

Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned

Hans-Ulrich Prokosch^{a,1}, Marie Gebhardt^b, Julian Gruendner^a, Philip Kleinert^b,
Karoline Buckow^b, Lorenz Rosenau^c and Sebastian C. Semler^b

^a*Chair of Medical Informatics, University of Erlangen-Nürnberg, Erlangen, Germany*

^b*TMF - Technology, Methods, and Infrastructure for Networked Medical Research, Berlin, Germany*

^c*IT Center for Clinical Research, University of Lübeck, Lübeck, Germany*

Abstract. Harmonizing medical data sharing frameworks is challenging. Data collection and formats follow local solutions in individual hospitals; thus, interoperability is not guaranteed. The German Medical Informatics Initiative (MII) aims to provide a Germany-wide, federated, large-scale data sharing network. In the last five years, numerous efforts have been successfully completed to implement the regulatory framework and software components for securely interacting with decentralized and centralized data sharing processes. 31 German university hospitals have today established local data integration centers that are connected to the central German Portal for Medical Research Data (FDPG). Here, we present milestones and associated major achievements of various MII working groups and subprojects which led to the current status. Further, we describe major obstacles and the lessons learned during its routine application in the last six months.

Keywords. Large scale data sharing, real world data, federated data network

1. Introduction

To leverage the power of real world data from all German university hospitals, the medical informatics initiative (MII) was initiated by the German Ministry of Research and Education [1]. Four consortia (DIFUTURE [2], HiGHmed [3], MIRACUM [4], and SMITH [5]) and a national coordination office [6] received funding from 2018 to 2022 to unlock the heterogeneous data silos across the German university hospital landscape and to enable large scale data sharing throughout Germany. Each consortium had independently defined its concept and practical approach towards data sharing within the partner sites of the consortium. Additionally, the MII national steering committee (NSC: comprising coordinators of each consortium, representatives from the national MII coordination office, and the Federal Ministry of Education and Research) was created as an overarching governance structure. Through these structures, a Germany-wide data sharing network across German university hospitals' data integration centers (DIC) was

¹ Corresponding Author: Hans-Ulrich Prokosch, Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Wetterkreuz 15, 91058 Erlangen, Germany; E-mail: Hans-Ulrich.Prokosch@fau.de.

established. The NSC initiated cross-consortia working groups (WGs) focused on data sharing concepts, patient consent, interoperability, and communication. The final aim of the MII was to support clinical and translational medical researchers by providing access to clinical patient data from all German university hospitals through one central entry portal. The portal enables researchers to:

- get a general overview of data and data types available across all German university hospitals within their local DIC,
- characterize their cohort of interest for planned research analysis,
- perform feasibility queries to retrieve the size of matching patient datasets,
- define data use proposals to request data from all integrated DIC and finally receive the proposed type of access for their research projects (e.g., central or federated analysis).

Such a portal (the German Portal for Medical Research Data = Deutsches Forschungsdatenportal für Gesundheit = FDPG) has been implemented and provided to the German university medicine research community as a beta-release in October 2022 and will be opened for general use in the first quarter of 2023.

The objective of this publication is to illustrate the general MII data sharing framework, the technical status of the FDPG, and the lessons learned from a projectathon in which the FDPG was used as data sharing framework.

2. Methods

Important milestones towards the FDPG have been achieved within the MII working groups “data sharing,” “consent,” and “interoperability,” which tackled the numerous harmonization/ standardization tasks required to align all German university hospitals in a joined, large scale data sharing network. Major results are published on the MII website [7]. The FDPG framework consists of three modules. First, a Germany-wide feasibility tool, which was developed and successfully deployed as a sub-project of (1) the CODEX project (COVID-19 data exchange platform: 2020/2021) [8] and (2) the MII- project ABIDE_MI (Aligning Biobanks and Data Integration Centers efficiently: 2021/2023) [9]. Second, The FDPG research proposal management module was developed by a commercial software development partner (Appsfactory GmbH). A third module, a transparency register was set up as a webpage designed to make research projects visible and understandable to patients. Parts of the FDPG middleware are based on consortia concepts and have been integrated into the FDPG ecosystem. The project partners have delivered central components for the portal and matching components for the decentral sites.

The decentral regulatory foundation and the local technical infrastructures were iteratively established at 31 German university hospitals within their DIC. This includes the establishment of associated data (and biosample) use and access committees (UACs) as well as local trusted third parties. To verify the practicability and performance of the FDPG, a MII-wide projectathon was pursued to test data sharing concepts and tools. Four research projects were prepared, submitted, and managed using the FDPG platform.

