptual Models (CMs). In view of the software developing goal, we are going to base our work on the similarities between both methods. As for making the approach FAIR, the second focus of this work lies on FAIR Digital Objects (FDOs) [5] and Persistent Identifiers (PIDs) and how they shall be applied together with KGs and the underlying models.

2 Ontologies and Conceptual Modeling

In [2] and [3], ontologies are seen as being relatively generic in the sense of being application independent. Furthermore, CMs are regarded as being application-specific and implementation-oriented, and are expressed e.g. by object-oriented programming languages [4]. The authors in [3] also argue that there is no strict line between both levels. Ontologies provide vocabularies with terms to designate their concepts as well as taxonomies forming class hierarchies. They also model relationships between the concepts.

According to these characteristics we stress the similarities and state a flowing transition between ontologies of general domains that span across various specific subject domains and the CMs describing them (see also Figure 1). These application-specific domains are the areas for research, data analysis, and operational software close to the relevant real world entities. They are associated with ontology domains [4] and, in principle, share the same kind of semantic structure of classes, inheritance and relationships between classes. We find the same structures in instances of class models or schemas. Classes of the ontologies, CMs for specific subject domains, and the instance models establish a KG that can be traversed searching for associated research data.

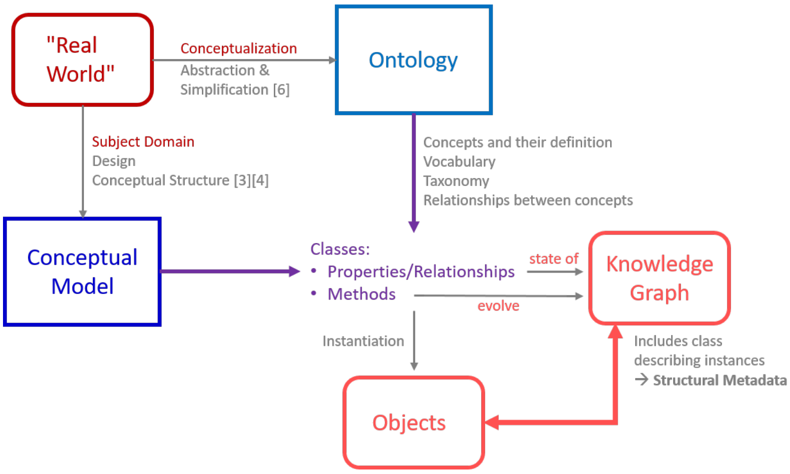

Figure 1. Ontologies and Conceptual Models [3][4][6]

3 FAIR Digital Objects

In the next step, we look at FDOs and how they are designed in terms of PIDs, all kinds of metadata, and of digital resources for which they are a publishing element and access point (see Figure 2). As all KG objects, FDOs possess each a PID and a link to the respective schema data (structural metadata). A link to the digital resource, the FDO's payload, is not given directly but encapsulated in an object called *data registry entry*. Additionally, it encapsulates a fourth kind of metadata, administrative metadata which are introduced provisionally as *location* and *access regulations*. The descriptive and provenance metadata, which are relevant for data searches, directly reference respective KG objects. Arbitrary numbers of descriptive metadata are stored in a collection, provenance data may point to a single entrance object to a provenance graph or to various separate provenance objects.

Our FDO design can be compared with the HMC preferred FDO that contains an extra set of primitive-typed single metadata called Kernel Information Profile (KIP) [7]. This main difference is not very prominent because each KIP property can be easily transferred to our metadata collection and both can be made compatible. Furthermore, the approach given here can be extended by admitting a hierarchy of FDO classes inheriting from an abstract base FDO type. Thus, many FDO designs exist in parallel and nevertheless seamlessly integrate themselves in the surrounding conception. Consensus on a minimal set of metadata (like KIP) may not be accepted everywhere and for

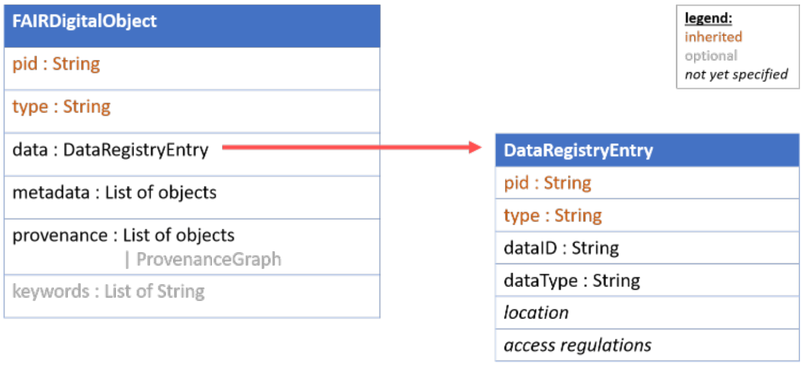

Figure 2. FDO and its connection to data

all times. Such restrictions should be decided on the implementation level, where the application of specific FDO types can be forced.

Regarding identification, we suggest to rely on a PID system designed in accordance to the Handle System [8], for the first time described in [9], or even build upon this system right from the start. It is crucial to have PIDs without any semantics. They are just unique and permanent.

