








Flexible Metadata Structuring for Research Data Management Through the FAIRDOM-SEEK Platform

Implementing Tailored and Complex Metadata Schemes in SEEK

Xiaoming Hu¹  <https://orcid.org/0000-0002-8318-3222>, Stuart Owen²  <https://orcid.org/0000-0003-2130-0865>, Frank Meineke³  <https://orcid.org/0000-0002-9256-7543>, Finn Bacall²  <https://orcid.org/0000-0002-0048-3300>, Carole Goble²  <https://orcid.org/0000-1111-2222-3333>, Wolfgang Müller¹  <https://orcid.org/0000-0002-4980-3512>, and Martin Golebiewski^{1*}  <https://orcid.org/0000-0002-8683-7084>

¹ Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany

² University of Manchester, UK

³ University Leipzig, Germany

*Correspondence: Martin Golebiewski, martin.golebiewski@h-its.org

Abstract

Modern research projects increasingly require hybrid metadata approaches that balance adherence to domain-overarching, as well as domain-specific community standards with flexibility for project- or resource-specific metadata. The FAIRDOM-SEEK platform [1] is a widely used research data management system designed to support diverse domains, from systems biology to health research data, by integrating standardized metadata models (e.g., the ISA framework [2]) with customizable extensions. To address this need, we introduce the Extended Metadata feature in SEEK, which allows researchers to extend core metadata schemas with user-defined fields, hierarchies, and semantic annotations while ensuring interoperability with domain-specific standards. We demonstrate this capability through two use cases:

1. NFDI4Health Local Data Hubs (LDH) [3],[4]:

In the context of the German National Research Data Infrastructure for Personal Health Data (NFDI4Health [5]), we have developed Local Data Hubs (LDH) based on the SEEK platform. These hubs support federated data structuring and sharing for sensitive health data from clinical trials, epidemiological studies, and public health research and allow to connect local platforms to the central metadata repository of NFDI4Health, the German Health Study Hub. Given the complexity of the NFDI4Health metadata schema (MDS) [6], the SEEK-based LDH software utilizes the Extended Metadata feature to fully represent the schema, allowing for flexible project-defined metadata extensions.

2. FAIR Data Station (FAIR-DS) [7]:

Based on the ISA-framework, with the addition of Observation units from MIAPPE [8], the FAIR-DS is a web application that enables users to create and manage metadata according to FAIR principles. Using packages and terms configured through the UI, it generates Excel spreadsheets which are then populated to gather the metadata. FAIR-DS is then used to validate the metadata and generates RDF datasets repre-

senting the content. SEEK has been updated to allow Extended Metadata and Sample Types to be configured automatically via these RDF datasets, and also the content can be imported, and updated, in a single action.

The Extended Metadata feature allows users to define additional metadata attributes to be tailored to specific data types, ensuring compliance with standards. When creating a resource, users can select an Extended Metadata type from a dropdown menu, dynamically triggering the rendering of associated metadata input forms within the web interface. This enables seamless integration of resource-specific metadata (e.g., clinical trial study metadata) alongside core descriptive fields.

Currently, only instance administrators can create, manage (enable/disable), and delete additional attributes for specific resource types (e.g., ISA items such as Investigation, Study, Assay, as well as Projects and Models) based on specific schemas (e.g., the NFDI4Health MDS). Attribute types range from simple (e.g., string, text, date, integer, Boolean) to complex (e.g., controlled vocabularies linked to ontologies, nested hierarchical structures), with validation rules for mandatory or optional fields. Regular expressions are introduced to ensure correct input formatting.

Metadata schemas can be created through backend seed files, JSON uploads, or FAIR-DS RDF imports. These schemas are programmatically accessible via the SEEK REST API, enabling automated metadata creation and retrieval. This ensures interoperability with external tools while adhering to FAIR data principles.

Keywords: FAIR data, metadata schema, research data management, FAIRDOM-SEEK

Resources

FAIRDOM SEEK Sources, <https://github.com/seek4science/seek>

Author contributions

XH developed and implemented the code, supported by SO, FB, FM and WM. FM coordinates the LDH development, supported by MG. MG and FM conceptualized the extended metadata feature in SEEK, supported by XH and WM. SO is responsible for the FAIR-DS part. XH wrote the initial abstract draft and all authors contributed to reviewing and editing.