3. Results

3.1. NSC working group results

The regulatory framework for MII-wide data sharing has been described in UML-based process maps and associated legal documents, such as data/biosample use regulations, contracts, and the data use project proposal form. Results have been discussed and agreed upon in the cross-consortia WG “data sharing,” shared with and consented to by the legal departments of all German university hospitals, and finally approved by the NSC.

To allow the use of patient data, documented primarily within hospitals’ patient care processes for medical research, a comprehensive discussion and consensus process has been initiated by the WG “consent” to define a template for a modular Broad Consent patient information/ consent form. After numerous rounds of discussion and revision, it was approved by all German state data protection officers and a WG of all German ethics committees in 2020 [10]. In summer 2021, German university hospitals have started “rollout projects” to implement the patient information and broad consent collection process into local workflows. About 100,000 patients have already signed this consent.

The MII-wide “information model” for a harmonized representation of data to be shared across all German DIC is one of the major achievements of the WG “interoperability.” HL7 FHIR resources form the basis of the MII CDS. Until today, official versions of the six basic CDS modules - patient, encounter, diagnosis, procedure, laboratory data, and medication data, as well as three extension modules - consent, biosample, and intensive care medicine, have been developed, balloted, and approved by the German HL7 community. Implementation guides (IG) are available as simplifier IGs [11]. Data from 7.6 million patients are currently available, including data items for more than 152 million lab results, 85 million diagnoses, and 37 million procedures.

3.2. CODEX and ABIDE_MI results

The COVID-19 data exchange platform (CODEX) was initiated as a joint action of the four MII consortia to tackle the COVID-19 pandemic [8]. One of the platform components was a central feasibility portal accessing the federated FHIR servers in German university hospitals. The architecture and technology for those developments were based on the MII concepts and design principles supporting smooth scalability even within the more complex and comprehensive structures of the MII. Thus, the MII ABIDE_MI project extended the small underlying datasets from COVID-19 patients to encompass comprehensive clinical data from all hospital patients based on the six basic modules of the CDS, the consent module, and the biosample module.

In October 2022, 27 DIC were able to integrate their FHIR servers with routinely collected clinical data into the FDPG framework. Eight biobanks integrated biosample information into DIC FHIR servers. The comprehensive architecture of the feasibility tool, its middleware as well as decentralized local components has been described by Gruendner et al. [12] and Rosenau et al. [13]. The complete system was demonstrated at the MII symposium.

3.3. Projectathon results

New processes and tools for the infrastructure are regularly evaluated in MII-projectathons. The seventh MII-projectathon was dedicated to evaluating the regulatory

framework and the FDPG-tools in combination with tasks carried out by administrative staff, as well as the data preparation and delivery process during the execution of four real research projects (comprising federated as well as centralized analysis). 31 German university hospitals were requested to deliver data. 30 DIC responded to the proposed projects, with 21 DIC participating in at least one project. Overall 42,7 % (53 out of 124) responses among the four projects were positive. Thus, sufficient data could be provided for all projects. The main reasons for not providing data were the need for patient consent forms available and the missing implementation of MII CDS modules. All contracts were administered and distributed to participating sites. Currently, signatures are being collected. Data sharing will be initiated once projects are published in the transparency registry.

Regarding the application of the research proposal management module: we learned that too much manual rework and communication to applicants was still necessary by FDPG staff to finalize project proposals with the researchers. This is mainly due to the current lack of FAQs and information in the FDPG web interface. For instance, free text data entry fields for project proposals were unclear and required further clarification. Similar problems emerged regarding user interfaces designed for DIC and UACs. Implemented processes were still novel and required additional clarification by FDPG staff. Integrating decision-making processes into local DIC processes does function but will need to be further optimized in the future. A time critical issue is the laborious process of signing contracts, as not all university hospitals accept digital signatures. Paper-based signature collection – considering the large number of contract partners – is time consuming and represents a considerable burden for all involved. Further, free text descriptions of the required data were not specific enough, requiring clarification cycles between DIC, FDPG staff, and data requestor. In a future release of the FDPG framework, we will provide users with a hierarchical data catalogue, allowing the precise selection of the required data elements to avoid misunderstandings between the requestor and the data provider. We expect additional optimization requirements concerning the processing pipeline with respect to data preparation and delivery when all projects are finished. However, beyond all technical support, there will need to be further organizational process optimization and governance-level decisions to appropriately address the important non-technical issues in data sharing mediated through the FDPG.

