nfdi
Nationale
Forschungsdaten
Infrastruktur

Section
(Meta)data,
Terminologies,
Provenance

Working Group Charter
Terminology Services

Name of the working group

Terminology Services

Acronym

section-metadata-wg-ts

Contact persons

Roman Baum (Deutsche Zentralbibliothek für Medizin; NFDI4Health)
baum@zbmed.de

Luca Ghiringhelli (Fritz-Haber-Institut der Max-Planck-Gesellschaft; FAIRmat)
ghiringhelli@fhi-berlin.mpg.de

Naouel Karam (Fraunhofer FOKUS; NFDI4Biodiversity)
karam@infai.org

Authors

Ivonne Anders (0000-0001-7337-3009), Kolja Bailly, Roman Baum (0000-0001-5246-9351), Felix Engel (0000-0002-3060-7052), Felix Ernst (0000-0002-2102-4170), Luca M. Ghiringhelli (0000-0001-5099-3029), Martin Golebiewski (0000-0002-8683-7084), Angela Kailus (0000-0003-1514-2258), Naouel Karam (0000-0002-6762-6417), Jürgen Kett (0000-0001-9421-1447), Florian Kraus, Matthias Löbe (0000-0002-2344-0426), Katrin Moeller (0000-0003-4090-5667), Akhil Patil (0000-0003-0035-8427), Lozana Rossenova (0000-0002-5190-1867), Jakob Voß (0000-0002-7613-4123), Claus Weiland (0000-0003-0351-6523), Mehtap Özaslan (0000-0001-8545-7576)

Version 1.0

2022-06-25

Summary

The core of FAIR metadata for research data management (RDM) are terminologies. Terminologies have various distinct characteristics: they are often domain specific (e.g. Chemistry, Biology or Information Science use their own set of terminologies), and they evolve continuously over time and they must be accepted and aligned (mapping) across communities. The overarching goal of the working group "Terminology Services" is to avoid siloed developments, share knowledge and reduce the barrier of using terminologies.

Zusammenfassung

Das Kernstück der FAIRer Metadaten für die Verwaltung von Forschungsdaten (RDM) sind Terminologien. Terminologien weisen verschiedene Merkmale auf: Sie sind oft bereichsspezifisch (z. B. verwenden Chemie, Biologie oder Informationswissenschaft ihre eigenen Terminologien) und sie entwickeln sich im Laufe der Zeit ständig weiter und müssen gemeinschaftsübergreifend akzeptiert und abgestimmt werden (gemappt). Das übergeordnete Ziel der Arbeitsgruppe "Terminology Services" ist es, isolierte Entwicklungen zu vermeiden, Wissen zu teilen und die Hürden für die Verwendung von Terminologien zu reduzieren.

1. Motivation

Semantic interoperability is a cornerstone of implementing the FAIR principles. Within that concept, we can identify several distinct but interrelated components that can benefit from cross-NFDI coordination: the harmonization and/or mapping of terminologies, the delivery of FAIR terminology software services, as well as the deployment and running of domain-specific knowledge graphs based on interoperable technical frameworks.

The core of metadata for research data management (RDM) are terminologies. Terminologies have various distinct characteristics: they are often domain specific (e.g. Chemistry, Biology or Information Science use their own set of terminologies), and even community specific, they evolve continuously over time and they must be accepted and aligned (mapping) across communities to avoid siloed developments.

A terminology service is a web based platform that addresses all above introduced characteristics and makes these vocabularies accessible for third party RDM applications. Some terminology services have been developed and deployed across different NFDI consortia. These services make use of various base technologies and are not interconnected and hence not fully aligned to each other. Essentially, the various services from the different NFDI consortia form information silos again as long as they cannot be linked together and used uniformly. The working group (WG) will focus on terminology services which provide terms with FAIR stable identifiers (e.g. IRIs)

2. Objectives

The overarching goal of the working group "Terminology Services" is an alignment of the existing terminology services across NFDI consortia to counteract siloed solutions and design and comprise the semantic key services to enable efficient cross-domain ontology use in the RDM. To reach this objective the following sub-objectives must be met:

- Provide a centralized consortia-agreed, federated terminology service access point to align the various services on an infrastructure level. In cooperation with section "Common Infrastructures", WG "Ontology Harmonization and Mapping" and WG "Search and Harvesting": provision of requirements for terminology services to be used in NFDI, that is platform independent service interfaces for access and curation of a broad set of structured terminology resources. We assume a heterogeneous situation for the different disciplines.
- Implementation, documentation and deployment of a federated terminology service access point conforming to the common standards
- Introduction of service alignment processes across the NFDI consortia
- Implementation of general functionalities
  - uniform terminology access
  - subscription to terminologies
  - semantification of unstructured (e.g. tabular) data and semantic enrichment of textual (automatic annotation/tagging/indexing)
  - facilitate extensibility of available terminologies
  - access to different versions of terminologies
  - provide interoperability for terminologies in different languages

- provide best practices and tools for assessing and improving the correctness and validity of terminologies. This includes, in particular, processing quality metrics such as those of consistency, comprehensiveness, conciseness, and usability, in a feedback loop with the involved consortia.

