Final Report
FAIR Workflows: Connecting Engineering Research Outputs using PID Infrastructure

1. General Information

Applicant: Matthew Buys, Executive Director

Organisation: DataCite – International Data Citation Initiative e.V.

Title of the project:
FAIR Workflows: Connecting Engineering Research Outputs using PID Infrastructure

Total funding: 49.725,00 Euro

List of published work results for this project:

- Garza, K., & Laing, B. (2023). NFDI - Instrument Page (Version 1.0.0) [Computer software]. DataCite. https://doi.org/10.14454/FSN5-EC40
- Stathis, Kelly, Ross, Cody, Dreyer, Britta, & Vierkant, Paul. (2022). DataCite Metadata Schema 4.4 to Schema.org Mapping (1.0). Zenodo. https://doi.org/10.5281/zenodo.7661399
- Vierkant, Paul, Stathis, Kelly, & Dreyer, Britta. (2022). Best Practice Guidelines for the DOI Registration of Research Outputs (1.0). Zenodo. https://doi.org/10.5281/zenodo.7661294

Acknowledgement

The authors: Matthew Buys, Kristian Garza, Paul Vierkant, Britta Dreyer, Cody Ross, Kelly Stathis, and Bryceson Laing would like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the framework of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) - project number 442146713.

2. Executive Summary

This report describes the results and outcomes of the project focused on enabling and supporting the use of Persistent Identifiers (PIDs) and metadata within NFDI4Ing to implement the FAIR principles for research data, and building interactive dashboards to visualize the consortium output, links, and data reuse. The project was successful in achieving its objectives, as it provided a comprehensive solution for standardizing the deposit, description, and exchange of research outputs and resources related to instruments. The project delivered a set of guidelines for integrating discipline-specific metadata in a generic schema, supporting interoperability between systems and disciplines. The project also contributed to addressing the problem of limited standardization of PID metadata by providing PID services tailored to the requirements of the engineering community, in collaboration with DataCite and the TIB DOI Konsortium.

3. Work and result report

Preamble

The main objective of the project was to enable and support the use of PIDs (persistent identifiers) and metadata within NFDI4Ing to implement the FAIR principles for research outputs and resources. The project aimed to provide a comprehensive solution for standardizing the deposit, description, and exchange of metadata and contribute to addressing the lack of standardization of research outputs and resource metadata by providing PID services tailored to the requirements of the engineering community. In this report, we present the project results achieved, discuss their relevance to the state of research, possible application perspectives, and conceivable follow-up efforts.

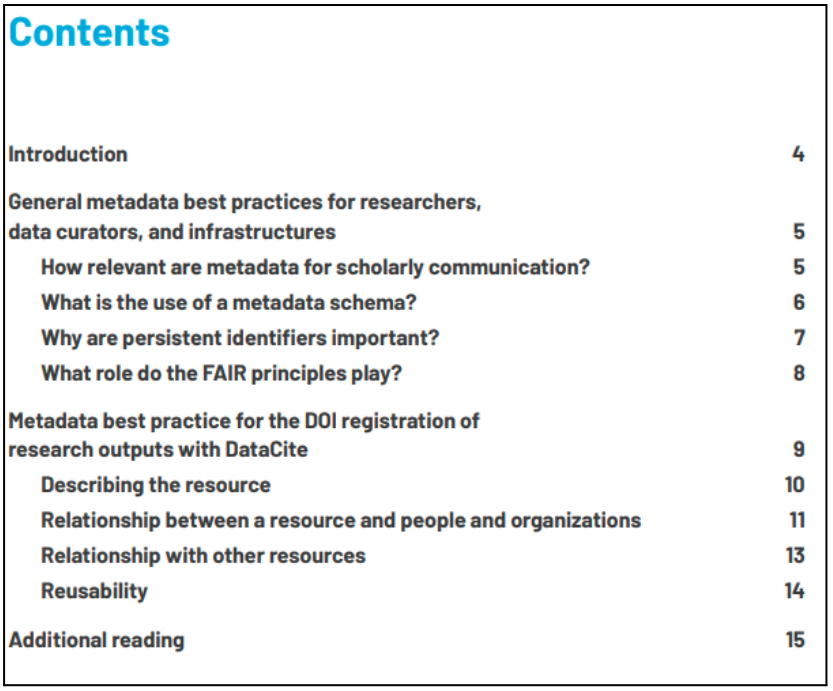
Project Results

The project successfully achieved its objectives by providing a comprehensive solution for standardizing the deposit, description, and exchange of research outputs and resources. A set of guidelines for integrating discipline-specific metadata in a generic schema was developed, supporting interoperability between systems and disciplines. The project contributed to addressing the problem of limited standardization of research outputs and resources by providing PID services tailored to the requirements of the engineering community, in collaboration with DataCite and the TIB DOI Konsortium. The project built interactive dashboards to visualize the consortium outputs and resources, links, and reuse, monitoring relevant KPIs for the FAIRness of engineering data management.

