

## NFDI4Cat Research Data Management System: A Proposal for Enhanced Data Sharing in Catalysis Research

### Authors:

Nadiia Huskova<sup>1\*</sup>, Taras Petrenko<sup>1</sup>, Yuliia Dikova<sup>1</sup>, Volodymyr Kushnarenko<sup>1</sup>, Preston Rodrigues<sup>1</sup>, and Thomas Bönisch<sup>1</sup>

\*nadiia.huskova@hirs.de

<sup>1</sup>High Performance Computing Center Stuttgart (HLRS), Stuttgart, Germany

### Abstract:

Efficient collaboration and robust data sharing are important in every research field, particularly in catalysis studies where interdisciplinary collaboration is essential. The National Research Data Infrastructure for Catalysis (NFDI4Cat) addresses these imperatives by developing a centralized research data management (RDM) system. Led by the Fraunhofer Institute For Open Communication Systems in Berlin (FOKUS) and High-Performance Computing Center in Stuttgart (HLRS), one of the NFDI4Cat tasks areas has made a significant effort in developing RDM system for the catalysis-related sciences. However, to fully realize the potential of data sharing between research groups, the central repository must be seamlessly connected with other components.

The central repository, maintained by HLRS, acts as a central hub for storing and organizing a wide range of research data from various domains of catalysis, including biocatalysis, homogeneous catalysis, and heterogeneous catalysis. Beside the data sets themselves, important details about the tools and equipment are also kept. This ensures that the data are well-contextualized which makes it easier for others to reuse and reproduce the results of previous studies.

To increase the findability and understandability of data, NFDI4Cat aims to implement an approach where the storing of data sets in the repository requires only essential metadata, while more comprehensive metadata critical for understanding are prepared within a semantically rich RDF framework. Special attention is paid to the generation and collection of RDF data. In various domains, this task is typically performed manually, which constitutes a significant obstacle for wide adoption of the methodology in the research community. For this purpose, RDF data are produced interactively via an appropriately designed web application. The core of this application is the so-called TRIQ program developed at HLRS, which performs a questionnaire based on specific logical patterns within a reference ontology. The RDF data are stored in a appropriate triple store, and can be explored via a SPARQL endpoint. In addition, the usability of RDF data can be greatly enhanced by converting RDF data sets into HTML pages that can be navigated through a web browser.

This methodology enables high interoperability across diverse data sets and fosters deep insights into the data. In particular, the developed RDM system helps to acquire more detailed information beyond what is usually given by a researcher, as well as describe complex research processes and capture specific knowledge in the field. The software is currently under development and supposed to be a user-friendly solution. Furthermore, this approach integrates data, vocabularies, and ontologies into a unified, interconnected framework.

In conclusion, the infrastructure proposed, not only enhances data management and sharing within the catalysis research community but also holds promise for a broader adoption across various domains. The scalability and interoperability of the system offers a potential approach for improving research data management practices.