4 Example Use Case: PV System

In [10], an application for managing and accessing a photovoltaic (PV) system and its sensor data at KIT is presented. The concept defines FDOs both for the system and its components, as well as for further metadata objects. These further metadata are given by third-party standards (IEC 61850[1], GeoJSON[2], and SensorML[3]). A proprietary PV ontology describes the PV system components and their composition.

The PV system has now been chosen as the first use case to accompany our development work. As a first step in realizing the proposed ontology- and FDO-based energy data management, it is planned to substitute the various FDOs representing single metadata objects with a single FDO representing a measured data set or other digital resource and referencing relevant metadata. Furthermore, the third-party objects will be integrated in the PV system components directly.

It is envisaged to build on the standards JSON[4] for serialization, RDF[5]/OWL[6], and JSON-LD[7] as data formats, SHACL[8] for defining constraints on RDF graphs, SPARQL[9] and SQL as database access languages, and on the programming languages Java and Python.

[1] https://webstore.iec.ch/publication/6028
[2] https://geojson.org/
[3] https://www.ogc.org/standards/sensorml
[4] https://www.json.org/json-en.html
[5] https://www.w3.org/RDF/
[6] https://www.w3.org/OWL/
[7] https://www.w3.org/TR/json-ld/
[8] https://www.w3.org/TR/shacl/
[9] https://www.w3.org/TR/sparql11-query/

5 Conclusion

This work on FAIR energy data management proposes to focus on Conceptual Models as an extension of a purely ontological view on energy metadata. It aims on a uniform object-oriented software design for data management and useful tools, e.g. for creating new metadata. A Knowledge Graph connects - especially the descriptive - metadata objects, and the overall system connects them with the research data, the administrative and the structural metadata. Fair Digital Objects (FDO) are connected to and thereby extending the KG. They also reference digital resources, mainly data sets from measurements or simulations. All KG node elements are also connected to structural metadata describing their classes. Data search and retrieval is done via the FDOs as an alternative to navigating across the KG.

Author contributions

Karl-Uwe Stucky: Conceptualization, Investigation, Methodology, Writing – original draft
Andreas Schmidt: Conceptualization, Writing – review & editing
Veit Hagenmeyer, Anis Koubaa, Nan Liu, Wolfgang Süß: Writing – review & editing

Competing interests

The authors declare that they have no competing interests.

Funding

The authors' work on ontologies is conducted as a contribution to the NFDI4Energy project in the German National Research Data Infrastructure, a special funding initiative of the Deutsche Forschungsgemeinschaft DFG.

Acknowledgements

This publication was supported within the Hub Energy of the Helmholtz Metadata Collaboration (HMC).

References

[1] M. D. Wilkinson, M. Dumontier, I. J. J. Aalbersberg, *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, 2016. DOI: 10.1038/sdata.2016.18.

[2] P. Spyns, R. Meersman, and M. Jarrar, "Data modelling versus ontology engineering," *ACM SIGMOD Record*, vol. 31, no. 4, p. 12, 2002, ISSN: 01635808. DOI: 10.1145/637411.637413.

[3] M. Jarrar, J. Demey, and R. Meersman, "On using conceptual data modeling for ontology engineering," in *Journal on Data Semantics I*, S. Spaccapietra, S. March, and K. Aberer, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 185–207, ISBN: 978-3-540-39733-5. DOI: 10.1007/978-3-540-39733-5{\textunderscore}8. [Online]. Available: https://doi.org/10.1007/978-3-540-39733-5_8.

[4] M. R. Kogalovsky and L. A. Kalinichenko, “Conceptual and ontological modeling in information systems,” *Programming and Computer Software*, vol. 35, no. 5, pp. 241–256, 2009, ISSN: 0361-7688. DOI: 10.1134/S0361768809050016.

[5] S. Collins, F. Genova, N. Harrower, S. Hodson, S. Jones, L. Laaksonen, D. Mietchen, R. Petrauskaite, P. Wittenburg, *Turning FAIR data into reality: Final report and action plan from the European Commission expert group on FAIR data*. Luxembourg: Publications Office of the European Union, 2018, ISBN: 9279965468. DOI: 10.2777/1524.

[6] T. R. Gruber, “A translation approach to portable ontology specifications,” *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, 1993, ISSN: 10428143. DOI: 10.1006/knac.1993.1008.

[7] C. Curdt, *Helmholtz Kernel Information Profile*. HMC Office, GEOMAR Helmholtz Centre for Ocean Research Kiel, 2022. DOI: 10.3289/HMC_publ_03.

[8] S. X. Sun, L. Lannom, and B. P. Boesch, “Handle system overview,” *RFC*, vol. 3650, pp. 1–21, 2003. [Online]. Available: https://www.rfc-editor.org/rfc/rfc3650.txt (visited on 11/05/2023).

[9] R. Kahn and R. Wilensky, “A framework for distributed digital object services,” *International Journal on Digital Libraries*, vol. 6, no. 2, pp. 115–123, 2006, ISSN: 1432-1300. DOI: 10.1007/s00799-005-0128-x. [Online]. Available: https://doi.org/10.1007/s00799-005-0128-x.

[10] J. Schweikert, K.-U. Stucky, W. Süß, and V. Hagenmeyer, “A photovoltaic system model integrating fair digital objects and ontologies,” IEEE, 2022. DOI: 10.1109/sege55279.2022.9889768.