Below is an overview of the three key work results from this project:

I. Best Practice Guidelines[1]

The project team created the Best Practice Guidelines for the DOI Registration of Research Outputs to provide metadata best practices for researchers, data curators, and research infrastructures. These guidelines use the DataCite Metadata Schema to support the implementation of the FAIR principles for research outputs and resources. The document outlines the challenges of standardization in research and metadata management and the importance of metadata in making research outputs findable, accessible, interoperable, and reusable. The document aims to complement the DataCite Metadata Schema by giving concise hands-on best practice guidance on specific metadata properties.



Contents

Fig 1. Best Practice Guideline document contents

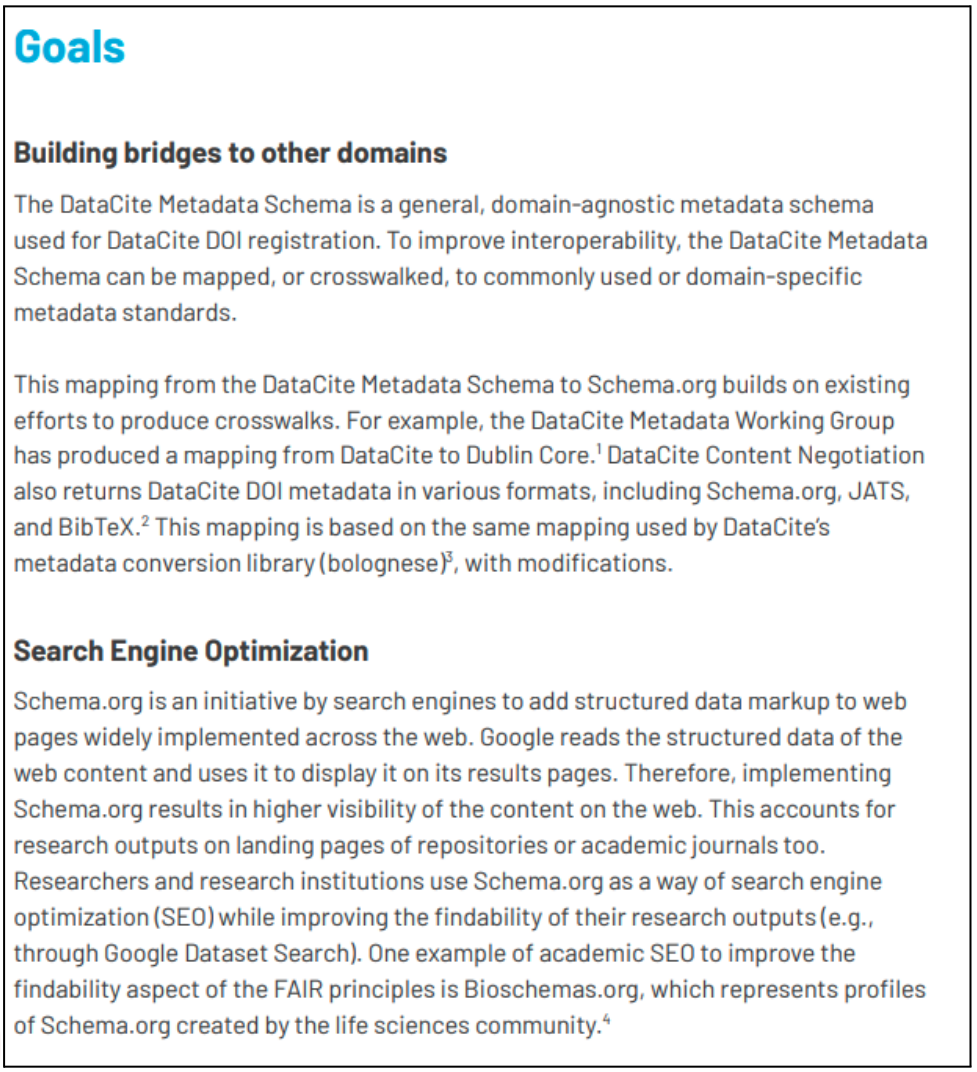
II. Schema.org Mapping[2]

One of the goals of the project was to map the DataCite Metadata Schema to Schema.org, which is widely implemented on the web to improve search engine optimization (SEO) and the discoverability of research outputs and resources. The mapping allows for crosswalks between the domain-agnostic DataCite Metadata Schema and specialized community standards while supporting the use of persistent identifiers

