

# Using NFDI4Health Local Data Hub and Metadata-Schema for clinical research information management

Frank Meineke<sup>1</sup>, Matthias Löbe<sup>1</sup>, René Hänsel<sup>1</sup>, Masoud Abedi<sup>1</sup>, Xiaoming Hu<sup>2</sup> and Martin Golebiewski<sup>2</sup>

<sup>1</sup>Institute for Medical Informatics, Statistics, and Epidemiology (IMISE), Leipzig University, Leipzig, Germany

<sup>2</sup>Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany

## Abstract

The German National Research Data Infrastructure for Personal Health Data (NFDI4Health) aims at making structured health data internationally searchable and accessible according to the FAIR guiding principles. We have developed an overarching metadata schema (MDS) for clinical trials and epidemiological studies and have implemented it in the Local Data Hub (LDH) software. LDHs installations are provided as data management platforms for data-holding organizations to organize comprehensive local research information on clinical and epidemiological research projects. Metadata data can be entered into the LDHs and via direct transmissions published to the central German Health Study Hub of NFDI4Health. LDHs are locally based low-level entry points into research data and information sharing. We will introduce the NFDI4Health metadata schema and demonstrate the LDH system as well as report on initial experiences from the first roll-outs.

## Keywords

Data collection, clinical trial, cohort studies, research design, research data management, representation of metadata

## 1. Introduction

The National Research Data Infrastructure for Personal Health Data (NFDI4Health) is a long-term initiative of the German National Research Infrastructures network (NFDI), funded jointly by federal and state governments via the German Research Foundation (DFG). To make clinical, epidemiological, and public health research data FAIR, NFDI4Health developed three key components: i) a common metadata schema (MDS) for (meta-)data publication from health studies, ii) the German Central Health Study Hub (GCHSH), and iii) the Local Data Hub (LDH), a web-based platform for local research data management (RDM).

The MDS [1] unifies metadata requirements across epidemiological and clinical domains, addressing the absence of a consensus schema for epidemiology. It is modularly designed based on user needs, supporting metadata publication for studies and their associated resources (e.g., protocols, instruments, and documents). Many of the more than 200 data items were primarily adapted from established metadata standards and schemes, including the ones from DataCite, ClinicalTrials.gov, DRKS, or Maelstrom. The GCHSH [2] now contains over 27,000 publicly accessible entries, predominantly interventional studies, using this MDS. Coordination centres for clinical trials register metadata in national and international registries and typically don't have tools for consistent research data management (RDM tools) in use. Public Health and Epidemiologic Centres typically operate platforms with application procedures tailored to their specific need of data sharing. However, the solution for a comprehensive presentation of research activities (including project structures, participating institutions, researchers, funding programs, publications, standard operating procedures, study metadata, and data) into one unified system with interface to GCHSH remained unaddressed.

---

SWAT4HCLS 2025: 16th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences 2025

ORCID: 0000-0002-9256-7543 (F. Meineke); 0000-0002-2344-0426 (M. Löbe); 0000-0001-8344-0658 (R. Hänsel); 0000-0002-8318-3222 (X. Hu); 0000-0002-8683-7084 (M. Golebiewski)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 2. The Local Data Hub

LDH [3] is a web-based cataloguing platform that provides a comprehensive view of the research activities of a Data Holding Organisation (DHO). It includes metadata about studies, as well as information about people, institutions, research programmes and supporting documents such as study protocols and statistical analysis plans or publications references.

The NFDI4Health LDH software is based on FAIRDOM-SEEK [4], which has proven to be a robust RDM tool in the biomedical and life science domains for many years. We have adapted and extended it across domains for clinical and epidemiological projects.

In particular, we have extended SEEK to flexibly handle complex, hierarchical and tailored metadata. This extension of SEEK we have applied to fully integrate the NFDI4Health MDS into the SEEK-based LDH software. The metadata structured in this way in an installation of the LDH software then can be published directly to the GCHSH via an API interface. Unlike GCHSH, access to resources in LDH can be either public or fine-grained, restricted to users or groups.

## 3. Results and Discussion

For the first roll-out phase, we launched a small tender offering DHOs the opportunity to use the system temporarily as a software-as-a-service. Applicants came not so much from large epidemiological studies and cohorts, but rather from clinical study centres and the Medical Data Integration Centres (MedDIC). 8 of 9 sites successfully published each at least 6 projects. We got feedback to improve real-time checks to prevent validity check errors in GCHSH publishing phase.

The LDH proved to be a visible commitment of the DHOs to research transparency and FAIRness. The use of RDM solutions such as the Local Data Hub is an intervention in the routine processes and publication culture of an institution and requires institutional commitment.

## Acknowledgments

Software used and developed (SEEK base system, LDH extensions and ready to start software container-based deployment) is published as open source.<sup>1</sup>

*This work was done as part of the NFDI4Health Consortium ([www.nfdi4health.de](http://www.nfdi4health.de)). We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 442326535 (NFDI4Health), as well as from the Klaus Tschira Foundation (KTS).*

## Declaration on Generative AI

The authors have not employed any Generative AI tools.

## References

- [1] Metadata schema of the NFDI4Health and the NFDI4Health task force COVID-19 (v3\_2), <https://doi.org/10.4126/FRL01-006453422>, 2023. Accessed: 2026-2-27.
- [2] J. Fluck, M. Golebiewski, J. Darms, Data publication for personalised health data, Proc Conf Res Data Infrastr 1 (2023).
- [3] X. Hu, H. Abaza, R. Hänsel, M. Abedi, M. Golebiewski, W. Müller, F. Meineke, NFDI4Health local data hubs implementing a tailored metadata schema for health data, Stud. Health Technol. Inform. 317 (2024) 115–122.

---

<sup>1</sup><https://github.com/nfdi4health/ldh-deployment> and <https://github.com/nfdi4health/ldh/releases>

- [4] K. Wolstencroft, S. Owen, O. Krebs, Q. Nguyen, N. J. Stanford, M. Golebiewski, A. Weidemann, M. Bittkowski, L. An, D. Shockley, J. L. Snoep, W. Mueller, C. Goble, SEEK: a systems biology data and model management platform, *BMC Syst. Biol.* 9 (2015) 33.