Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240676

Using Health IT Standards to Facilitate Metadata Exchange Between Research Data Management Systems

Matthias LÖBE [a,1] Xiaoming HU [b] and Sophie A. I. KLOPFENSTEIN [c]
[a] *Institute for Medical Informatics (IMISE), University of Leipzig, Germany*
[b] *Heidelberg Institute for Theoretical Studies (HITS), Heidelberg, Germany*
[c] *Core Facility Digital Medicine and Interoperability, Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Germany*
ORCiD ID: M. Löbe https://orcid.org/0000-0002-2344-0426,
X. Hu https://orcid.org/0000-0002-8318-3222,
S. A. I. Klopfenstein https://orcid.org/0000-0002-8470-2258

Abstract. The National Research Data Infrastructure for Personal Health Data (NFDI4Health) uses Local Data Hubs (LDHs) to manage locally research studies, documents and sensitive personal data to support controlled data sharing. While research data management (RDM) systems facilitate the storage and preparation of data and metadata as well as organizational access, they often lack support for interoperability standards of the application domain. To support the exchange with external registries of research studies, we chose 17 attributes to characterize the most relevant aspects of clinical trials (in the following named "metadata profile"). We implemented the metadata profile in the RDM system FAIRDOM SEEK using core attributes and SEEK's extended metadata feature and created a mapping conforming to the Health Level 7 Fast Healthcare Interoperability Resources (FHIR) standard version R4. Finally, we implemented a prototype application interface for exports in FHIR-JSON format. We plan to extend the interface to serve central registries and support specific FHIR Implementation Guides from various use cases.

Keywords. RDM, SEEK, HL7 FHIR, Clinical research, FAIR, Metadata standards, Interoperability

1. Introduction

A research data management (RDM) system fulfills a variety of functionalities to support the entire life cycle of research data. The most important point is undoubtedly the persistent storage of data and metadata in a repository. The inclusion of new assets should be governed by consensus-based processes and conventions in order to capture all important information on the one hand and to minimize the effort for data producers on the other. However, further points are crucial for the broad reuse of research data repositories, including by third parties: In particular, the findability of datasets through indexing or suitable metadata, functions for access control and security to ensure that only authorized users can access certain data, support for international standards for data

[1] Corresponding Author: Matthias Löbe; E-mail: matthias.loebe@imise.uni-leipzig.de.

exchange and analysis and the citability of artefacts gained in importance in recent years as part of the FAIR (findability, accessibility, interoperability and reusability) principles [1].

The National Research Data Infrastructure for Personal Health Data (NFDI4Health) is one of 27 scientific consortia that aim to establish overarching infrastructures for scientists in Germany. In addition to central services and offers for privacy-preserving distributed computing [2], local data repositories–so called Local Data Hubs (LDHs)–manage sensitive data requiring special protection. Such sensitive data is generated in many areas of health research, and cannot be simply shared with regard to legal requirements and the consent of the subject/participants. The LDHs are enhancements of the FAIRDOM-SEEK software, which is widely used in the field of bioinformatics [3]. One large instance is the Leipzig Health Atlas [4]. SEEK is based on an adaptation of the Investigation-Study-Assay model (ISA). ISA is a framework for describing the experimental metadata of biological and biomedical studies, ensuring consistency and facilitating data sharing and integration. For research studies, a so-called archive information package according to ISO 14721, also known as the Open Archival Information System (OAIS) Reference Model, was developed, which defines important data packages and documents for the understanding of the experiment by third-party researchers [5]. An extension to include W3C DCAT (a standard for describing and sharing metadata about data catalogs and datasets) has already been implemented as a prototype [6]. Although SEEK has a JSON API, it only supports specific web or health information technology (IT) standards to a limited extent. SEEK's extended metadata feature enables to create and curate additional metadata, ensuring adaptability to project requirements and the integration of further standards. However, in recent years, Health Level 7 Fast Healthcare Interoperability Resources (FHIR®) has established itself as the most important standard in the field of health(care) IT, and presents ongoing developments in the area of (clinical) research [7].

As part of the use of LDHs as registers for local data use projects in data integration centres at university hospitals, there is a need to be able to reference research studies directly and process them automatically, for example when patients are recruited as study participants. Infrastructures in healthcare are currently converting their information systems to FHIR as the main data exchange standard, so there is a requirement for FHIR-based communication. This paper presents the development of a prototype application interface enabling FHIR exports from SEEK in FHIR-JSON format.

2. Methods

A minimal set of relevant study metadata (in the following named "metadata profile") was defined in a consensus-based iterative process based on expert requirements and FHIR knowledge. The metadata profile was implemented to SEEK using core attributes and SEEK's extended metadata feature via a seed file. We mapped the metadata profile to FHIR Release 4 (R4) and stored the map in SEEK. A prototype application interface was developed to enable users to enter study metadata through a user-friendly web-based interface. We integrated the FHIR map into the export function of SEEK to support the creation of FHIR-JSON files compliant with "ResearchStudy", "Practitioner", "Organization" and "Group" FHIR resources. We validated the resulting structures with prominent FHIR validators.