[1] Vierkant, Paul, Stathis, Kelly, & Dreyer, Britta. (2022). Best Practice Guidelines for the DOI Registration of Research Outputs (1.0). Zenodo. https://doi.org/10.5281/zenodo.7661294

[2] Stathis, Kelly, Ross, Cody, Dreyer, Britta, & Vierkant, Paul. (2022). DataCite Metadata Schema 4.4 to Schema.org Mapping (1.0). Zenodo. https://doi.org/10.5281/zenodo.7661399

(PIDs) and metadata within NFDI4Ing. The mapping was challenging due to the differences in the type hierarchy and the lack of equivalent properties between the two schemas. This being said, the mapping provided may serve as an initial starting point for implementation, but repositories should review the Schema.org properties in detail and customize their mappings as needed.



Goals

Building bridges to other domains

The DataCite Metadata Schema is a general, domain-agnostic metadata schema used for DataCite DOI registration. To improve interoperability, the DataCite Metadata Schema can be mapped, or crosswalked, to commonly used or domain-specific metadata standards.

This mapping from the DataCite Metadata Schema to Schema.org builds on existing efforts to produce crosswalks. For example, the DataCite Metadata Working Group has produced a mapping from DataCite to Dublin Core.[1] DataCite Content Negotiation also returns DataCite DOI metadata in various formats, including Schema.org, JATS, and BibTeX.[2] This mapping is based on the same mapping used by DataCite's metadata conversion library (bolognese)[3], with modifications.

Search Engine Optimization

Schema.org is an initiative by search engines to add structured data markup to web pages widely implemented across the web. Google reads the structured data of the web content and uses it to display it on its results pages. Therefore, implementing Schema.org results in higher visibility of the content on the web. This accounts for research outputs on landing pages of repositories or academic journals too. Researchers and research institutions use Schema.org as a way of search engine optimization (SEO) while improving the findability of their research outputs (e.g., through Google Dataset Search). One example of academic SEO to improve the findability aspect of the FAIR principles is Bioschemas.org, which represents profiles of Schema.org created by the life sciences community.[4]

Fig 2. Goals of the schema.org mapping

III. Interactive Dashboard[3]

The project developed an interactive dashboard, the NFDI – Instrument Page, to create and visualize connections in the PID graph representing the relationships between instruments and their associated resources. This tool provides an overview of connections between publications, datasets, and instruments and helps improve the findability, accessibility,

[3] Garza, K., & Laing, B. (2023). NFDI – Instrument Page (Version 1.0.0) [Computer software]. DataCite. https://doi.org/10.14454/FSN5-EC40

interoperability, and reusability (FAIR) of research data in the engineering sciences by implementing PIDs and metadata as essential components.

NFDI User Story Instrument	As an instrument manager, I want to be able to see the citations of my instrument, so that I see a complete picture of reuse.

Goal: By using this notebook, for a given instrument id you should be able to display:

- Counts of citations
- List of all the datasets created using the instrument
- List of all the publications that reference the instrument
- List of all the authors that have been involved with the use of the instrument
- A list and a bar chart showing how the related works counts have been linked to the instrument

