

NFDI4DSO: Towards a BFO Compliant Ontology for Data Science

Genet Asefa Gesese^{1,2,*}, Jörg Waitelonis^{1,2}, Zongxiong Chen³, Sonja Schimmler³ and Harald Sack^{1,2}

¹*FIZ Karlsruhe, Leibniz Institute for Information Infrastructure, Germany*

²*Karlsruhe Institute of Technology, KIT, Germany*

³*Fraunhofer FOKUS, Berlin, Germany*

Abstract

The NFDI4DataScience (NFDI4DS) project aims to enhance the accessibility and interoperability of research data within Data Science (DS) and Artificial Intelligence (AI) by connecting digital artifacts and ensuring they adhere to FAIR (Findable, Accessible, Interoperable, and Reusable) principles. To this end, this poster introduces the NFDI4DS Ontology, which describes resources in DS and AI and models the structure of the NFDI4DS consortium. Built upon the NFDICore ontology and mapped to the Basic Formal Ontology (BFO), this ontology serves as the foundation for the NFDI4DS knowledge graph currently under development.

Keywords

Data Science, Artificial Intelligence, Ontology, Knowledge Graph, NFDI4DS

1. Introduction

The German National Research Data Infrastructure (NFDI)¹ is a non-profit association founded to coordinate the activities for establishing a national research data infrastructure. It comprises 26 consortia spanning a wide range of scientific disciplines, from cultural sciences, social sciences, humanities and engineering to life sciences and natural sciences. The NFDI consortia share common goals and concepts, such as their members, structure, data repositories, and services [1]. To enhance interoperability across these consortia, the NFDICore ontology² has been developed. It acts as a mid-level ontology for representing metadata related to NFDI

SEMANTICS 2024: 20th International Conference on Semantic Systems, September 17–19, 2024, Amsterdam, The Netherlands

*Corresponding author.

✉ genet-asefa.gesese@fiz-karlsruhe.de (G. A. Gesese); Joerg.Waitelonis@fiz-Karlsruhe.de (J. Waitelonis);

zongxiong.chen@fokus.fraunhofer.de (Z. Chen); sonja.schimmler@fokus.fraunhofer.de (S. Schimmler);

harald.sack@fiz-karlsruhe.de (H. Sack)

🌐 <https://tinyurl.com/3cx37b9x> (G. A. Gesese); <https://shorturl.at/UwDND> (J. Waitelonis);

<https://www.fokus.fraunhofer.de/009785fd54551039> (S. Schimmler); https://www.aifb.kit.edu/web/Harald_Sack (H. Sack)

🆔 0000-0003-3807-7145 (G. A. Gesese); 0000-0001-7192-7143 (J. Waitelonis); 0000-0003-2452-0572 (Z. Chen);

0000-0002-8786-7250 (S. Schimmler); 0000-0001-7069-9804 (H. Sack)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹<https://www.nfdi.de/>

²<https://ise-fizkarlsruhe.github.io/nfdicore/2.0.0/>

resources, including individuals, organizations, projects, data portals, and more. NFDICore provides mappings to a broad range of standards across different domains, such as the Basic Formal Ontology (BFO) [2] and Schema.org [3] to advance knowledge representation, data exchange, and collaboration across diverse domains. To address domain-specific research questions for each consortium, NFDICore follows a modular architecture. Examples for modular extensions include the NFDI4Culture ontology module CTO³[4] and the NFDI-MatWerk ontology module MWO⁴, which are specifically designed for the cultural heritage and materials science domains, respectively. In this paper, we present an ontology named NFDI4DSO for the data science domain as a domain-specific modular extension of NFDICore.

NFDI4DataScience (NFDI4DS)⁵ is one of the NFDI consortia and its project aims to enhance the accessibility and interoperability of research data in the domain of Data Science (DS) and Artificial Intelligence (AI). Data Science (DS) is a multidisciplinary field combining different aspects of mathematics, statistics, computer science, and domain-specific knowledge to extract meaningful insights from diverse data sources. DS and Artificial Intelligence (AI) involve various artifacts, e.g., datasets, models, ontologies, code repositories, execution platforms, repositories, etc. The project achieves this by linking digital artifacts and ensuring their FAIR (Findable, Accessible, Interoperable, and Reusable) accessibility, thereby fostering collaboration across various DS and AI platforms. To this end, the NFDI4DS Ontology (NFDI4DSO) is built.