3. Results

The metadata profile contains 17 attributes to describe studies: We used seven core attributes and thirteen extended metadata attributes to represent it. Table 1 presents the mapping between FHIR R4 elements and the core SEEK attributes supplemented by attributes from the metadata profile.

Table 1. Elements and references used from FHIR ResearchStudy populated from SEEK core attributes or user-defined metadata profile

FHIR R4 element / reference	SEEK core attribute	Extended metadata profile attribute
ResearchStudy	Investigation and / or Study	
--identifier	Study.ID	study_identifier
--title	Study.Title	
--partOf	Study.Investigation	
--status		study_status
--category		study_type
--condition		study_condition
--contact		study_homepage
--keyword	Study.Tags	
--description	Study.Description	
--enrollment (Group)		study_sample_size
		study_start_date
--period		study_end_date
--sponsor (Organization)		study_sponsor
--principalInvestigator (Practitioner)		study_pi
(Extension)		study_acronym
(Extension)		study_sites_number
(Extension)		study_dmp

Currently, the prototype uses the ResearchStudy, Group, Organization, Practitioner FHIR resources. The code can be retrieved from the GitHub repository [8].

Figure 1 shows the overview page of a research study in an LDH including the metadata profile (without other artifacts such as data sets or study documents).

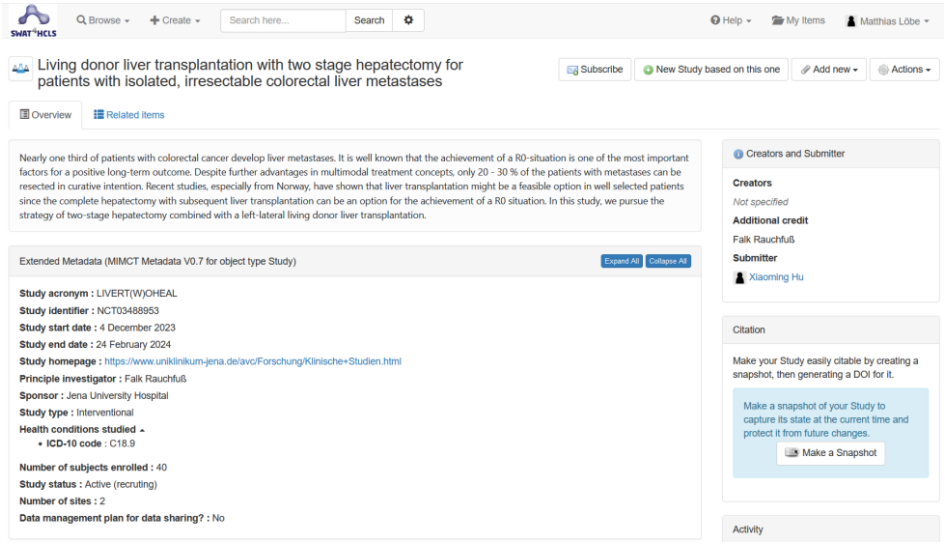

Figure 1. Clinical study in a SEEK Local Data Hub (cropped to emphasize the relevant content)

Figure 2 shows an excerpt from the generated FHIR structures that can be retrieved and consumed by external agents. While ResearchStudy is an independent resource, the others are exported as contained resources.

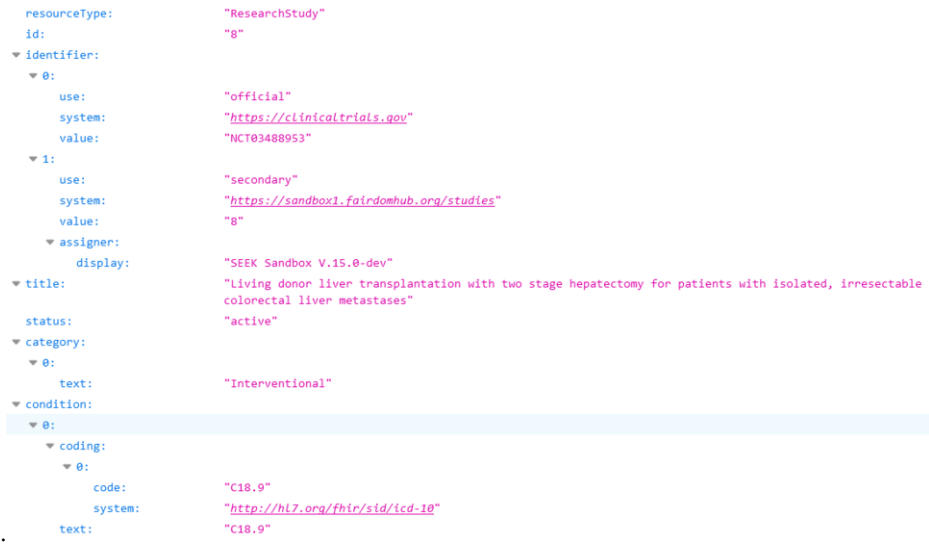

Figure 2. Representation of a clinical study in FHIR R4.0.1 (excerpt)

4. Discussion

RDM and FHIR are two separate but related concepts in the field of healthcare and medical research. RDM systems are more oriented towards the human user and tend to address the findability, access and secondary use of data. FHIR-based systems focus on smooth, permanent machine data exchange and interoperability. RDM platforms can use FHIR to enable seamless integration with other healthcare systems. The presented export functions as a READ API and serves as a FHIR-compliant data source that can be seamlessly integrated into other systems. It allows any LDH to expose their studies so that they can be captured by FHIR-based study registries.