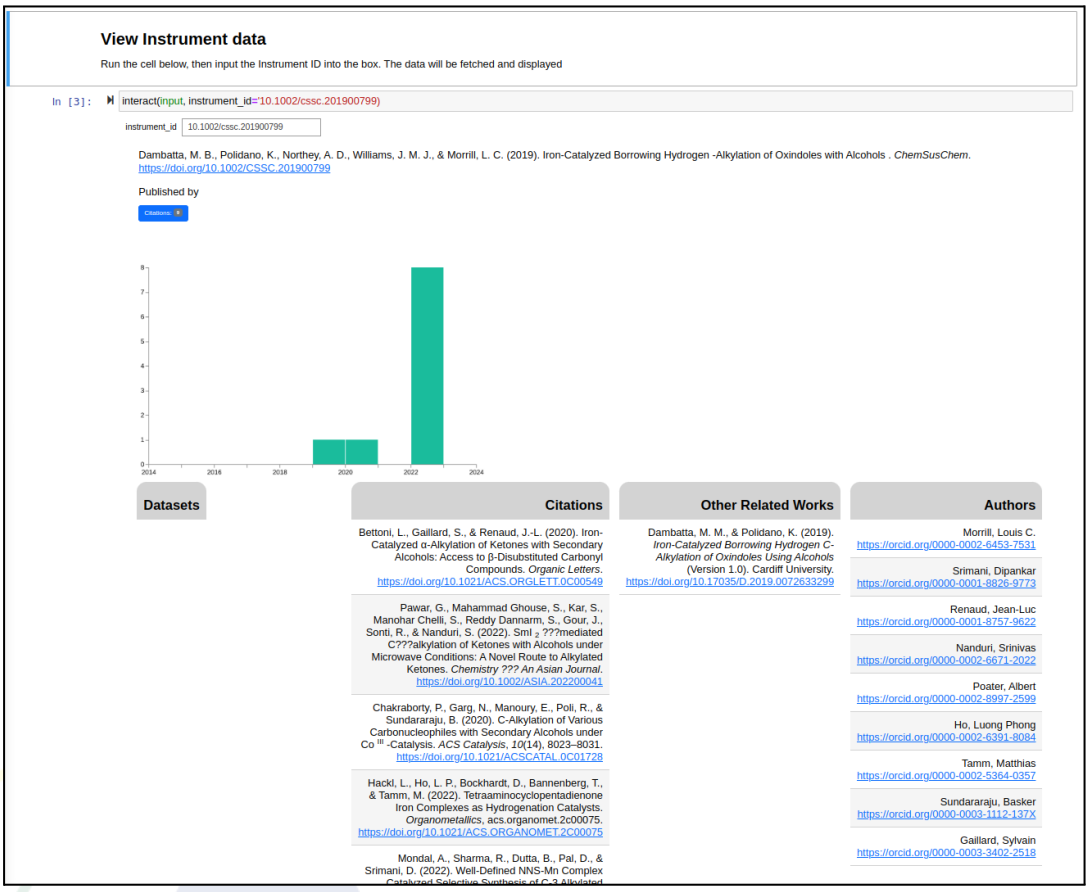


Fig 3. Example screenshot from the Jupyter notebook

Potential Applications and Future Developments

The project results have several potential applications in research data management as they provide a standardized and comprehensive solution for storing, describing, and exchanging research output and resource metadata. The project can be extended to cover other domains and disciplines, providing a generic schema for subject-specific metadata, enabling easier interoperability between systems and disciplines. The project can be further developed to integrate with the larger context of the National Research Data Infrastructure (NFDI) in general, and the NFDI4Ing consortium in particular. The project team has made the code and outputs openly available to facilitate PID cross-linking for overarching value-added services, such as knowledge graphs, as part of the larger ecosystem of research data management and infrastructure.

The results of this project have the potential to benefit the broader scientific community by providing a standardized and comprehensive solution for storing, describing, and exchanging research output and resource metadata. Additionally, the project's interactive dashboards and guidelines for integrating discipline-specific metadata in a generic schema can be used to improve the findability, accessibility, interoperability, and reusability (FAIR) of research data in other domains and disciplines beyond engineering.

The project team plans to maintain the interactive dashboards and guidelines to keep them up-to-date and relevant for the engineering community. They also plan to explore opportunities to collaborate with other National Research Data Infrastructure (NFDI) consortia and initiatives to improve research data management and FAIRness across different domains.

In summary, this project successfully achieved its objectives by providing a comprehensive solution for standardizing the deposit, description, and exchange of research outputs and resources. The project's interactive dashboards and guidelines for integrating discipline-specific metadata in a generic schema are valuable resources that can be used to improve the FAIRness of research data in the engineering sciences and potentially other domains and disciplines. The results of this project demonstrate the potential for collaboration between research communities and infrastructure providers to improve research data management and FAIRness.

Cooperation with the NFDI4Ing consortium

The project was developed in close cooperation with the NFDI4Ing consortium, providing tailored PID services to meet the specific requirements of the engineering community. The project results can be integrated into the larger context of the National Research Data Infrastructure (NFDI) in general, providing a standardized and comprehensive solution for storing, describing, and exchanging research data and its metadata.

Third-party contributions

While the core work of the project was carried out by our internal team at DataCite, we did receive contributions from a number of external parties. These included:

- TIB Konsortium as the DataCite DOI provider for the engineering sciences in Germany has contributed its metadata expertise and understanding of the challenges of implementing international standards and best practices in the DOI registration process. Furthermore, the TIB Konsortium identified use cases in the engineering sciences needs and the requirements that were central to the deliverables and success of this project.

We are grateful for the contributions of these third parties, which helped to enhance the scope and impact of our work.

Estimation of the range or degree of use of the project results

While it is still too early to fully assess the range or degree of use of the project results, we are optimistic that our work will have a significant impact on the engineering community and the broader NFDI ecosystem. We have already received positive feedback from several stakeholders, and we believe that our findings will be of interest to a wide range of researchers, practitioners, and policymakers.

As part of our dissemination efforts, we plan to make our results widely available through a variety of channels. By doing so, we hope to ensure that our work has a lasting impact on the field of engineering and beyond.