2. The NFDI4DataScience Ontology (NFDI4DSO)

As mentioned earlier, NFDI4DSO is created in a modular fashion, building upon NFDICore. Similar to NFDICore, the NFDI4DSO ontology is developed using a bottom-up, iterative, user-centered approach. NFDICore comprises 51 classes, 55 object properties, 8 data properties, 18 annotation properties, and 5 SWRL rules [5] (for details refer to NFDICore documentation⁶). In NFDI4DSO, in addition to what is provided in NFDICore, 42 classes, 38 object properties, 9 data properties, and 8 SWRL rules are added. The NFDI4DSO ontology not only describes various data science artifacts but also provides information about the resources of the NFDI4DS Consortium, such as personas, consortium members, spokespersons, and task area leads. AS in NFDICore, the classes introduced in NFDI4DSO are also mapped to the top-level ontology BFO and also other ontologies such as schema.org, the FaBiO ontology [6], and the Conference Ontology⁷.

NFDI4DSO contains various kinds of classes such as processes, roles, and independent continuants. For instance, Figure 1 depicts how NFDI4DSO represents the relationship between the independent continuant *nfdi4dso:SonjaSchimmler* and her spokesperson role *nfdi4dso:SpokespersonRole* by mapping it to BFO. By using roles and processes, NFDI4DSO enables a detailed representation of the relationship between different entities enhancing the ontology's level of expressivity. On the other hand, to support easier integration and use of less complex relations, shortcuts are also introduced to simplify the ontology by implementing easy-to-use

³<https://gitlab.rlp.net/adwmainz/nfdi4culture/knowledge-graph/culture-ontology>

⁴<https://git.rwth-aachen.de/nfdi-matwerk/ta-oms/mwo>

⁵<https://www.nfdi4datascience.de/>

⁶<https://ise-fizkarlsruhe.github.io/nfdicore/>

⁷<http://www.scholarlydata.org/ontology/doc/#toc>

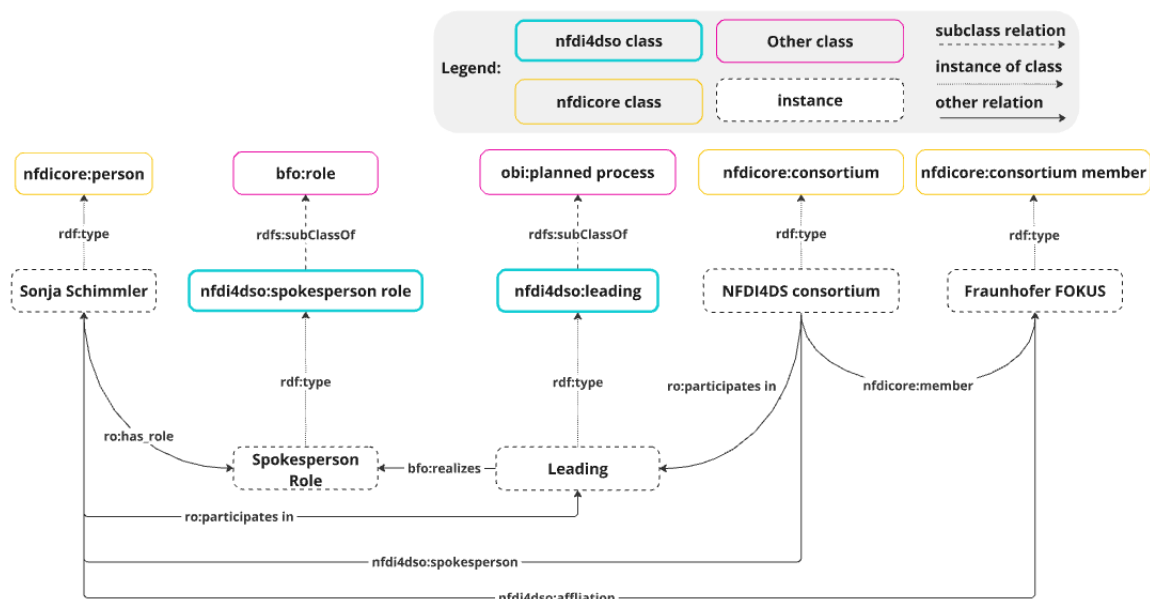


Figure 1: Example of representing roles where the prefixes ro and obi represent <http://purl.obolibrary.org/obo/ro.owl> and <http://purl.obolibrary.org/obo/obi.owl> ontologies, respectively.