While an ISA-to-FHIR mapping has already been developed in the past [9], it could not be used for this study, as SEEK deviates in many areas from the original ISA model. Our prototype currently covers only the basic attributes required for a minimal FHIR ResearchStudy as requested by clinical users. We plan to include further artifacts to support the metadata of e.g., study documents and datasets that are available for a study in the LDH. Furthermore, it would be useful to provide an export of a complete study as a JSON bundle, which can be uploaded directly to a target FHIR server and takes into account existing artifacts such as existing persons and organizations (conditional creates). Supporting profiles from Implementation Guides of popular use cases instead of manual mapping would be worth considering. However, detailed study metadata should be stored in a structured manner in special trial registries rather than in SEEK.

Interventional clinical studies are predominantly recorded in study registries, but the majority of academically driven research studies are not. Furthermore, the current comprehensive further development of FHIR in the area of (clinical) research is

challenging: FHIR R5, the current version, brings major improvements, but many projects are still dependent on R4 to remain compatible with national base profiles and other FHIR data. In addition, the new, richer constructs allow for multiple alternative modeling options, and best practices have yet to emerge.

5. Conclusions

We are currently in the process of coordinating with the German Medical Informatics Initiative's Medical Research Data Portal [10], which, as a national transparency register, aims to track all research projects with patient data from university hospitals. This aligns with RDM's growing focus on patients and subjects as active participants that want to know what research projects their data is being used for and being involved in decisions about its use.

Acknowledgments and Competing Interests

This work was done as part of the NFDI4Health project (www.nfdi4health.de). We gratefully acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project Number 442326535.

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, et. al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. doi: 10.1038/sdata.2016.18

[2] NFDI4Health. NFDI4Health National Research Data Infrastructure for Personal Health Data [Internet]. 2024. Available from: https://www.nfdi4health.de/en/service.html

[3] Wolstencroft K, Owen S, Krebs O, Nguyen Q, Stanford NJ, Golebiewski M, Weidemann A, Bittkowski M, An L, Shockley D, Snoep JL, Mueller W, Goble C (2015) SEEK: a systems biology data and model management platform. BMC Syst Biol 9, 33.doi: 10.1186/s12918-015-0174-y

[4] Kirsten T, Meineke F, Löffler-Wirth H, Uciteli A, Beger C, Stäubert S, Löbe M, Hänsel R, Rauscher FG, Schuster JC, Peschel T, Herre H, Wagner J, Zachariae S, Engel C, Scholz M, Rahm E, Binder H, Löffler M (2022) The Leipzig Health Atlas - An open platform to present, archive and share bio-medical data, analyses and models online. Methods Inf Med. doi: 10.1055/a-1914-1985

[5] Löbe M, Kuntz A, Henke C, Meinke F, Sax U, Winter A (2023) Concept for a Basic ISO 14721 Archive Information Package for Clinical Studies. Studies in health technology and informatics 302, 721–725. doi: 10.3233/SHTI230247

[6] Löbe M, Ulrich H, Beger C, Bender T, Bauer C, Sax U, Ingenerf J, Winter A (2022) Improving Findability of Digital Assets in Research Data Repositories Using the W3C DCAT Vocabulary. Studies in health technology and informatics 290, 61–65. doi: 10.3233/SHTI220032

[7] Duda SN, Kennedy N, Conway D, Cheng AC, Nguyen V, Zayas-Cabán T, Harris PA (2022) HL7 FHIR-based tools and initiatives to support clinical research: a scoping review. J Am Med Inform Assoc 29, 1642–1653. doi: 10.1093/jamia/ocac105

[8] https://github.com/nfdi4health/SEEKonFHIR

[9] Klopfenstein SAI, Sass J, Vorisek CN, Jorczik F, Schmidt CO, Löbe M, Golebiewski M, Abaza H, Thun S. Bringing Communities Together: Mapping the Investigation-Study-Assay-Model (ISA) to Fast Healthcare Interoperability Resources (FHIR). Stud Health Technol Inform. 2024 Jan 25;310:18-22. doi: 10.3233/SHTI230919. PMID: 38269757.

[10] Prokosch H-U, Gebhardt M, Gruendner J, Kleinert P, Buckow K, Rosenau L, Semler SC (2023) Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned. Studies in health technology and informatics 302, 307–311. doi: 10.3233/SHTI230124