Competing interests

The authors declare that they have no competing interests.

Funding

We greatly appreciate the funding from the Deutsche Forschungsgemeinschaft (DFG) through projects no. 442326535 (NFDI4health), 451265285 (NFDI4health TF COVID19), and 315072261 (NMDR2), as well as from the Klaus Tschira Foundation (KTS).

Acknowledgement

We thank the whole FAIRDOM community and the NFDI4Health consortium for their valuable input and feedback.

References

- [1] 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 Systems Biology, 9, 33, July 2015, doi: <https://doi.org/10.1186/s12918-015-0174-y>.
- [2] Sansone, S.-A., Rocca-Serra, P., Gonzalez-Beltran, A., Johnson, D., & ISA Community. ISA Model and Serialization Specifications 1.0. Zenodo. 2016. <https://doi.org/10.5281/zenodo.163640>.
- [3] Meineke, F., Golebiewski, M., Hu, X., Kirsten, T., Löbe, M., Klammt, S., ... Müller, W. (2023). NFDI4Health Local Data Hubs for Finding and Accessing Health Data: Making Distributed Data Accessible Through a SEEK-Based Platform. *Proceedings of the Conference on Research Data Infrastructure*, 1. <https://doi.org/10.52825/cordi.v1i.375>.
- [4] Hu X, Abaza H, Hänsel R, Abedi M, Golebiewski M, Müller W, Meineke F. NFDI4Health Local Data Hubs Implementing a Tailored Metadata Schema for Health Data. *Stud Health Technol Inform*. 2024 Aug 30;317:115-122. <https://doi.org/10.3233/SHTI240845>.
- [5] Fluck, J., Lindstädt, B., Ahrens, W., Beyan, O., Buchner, B., Darms, J., Depping, R., Dierkes, J., Neuhausen, H., Müller, W., Zeeb, H., Golebiewski, M., Löffler, M., Löbe, M., Meineke, F., Klammt, S., Fröhlich, H., Hahn, H., Schulze, M., Pischon, T., Nöthlings, U., Sax, U., Kusch, H., Grabenhenrich, L., Schmidt, C.O., Waltemath, D., Semler, S., Gehrke, J., Kirsten, T., Praßer, F., Thun, S., Wieler, L., Pigeot, I., "NFDI4Health – Nationale Forschungsdateninfrastruktur für personenbezogene Gesundheitsdaten", BFDm, Nr. 2, S. 72–85, Juli 2021, doi: <https://doi.org/10.17192/bfdm.2021.2.8331>.
- [6] Abaza H, Shutsko A, Klopfenstein SAI, Vorisek CN, Schmidt CO, Brünings-Kuppe C, Clemens V, Darms J, Hanß S, Intemann T, Jannasch F, Kasbohm E, Lindstädt B, Löbe M, Nimptsch K, Nöthlings U, Ocanto MG, Osei TB, Perrar I, Peters M, Pischon T, Sax U, Schulze MB, Schwarz F, Schwedhelm C, Thun S, Waltemath D, Wünsche H, Zeleke AA, Müller W, Golebiewski M. Towards a Domain-Overarching Metadata Schema for Making Health Research Studies FAIR: The NFDI4Health Metadata Schema. *JMIR Medical Informatics*. 2025 Jan 23;63906 (in press). <https://doi.org/10.2196/63906>.
- [7] Bart Nijse, Peter J Schaap, Jasper J Koehorst, FAIR data station for lightweight metadata management and validation of omics studies, *GigaScience*, Volume 12, 2023, giad014, <https://doi.org/10.1093/gigascience/giad014>.
- [8] Ówiek-Kupczyńska, H., Altmann, T., Arend, D. et al. Measures for interoperability of phenotypic data: minimum information requirements and formatting. *Plant Methods* 12, 44 (2016). <https://doi.org/10.1186/s13007-016-0144-4>.