Towards Common Metadata For Research Software in the NFDI

Stephan Ferenz[1, 2] https://orcid.org/0000-0001-9523-7227 , Leyla Jael Castro[3] https://orcid.org/0000-0003-3986-0510 ,
Hamideh Hajiabadi[4] https://orcid.org/0000-0002-5793-4563 , Matthias Löbe[5] hhttps://orcid.org/0000-0002-2344-0426 ,
Ulrik Stervbo[6] https://orcid.org/0000-0002-2831-8868 , Lu Gan[7,8] https://orcid.org/0000-0001-5844-3021
and Florian Thiery[9,10] https://orcid.org/0000-0002-3246-3531

[1] OFFIS – Institute for Information Technology, Oldenburg, Germany
[2] Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
[3] ZB MED Information Centre for Life Science, Cologne, Germany
[4] Karlsruhe Institute of Technology (KIT), Germany
[5] IMISE Universität Leipzig, Germany
[6] Ruhr Universität Bochum, Germany
[7] Heinrich Heine University Düsseldorf, Germany
[8] GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany
[9] Leibniz-Zentrum für Archäologie (LEIZA), Mainz, Germany
[10] Research Squirrel Engineers Network & CAA-DE e.V., Germany

*Correspondence: Stephan Ferenz, stephan.ferenz@uol.de

Abstract

Most scientific disciplines use research software to address relevant research questions, and in many cases this research software is tailored to the specific research questions. Thus, research software is defined as software created during the research process or for a research purpose [1]. Within the German National Research Infrastructure (NFDI), research software plays two important roles. First, the services and tools developed by the consortia can be considered research software. These are mostly related to infrastructure software as defined by Hasselbring et al., but not exclusively [2]. Second, research software that supports a research project is considered an important research artifact itself. Several consortia of the NFDI are consequently working to provide infrastructure for these artifacts. Their significance is evident in their crucial function of research software, for example, in natural sciences for climate and meteorological modeling, in social sciences for network analysis and sentiment analysis, in digital humanities for text mining and computational linguistics, and in engineering sciences for simulating complex systems. Whether being a service/tool or a research artifact, provenance and rich descriptions of research software are essential to improve scientific transparency, findability, citability, reproducibility, software security, and reusability. Therefore, to improve the FAIRness of research software [3], software metadata plays an important role in the implementation of the FAIR4RS principles [4], [5].

To enable good software metadata, common metadata standards play an important role. Consequently, the working group "Research Software Metadata" of the Section Metadata, Terminology, and Provenance of the NFDI focuses on metadata for research software. The overall goal of this working group is to provide a common core metadata vocabulary for research software that can be used to describe research software created in the various

research domains of the NFDI consortia. The work will create a unified framework that can be used across the different NFDI consortia generating research software artifacts. This effort also addresses key interoperability challenges, such as aligning different metadata schemas, ensuring consistency across disciplines, and integrating with existing standards such as CodeMeta [6], Citation File Format (CFF) [7] and schema.org.

To gain an overview of research software metadata schemas already in use in the different consortia, we have conducted a survey on the implementation of research software metadata in the NFDI consortia [8]. We found that only a few domain-specific metadata schemas for research software were in use by different consortia. In contrast, generic metadata schemas such as CodeMeta, Bioschemas, schema.org, and the CFF were most frequently used to describe research software. Relevant metadata elements for research software include algorithms, methods, software/hardware requirements, and tasks performed by the software. We subsequently performed an in depth analysis of the alignment of the generic metadata schemas by creating crosswalks, i.e. mapping the schemas onto each other. This crosswalk analysis provides the basis for further discussions with the wider RDM community regarding strategies for provenance and rich descriptions of research software.

Keywords: Research Software, Software Metadata

Author contributions

All authors contributed in the conceptualization, writing - original draft and writing - review & editing.

Funding

References

[1] M. Gruenpeter *et al.*, "Defining Research Software: a controversial discussion," Zenodo, Sep. 2021. doi: 10.5281/zenodo.5504016.

[2] W. Hasselbring *et al.*, "Multi-Dimensional Research Software Categorization," *Comput. Sci. Eng.*, pp. 1–10, 2025, doi: 10.1109/MCSE.2025.3555023.

[3] M. Barker *et al.*, "The FAIR for Research Software Principles after two years: An adoption update," Zenodo, Mar. 2024. doi: 10.5281/zenodo.10816032.

[4] N. P. Chue Hong *et al.*, "FAIR Principles for Research Software (FAIR4RS Principles)," FAIR4RS WG, May 2022. doi: 10.15497/RDA00068.

[5] L. J. Castro *et al.*, "'Research Software Metadata' - Working Group Charter (NFDI section-metadata)," Zenodo, Dec. 2023. doi: 10.5281/zenodo.10246218.

[6] M. B. Jones *et al.*, *CodeMeta: an exchange schema for software metadata. Version 3.0.*, 2023. Accessed: Jun. 22, 2021. [Online]. Available: https://codemeta.github.io/terms/

[7] S. Druskat *et al.*, *Citation File Format*, Aug. 09, 2021. doi: 10.5281/zenodo.5171937.

[8] L. J. Castro *et al.*, "Results on a Survey on Research Software Metadata in the NFDI Consortia," Zenodo, Aug. 2024. doi: 10.5281/zenodo.12704414.