



Section
(Meta)data,
Terminology,
Provenance

Working Group
Research Software
Metadata

Working Group Charter

Research Software Metadata

Version 3, 01.12.2023

Name of the working group

Research Software Metadata

Acronym

section-metadata-wg-rsmeta

Contact persons

Stephan Ferenz, Leyla Jael Castro

Authors

Leyla Jael Castro (0000-0003-3986-0510)

Stephan Ferenz (0000-0001-9523-7227)

Marc Fuhrmans (0000-0002-9826-018X)

Jan Göpfert (0000-0003-4722-4012)

Dorothea Iglezakis (0000-0002-8524-0569)

Oliver Karras (0000-0001-5336-6899)

Alexander Struck (0000-0002-1173-9228)

Summary

Research software is commonly used in various disciplines to perform research and as a research object itself. Therefore, it is covered by multiple NFDI consortia. To fulfill the FAIR criteria for research software, it needs to be described with meaningful and interoperable metadata enhancing transparency, reproducibility, and reusability in research. While there exist some metadata schemes for research software, e.g., CodeMeta, they have some limitations, e.g., missing elements and missing semantic interoperability.

To address these issues, this working group aims to provide a comprehensive metadata vocabulary for research software, compatible with existing frameworks such as schema.org and CodeMeta. Also, the working group will support all NFDI consortia in applying the vocabulary as well as develop domain-specific extensions if needed.

Zusammenfassung

Forschungssoftware wird in vielen Bereichen der Wissenschaft zur Forschung genutzt und stellt teilweise selbst ein Forschungsobjekt dar. Viele NFDI-Konsortien decken Forschungssoftware als wichtiges Forschungsartefakt mit ab und wollen, dass Forschungssoftware in Zukunft die FAIR-Kriterien erfüllt, um die Transparenz, Reproduzierbarkeit und Wiederverwendbarkeit in software-basierter Forschung zu erhöhen. Damit Forschungssoftware die FAIR-Kriterien erfüllt, muss sie mit aussagekräftigen und interoperablen Metadaten beschrieben werden. Dafür werden entsprechende Metadaten schemata benötigt. Es gibt zwar einige Metadaten schemata für Forschungssoftware, z. B. CodeMeta, aber sie weisen einige Einschränkungen auf, z. B. fehlende Elemente und fehlende semantische Interoperabilität. Um diese Probleme zu beheben, hat sich diese Arbeitsgruppe zum Ziel gesetzt, ein umfassendes Metadaten vokabular für Forschungssoftware bereitzustellen, das mit bestehenden Standards wie schema.org und CodeMeta kompatibel ist. Damit wird die Arbeitsgruppe alle NFDI-Konsortien bei der Anwendung und Etablierung von Metadaten schemata für Forschungssoftware unterstützen, z.B. bei Entwicklung von bereichsspezifischen Erweiterungen.

1. Motivation

“Research software is software employed in the scientific discovery process or a research object itself”¹. It is commonly used in various disciplines to conduct research, e.g., to solve equations, simulate complex systems, analyze or transform data, and statistically test models (e.g., climate or meteorological models). Provenance and rich descriptions of research software are important to improve scientific transparency, reproducibility, and reusability. To enhance the FAIRness of research software^{2,3}, software metadata play an important role.³

To improve its interoperability, these metadata should be based on a common metadata vocabulary. [CodeMeta](#) is a community-driven metadata standard for research software based on the classes [SoftwareApplication](#) and [SoftwareSourceCode](#) of schema.org. Various crosswalks to other metadata schemes already exist. CodeMeta contains multiple elements, some focusing on technical details like file size or supported operating systems, and others including administrative information like licenses. It supports the use of URIs for authors, contributors, and licenses. The content-specific metadata are limited to an application (sub-)category and keywords. CodeMeta is preparing a [new release 3.0](#) and is trying to incorporate its definitions into schema.org. The [Software Description Ontology](#) extends CodeMeta by SoftwareConfigurations defining the expected input and output and SoftwareImages to integrate Research Software in a Knowledge Graph⁴, but at the same time introduces complexity that complicates the description of Research Software.