4. Discussion and Outlook

The MII aims to share data across all German university hospitals for research. The MII Symposium and the seventh projectathon have shown that MII is close to this goal but far from being lean, fast, and efficient in its implementation of all processes, which is necessary to save operational resources and for enabling high-quality research. To date, due to the results of the CODEX and ABIDE_MI projects, the FDPG is able to give researchers an impression of what data they can request from the university hospitals. After conducting a feasibility analysis, researchers can write a project proposal, and the FDPG supports spreading this proposal to all participating DIC. Local processes have been established at all German university hospitals to review those proposals within their UACs and then accept or reject the proposed project and provide data. As a final regulatory step, a joint data use contract is signed by all data providing sites and the respective project coordinator based on the MII data use contract template. Data provided by the university hospitals for central data pooling and central data analysis were – in the

recent projectathon - collected by one DIC. The individual datasets were integrated into a single dataset and securely provided to the project coordinator. Those results illustrate the success of the presented approach. Most sites responded to the data requests with a high proportion of UACs approving data access. This showcases the acceptance of the processes, especially considering the novelty of this framework.

However, data harmonization across all universities remains challenging. The MII CDS serves as a good baseline but is insufficiently precise for the use-case of federated queries that requires strong data harmonization to fulfil performance requirements. Further restrictions on top of the MII CDS are therefore necessary to ensure interoperability, and shared tooling is required to guarantee data quality across sites. Integration of a FHIR terminology server proved essential to provide a hierarchic search ontology [13] and will in the future be crucial to verify data accuracy and quality. Further enhancements of the FDPG and extensions to the comprehensive FDPG ecosystem (to be implemented in central MII infrastructure projects of the MII 2023-2026 funding phase) will, e.g., add functionality to select required data items from an automatically generated data element catalogue based on the FHIR profiles of the MII CDS, to then automatically extract such data from the DIC FHIR servers and preprocess it for secure data delivery between all data providers, a central data integrating service and the final project coordinator. Further, the contract pipeline is being prepared to support digital signatures. All those steps need to be transparent to researchers and DIC staff, encapsulated within FDPG components and workflows.

References

- [1] Gehring S, Eulenfeld R. German Medical Informatics Initiative: Unlocking Data for Research and Health Care. *Methods Inf Med.* 2018;57(S 01):e46-e49.
- [2] Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data Integration for Future Medicine (DIFUTURE). *Methods Inf Med.* 2018 Jul;57(S 01):e57-e65.
- [3] Haarbrandt B, Schreiweis B, Rey S, et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. *Methods Inf Med.* 2018;57(S 01):e66-e81.
- [4] Prokosch HU, Acker T, Bernarding J, et al. MIRACUM: Medical Informatics in Research and Care in University Medicine. *Methods Inf Med.* 2018;57(S 01):e82-e91.
- [5] Winter A, Stäubert S, Ammon D, et al. Smart Medical Information Technology for Healthcare (SMITH). *Methods Inf Med.* 2018;57(S 01):e92-e105. doi:10.3414/ME18-02-0004
- [6] Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative. *Methods Inf Med.* 2018;57(S 01):e50-e56. doi:10.3414/ME18-03-0003.
- [7] <https://www.medizininformatik-initiative.de/en/start> [last visited on December 4th 2022].
- [8] Prokosch HU, Bahls T, Bialke M, et al. The COVID-19 Data Exchange Platform of the German University Medicine. *Stud Health Technol Inform.* 2022;294:674-678.
- [9] Prokosch HU, Baber R, Bollmann P, Gebhardt M, Gruendner J, Hummel M. Aligning Biobanks and Data Integration Centers Efficiently (ABIDE_MI). *Stud Health Technol Inform.* 2022;292:37-42.
- [10] Zenker S, Strech D, Ihrig K, et al. Data protection-compliant broad consent for secondary use of health care data and human biosamples for (bio)medical research: Towards a new German national standard. *J Biomed Inform.* 2022;131:104096.
- [11] <https://www.medizininformatik-initiative.de/de/uebersicht-ueber-versionen-der-kerndatensatz-module> [last visited on February 22, 2023].
- [12] Gruendner J, Deppenwiese N, Folz M, et al. The Architecture of a Feasibility Query Portal for Distributed COVID-19 Fast Healthcare Interoperability Resources (FHIR) Patient Data Repositories: Design and Implementation Study. *JMIR Med Inform.* 2022;10(5):e36709.
- [13] Rosenau L, Majeed RW, Ingenerf J, et al. Generation of a Fast Healthcare Interoperability Resources (FHIR)-based Ontology for Federated Feasibility Queries in the Context of COVID-19: Feasibility Study. *JMIR Med Inform.* 2022;10(4):e35789.