3. Work Plan

Q3 – Q4 2022: Evaluating the state of the art
- Collect and assess used terminology services within the NFDI consortia (and beyond)
- Evaluate ways to get measurements on how often a service or a specific resource of this service was used/requested
- Evaluate ways to provide Crowd-Engineering approaches (e.g., provide a form to extend a terminology and use the ORCID for provenance data to a specific concept)

Q1 2023: Overviews
- Create an overview of the service landscape and relevant activities, in and beyond all consortia

Q2 2023: Starting implementation, piloting and initializing the feedback process
- Coordinate cross-domain subgroups for skill sharing
- Provide a collaborative platform (e.g. git-repository)
- Coordinate cross-domain subgroups which will organize testing sessions and feedback gathering with concrete software services developed during different stages of the implementation milestone

Q2 – Q3 2023: Communication activities
- Give talks and provide documentation to working groups, NFDI consortias, and other groups of interests to get feedback
- Regular updates about WG activities
- Milestone NFDI Conference and Section Workshops Q2/Q3 2023

Q3 – Q4 2023: Dissemination and documentation of outcomes
- Publication of a white paper summarizing the WG activities
- Talks with WGs, NFDI consortias, and other groups of interests

Specific tasks:
- Survey of existing NFDI terminology services and functional offers. Oversee the delivery of a suite of tools that meet the requirements of providing terminology services across different knowledge domains, if applicable in collaboration with other WGs and sections, e.g., "Common infrastructure".
- Survey of existing terminology standards and services
- Provision of a scalable, federated infrastructure architecture and deployment plan
- Provision of a registry for terminology services involved including a set of minimal specifications for data models and APIs to exchange and access terminologies.
- Implementation and adjustment of terminology services to establish the federated infrastructure

- Survey of existing terminology service functional offer and identification of overarching functionalities. Ways to create a harmonized micro service for terminology services from different sources
- Integration of terminology/ontology mappings and alignments from other WGs and sections
- Provision of a versioning functionality for terminologies
- Evaluation of the provision for additional functionalities (e.g. contact to terminology creators)

4. Initial Membership List

Name	Consortium	Role in Workgroup	Email
Ivonne Anders	NFDI4Earth	Contributor	anders@dkrz.de
Kolja Bailly	NFDI4Culture	Contributor	Kolja.Bailly@tib.eu
Roman Baum	NFDI4Health	Co-Spokesperson	baum@zbmed.de
Felix Engel	NFDI4Ing	Contributor	felix.engel@tib.eu
Felix Ernst	NFDI4Ing	Contributor	felix.ernst@kit.edu
Luca Ghiringhelli	FAIRmat	Co-Spokesperson	ghiringhelli@fhi-berlin.mpg.de
Martin Golebiewski	NFDI4Health	Contributor	Martin.Golebiewski@h-its.org
Angela Kailus	NDFI4Culture	Contributor	angela.kailus@staff.uni-marburg.de
Naouel Karam	NFDI4Biodiversity	Co-Spokesperson	karam@infai.org
Jürgen Kett	Text+ NDFI4Culture	Contributor	j.kett@dnb.de
Florian Kraus	GHGA	Contributor	florian.kraus@qbic.uni-tuebingen.de
Matthias Löbe	NFDI4Health	Contributor	matthias.loebe@uni-leipzig.de
Katrin Moeller	NFDI4Memory	Contributor	katrin.moeller@geschichte.uni-halle.de
Akhil Patil	NFDI4Ing	Contributor	akhil.patil@dlr.de
Lozana Rossenova	NFDI4Culture	Contributor	Lozana.Rossenova@tib.eu
Jakob Voß	NFDI4Objects	Contributor	voss@gbv.de

Claus Weiland	NFDI4Earth	Contributor	cweiland@senckenberg.de
Mehtap Özaslan	NFDI4Cat	Contributor	m.oezaslan@tu-braunschweig.de

5. Adoption Plan

To ensure the success of the working group, acceptance in the consortia is elementary. To reach this goal the working group will provide support in the integration process of terminology services for use cases from the consortia. Due to this approach the use of terminology services in the consortia should increase.

Additionally the working group will get important feedback from the consortia about the terminology services.

Adoption plans for specific terminology services could differ, depending on the features and weaknesses of the specific service.

Represented Consortia

NFDI4Health
NFDI4Ing
NFDI4Cat
NFDI4Biodiversity
FAIRmat
NFDI4Earth
NFDI4Culture
NFDI4Objects
NFDI4Immuno
GHGA
Text+
NFDI4Memory

Connected to other NFDI working groups

Search & Harvesting
Semantic Interoperability: Ontology harmonization & mapping
Cookbooks, Guidance, Best Practices
Knowledge Graphs
Data Integration in section "Common Infrastructures"

Connected to other working groups outside NFDI

- Research Data Alliance Vocabulary Services Interest Group (VSSIG)
- Basic Register of Thesauri, Ontologies & Classifications (BARTOC)
- Linked Open Vocabularies (LOV)
- Networked Knowledge Organization Systems/Services/Structures (NKOS)