While CodeMeta already fulfills a lot of different requirements for research software metadata from different use cases in the NFDI consortia, it also has some limitations. For example, CodeMeta merges two distinct types from schema.org (but does not define the merge as a new type) and self-defined elements of CodeMeta are so far not part of an ontology, making it difficult to integrate it into semantic-web-based applications. Also, CodeMeta does not recommend nor make certain elements mandatory (or recommended), which would be necessary to ensure a certain quality standard for software described with such metadata. The elements of CodeMeta are not detailed enough to describe the purpose of the software, its interface definition (input, output), its maturity (e.g., prototype), its type (command-line, library, script), its quality, its sustainability⁵, and its connection to research data. In addition, CodeMeta does not recommend value vocabularies for certain elements, e.g., for programming languages, operating systems, etc., which would improve the interoperability and semantic usability of the resulting metadata a lot.

¹ W. Hasselbring, L. Carr, S. Hettrick, H. Packer, and T. Tiropanis, “From FAIR research data toward FAIR and open research software,” *it - Information Technology*, vol. 62, no. 1, pp. 39–47, Feb. 2020, doi: [10.1515/itit-2019-0040](#).

² M. Barker *et al.*, “Introducing the FAIR Principles for research software,” *Sci Data*, vol. 9, no. 1, Art. no. 1, Oct. 2022, doi: [10.1038/s41597-022-01710-x](#).

³ A.-L. Lamprecht *et al.*, “Towards FAIR principles for research software,” *Data Science*, vol. 3, no. 1, pp. 37–59, Jan. 2020, doi: [10.3233/DS-190026](#).

⁴ D. Garijo, M. Osorio, D. Khider, V. Ratnakar, and Y. Gil, “OKG-Soft: An Open Knowledge Graph with Machine Readable Scientific Software Metadata,” in *2019 15th International Conference on eScience (eScience)*, Sep. 2019, pp. 349–358. doi: [10.1109/eScience.2019.00046](#).

⁵ D. Wilkinson, L. Oliveira, D. Mossé, and B. Childers, “Evaluating Interactive Archives,” 2017.

Accessed: Jul. 18, 2023. [Online]. Available:

<https://www.semanticscholar.org/paper/Evaluating-Interactive-Archives-Wilkinson-Oliveira/e9ad6276399e9e59ce4984193b6f0ae64413919a>

2. Objectives

The overall goal of the working group is to provide a common metadata vocabulary for research software that can be used to develop application profiles to describe research software in the different research domains of the NFDI consortia. An application profile is a metadata scheme that reuses existing terms, e.g. from existing schemes or metadata vocabularies. The solution should be compatible with common existing semantic frameworks including schema.org and metadata already in use in NFDI consortia, and take into account established recommendations such as CodeMeta. To reach this goal, the following sub-objectives should be met:

1. Collect different use cases of research software metadata from the different consortia
2. Review existing terminologies for describing (research) software with regard to their suitability for the collected use cases and their adoption by existing software tools and platforms.
3. Define a metadata vocabulary for describing research software reusing existing vocabularies where possible to ensure semantic interoperability and provide new metadata elements where needed.
4. Define best practices to use and adopt the vocabulary based on the needs of communities, including recommendations for value vocabularies and mandatory elements.
5. Coordinate adoption of research software metadata in research software directories/catalogs in cooperation with the working group [Research Software in section Infra](#), in particular, INFRA-RSE-T03 “Concept for a Software Marketplace”.