direct shortcut properties, which can be expanded to fully-fledged BFO-compliant complex path expressions. For instance, in Figure 1, the shortcut relation *nfdi4dso:spokesperson* is provided and its corresponding SWRL⁸ rule is given below.

$$Person(?p) \wedge Consortium(?c) \wedge SpokespersonRole(?sr) \wedge Leading(?l) \wedge participates\ in(?p, ?l) \wedge participates\ in(?c, ?l) \wedge has\ role(?p, ?sr) \wedge realized\ in(?sr, ?l) \rightarrow spokesperson(?c, ?p)$$

Ontology Implementation The Protégé ontology editor⁹ for the OWL-based formalization of terminological knowledge has been used to develop and implement NFDI4DSO. Widoco¹⁰ has been used to create an enriched and customized documentation of the ontology automatically. The stable version of the ontology NFDI4DSO v1.0.0 is available at <https://github.com/ISE-FIZKarlsruhe/NFDI4DS-Ontology/tree/main> and the latest development version is at <https://github.com/ISE-FIZKarlsruhe/NFDI4DS-Ontology/tree/develop-1.0.1>.

3. NFDI4DSO in Use

The NFDI4DSO is designed to form the foundation of the NFDI4DS Knowledge Graph (NFDI4DS-KG), which is currently under development. The NFDI4DS-KG consists of two main components: the Research Information Graph (RIG) and the Research Data Graph (RDG). RIG includes metadata about the NFDI4DS consortium’s resources, persons, and organizations, while the RDG encompasses content-related index data from the consortium’s heterogeneous data sources.

⁸<https://ise-fizkarlsruhe.github.io/NFDI4DS-Ontology/#d4e7620>

⁹<https://protege.stanford.edu/>

¹⁰<https://github.com/dgarijo/Widoco>

?s	https://nfdi.fiz-karlsruhe.de/nfdi4dso/coSpokesperson	?e
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O nfdi4dso:Harald_Sack
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O nfdi4dso:Claudia_Wagner
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O nfdi4dso:Volker_Markl
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O nfdi4dso:Christoph_Lange-Bever
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O nfdi4dso:Marcel_R_Ackermann
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O nfdi4dso:Michael_Wagner
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O nfdi4dso:Klaus_Tochtermann
S · O NFDI4DS Consortium	S · O co-spokesperson	S · O Georg_Rehm

Figure 2: A screenshot of part of the SHMARQL interface with the list of NFDI4DS co-spokespersons (refer to <https://shorturl.at/eNb5e> to navigate it fully.)

RIG serves as the backend for the NFDI4DS web portal, facilitating interactive access and management of this data. Both RIG and RDG will be accessible and searchable via the NFDI4DS Registry platform. Additionally, the NFDI4DS consortium plans to collaborate with other NFDI consortia to further integrate domain-specific knowledge into the RDG seamlessly. Currently, the first version of the NFDI4DS-KG¹¹ with RIG is publicly available. For example, to view the list of co-spokespersons of the NFDI4DS Consortium, you can either navigate through the data using SHMARQL¹², as depicted in Figure 2 or query it using SPARQL, as shown in Figure 3.

4. Conclusion and Future Work

This paper presents the NFDI4DS Ontology and its use for the NFDI4DS-KG that is currently under-development. The ontology facilitates the representation and interoperability of data science artifacts within and outside of NFDI4DS. NFDI4DSO is built on top of the NFDICore ontology and mapped to BFO and other ontologies. In the future, there is a plan to perform extensive ontology evaluation using competency questions based on the persona definitions from the NFDI4DS consortium.

Acknowledgments

This publication was written by the NFDI consortium NFDI4DataScience in the context of the work of the association German National Research Data Infrastructure (NFDI) e.V.. NFDI is financed by the Federal Republic of Germany and the 16 federal states and funded by the Federal Ministry of Education and Research (BMBF) – funding code M532701 / the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number NFDI4DataScience (460234259).