3. Work Plan: Projected timeframes and milestones

WP1: Collect use cases (Q3 2023 - Q4 2023)

- Collect different use cases for research software metadata as well as contact persons from the different NFDI consortia
 - Start with use cases from consortia presented in this working group
 - Continue with use cases from all consortia
- Cluster the different use cases
- Involved: Dorothea Iglezakis, Leyla Jael Castro, Stephan Ferez, Christian Schmidt-Sonntag, Olaf Kaczmarek

WP2: Review existing terminologies (Q3 2023 - Q4 2023)

- Review literature on different metadata schemes, ontologies, and terminologies to describe research software metadata
- Connect with existing initiatives and experts on research software metadata
- Involved: Luca Ghiringhelli, Stephan Ferez, Dorothea Iglezakis, Leyla Jael Castro, Alexander Struck

WP3: Define metadata vocabulary (Q1 2024 - Q2 2024)

- Identify relevant terms for research software metadata from the existing terminologies based on the use cases
- Define additional terms not included in any existing terminology based on the identified use cases
- Involved: Luca Ghiringhelli, Stephan Ferenz, Leyla Jael Castro, Olaf Kaczmarek, Jan Göpfert, Oliver Karras, Andreas Czerniak

WP4: Define best practices (Q1 2024 - Q2 2024)

- Provide guidelines on how to adopt the defined metadata vocabulary
- Gather best practices on research software (e.g., wrt the use of the metadata)
- Involved: Leyla Jael Castro, Andreas Czerniak, Stephan Ferenz

WP5: Coordinate adaptation (Q3 2024 - Q4 2024)

- Get feedback from the consortia on the defined metadata vocabulary
- Get feedback from international experts on the defined metadata vocabulary
- Assist the consortia in applying the metadata vocabulary
- Support the consortia to develop domain-specific extensions
- Involved: Leyla Jael Castro, Stephan Ferenz

4. Initial Membership List

(At least 6 members from different institutions and at least 6 consortia)

	Name	Institution	NFDI consortia
1	Stephan Ferenz (co-spoke)	Universität Oldenburg	NFDI4Energy
2	Oliver Karras	TIB	NFDI4Ing
3	Dorothea Iglezakis	Universität Stuttgart	NFDI4Ing, MaRDI
4	Marc Fuhrmans	TU Darmstadt	NFDI4Ing
5	Jan Göpfert	FZ Jülich	NFDI4Ing
6	Leyla Jael Castro (co-spoke)	ZB Med	NFDI4DS
7	Christian Schmidt-Sonntag	Universität Bielefeld	PUNCH4NFDI
8	Olaf Kaczmarek	Universität Bielefeld	PUNCH4NFDI
9	Matthias Löbe	Universität Leipzig	NFDI4Health
10	Lars Kastner	TU Berlin	MaRDI
11	Luca Ghiringhelli	HU Berlin	FAIRmat
12	Andreas Czerniak	Universität Bielefeld	
13	Alexander Struck	HU Berlin	ExC MoA
14	Patrick Kuckertz	FZ Jülich	NFDI4Ing

5. Adoption Plan

Metadata for research software is needed by all consortia that deal with research software, either as a research outcome or as a means of research. Furthermore, it is also needed to develop and consume registries or repositories for research software. With our WP1, we will directly collect requirements from the consortia. Later on, with WP5, we will assist all consortia in applying our metadata vocabulary.

Represented Consortia

NFDI4Ing
NFDI4Energy
MaRDI
NFDI4DS
NFDI4Health
PUNCH4NFDI
FAIRmat

Connected to other NFDI working groups

- WG RSE in Common Infrastructure

Related projects and approaches within the NFDI

- Metadata schema for Software Management Plans (ZB MED, NFDI4DS, ELIXIR Software Best Practices Group)
- DataDesc (NFDI4Ing)

Related projects and approaches outside NFDI

- FAIRCORE4EOSC (also responsible for CodeMeta Governance)
 - Work Package 6 Research Software
- FAIR-IMPACT:
 - D4.4 - Guidelines for recommended metadata standard for research software within EOSC
 - D5.2 - Metrics for automated FAIR software assessment in a disciplinary context
- Software Citation Implementation Working Group
- FAIR for Research Software (FAIR4RS) WG of RDA
- SciCodes: <https://scicodes.net/>
- SoMeSci