¹¹<https://nfdi.fiz-karlsruhe.de/4ds/sparql>, <https://nfdi.fiz-karlsruhe.de/4ds/shmarql>

¹²<https://shorturl.at/eNb5e>

```

1 PREFIX foaf: <http://xmlns.com/foaf/0.1/>
2 PREFIX nfdi4dso: <https://nfdi.fiz-karlsruhe.de/nfdi4dso/>
3 SELECT (?p as ?CoSpokesperson) (?f as ?FirstName)
4 (?l as ?LastName) (?a as ?Affiliation)(?o as ?ORCID) WHERE {
5   nfdi4dso:NFDI4DS nfdi4dso:coSpokesperson ?p .
6   ?p nfdi4dso:hasORCID ?o.
7   ?p nfdi4dso:affiliation ?a.
8   ?p foaf:firstName ?f .
9   ?p foaf:lastName ?l .
10 }

```

Table Response 22 results in 0.038 seconds Simple view

	CoSpokesperson	FirstName	LastName	Affiliation	ORCID
1	nfdi4dso:Harald_Sack	Harald	Sack	nfdi4dso:KIT	0000-0001-7069-9804
2	nfdi4dso:Harald_Sack	Harald	Sack	nfdi4dso:FIZ	0000-0001-7069-9804
3	nfdi4dso:Volker_Markl	Volker	Markl	nfdi4dso:TUB	0009-0009-0964-026X
4	nfdi4dso:Christoph_Lange-Bever	Christoph	Lange-Bever	nfdi4dso:RWTH	0000-0001-9879-3827
5	nfdi4dso:Christoph_Lange-Bever	Christoph	Lange-Bever	nfdi4dso:FIT	0000-0001-9879-3827
6	nfdi4dso:Marcel_R_Ackermann	Marcel R.	Ackermann	nfdi4dso:LZI	0000-0001-7644-2495
7	nfdi4dso:Klaus_Tochtermann	Klaus	Tochtermann	nfdi4dso:CAU	0000-0003-2471-2697
8	nfdi4dso:Klaus_Tochtermann	Klaus	Tochtermann	nfdi4dso:ZBW	0000-0003-2471-2697
9	nfdi4dso:Georg_Rehm	Georg	Rehm	nfdi4dso:DFKI	0000-0002-7800-1893
10	nfdi4dso:Markus_Stocker	Markus	Stocker	nfdi4dso:TIB	0000-0001-5492-3212
11	nfdi4dso:Adamantios_Koumpis	Adamantios	Koumpis	nfdi4dso:RWTH	0000-0003-2661-7749
12	nfdi4dso:Dietrich_Rebholz-Schuhm...	Dietrich	Rebholz-Schuh...	nfdi4dso:ZB_MED	0000-0002-1018-0370
13	nfdi4dso:Dietrich_Rebholz-Schuhm...	Dietrich	Rebholz-Schuh...	nfdi4dso:UzK	0000-0002-1018-0370
14	nfdi4dso:Thomas_Neumuth	Thomas	Neumuth	nfdi4dso:ULEI	0000-0001-6999-5024

Figure 3: An example SPARQL query to provide a list of the co-spokespersons of the NFDI4DS Consortium. (It possible to query it live at: <https://nfdi.fiz-karlsruhe.de/4ds/sparql>)

References

- [1] H. Sack, T. Schrade, O. Bruns, E. Posthumus, T. Tietz, E. Norouzi, J. Waitelonis, H. Fliegl, L. Söhn, J. Tolksdorf, J. Steller, A. Azócar Guzmán, S. Fathalla, A. Zainul Ihsan, V. Hofmann, S. Sandfeld, F. Fritzen, A. Laadhar, S. Schimmler, P. Mutschke, Knowledge Graph Based RDM Solutions: NFDI4Culture-NFDI-MatWerk-NFDI4DataScience, in: 1st Conf. on Research Data Infrastructure, 2023.
- [2] J. N. Otte, J. Beverley, A. Ruttenberg, BFO: Basic formal ontology, Applied ontology (2022).
- [3] R. V. Guha, D. Brickley, S. Macbeth, Schema. org: evolution of structured data on the web, Communications of the ACM (2016).
- [4] T. Tietz, O. Bruns, L. Söhn, J. Tolksdorf, E. Posthumus, J. J. Steller, H. Fliegl, E. Norouzi, J. Waitelonis, T. Schrade, H. Sack, From Floppy Disks to 5-Star LOD: FAIR Research Infrastructure for NFDI4Culture, in: DaMaLOS, co-located with ESWC 2023, 2023.
- [5] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosf, M. Dean, et al., Swrl: A semantic web rule language combining owl and ruleml, W3C Member submission (2004).
- [6] S. Peroni, D. Shotton, FaBiO and CiTO: Ontologies for describing bibliographic resources and citations, Journal of Web Semantics (2012).