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

doi:10.3233/SHTI240751

Germany's Approach for the Secondary Use of Federated Real-World Data in the German Portal for Medical Research Data

Philip KLEINERT[a,1], Marie GEBHARDT [a], Karoline BUCKOW [a], Julian GRUENDNER [b], Thomas GANSLANDT [b], Hans-Ulrich PROKOSCH [b] and Sebastian C. SEMLER [a]
[a] *TMF - Technology, Methods, and Infrastructure for Networked Medical Research, Berlin, Germany*
[b] *Chair of Medical Informatics, University of Erlangen-Nürnberg, Erlangen, Germany*

Abstract. In the light of big data driven clinical research, fair access to real world clinical health data enables evidence to improve patient care. Germany's healthcare system provides an abundant data resource but unique challenges due to its federated nature, heterogeneity and high data-protection standards. The Medical Informatics Initiative (MII) developed concepts that are being implemented in the German Portal for Medical Research Data (FDPG) to grant access to distributed data-sources across state borders. The portal currently provides access to more than 10 million patient resources containing hundreds of millions of laboratory parameters, diagnostic reports, administered medications, procedures and specimens. Upcoming datasets include among others oncological data, molecular analysis results and microbiological findings. Here, we describe the philosophy, implementation and experience behind the framework: standardized access processes, interoperable fair data, software for in depth feasibility requests, tools to support researchers and hospital stakeholders alike as well as transparency measures to provide data use information for patients. Challenges remain to improve data quality and automatization of technical and organizational processes.

Keywords. Real-World-Data, Secondary Use, Clinical Research Informatics, Federated IT Infrastructure

1. Introduction

Secondary use of clinical health data is essential to generate evidence for improved healthcare of patients as well as decision making of clinicians and regulators.
Various national (PCORnet [1], SPHN [2,3]) and international (OHDSI [4]) networks were founded, each providing frameworks for accessing federated real world data. Germany faces unique challenges to make electronic health records available as strong data protection regulations as well as heterogeneity across the health care sectors and in state regulations exist. Data protection regulations vary across states and their interpretation varies across university hospitals. Varying electronic health record systems from different EHR vendors are applied in the German (university) hospital landscape and even hospitals using the same EHR system have typically stored their data in

[1] Corresponding Author: Philip Kleinert; E-mail: Philip.Kleinert@tmf-ev.de.

different internal data structure and formats, making data sharing between hospitals challenging. On the other hand, Germany's publicly funded (through social insurance contributions) health care system, that provides access to high quality healthcare to all individuals, is an abundant resource of detailed and in-depth patient records. While various public and private health insurance options exist, population wide insurance is guaranteed with every individual carrying an individual health insurance number.
The German Medical Informatics Initiative (MII) was founded in 2015 to overcome these challenges [5] and make patient care data sharing possible across all German university hospitals For this purpose data integration centers (DIC) were created at each university hospital to establish local organizational and technical tools and infrastructures (e.g. data extraction, harmonization and integration processes, data pseudonymization and managing patient consents) and linking those with one central data access portal (the German Portal for Medical Research Data, Forschungsdatenportal für Gesundheit = FDPG).

The objective of this publication is to describe the measures which led to the implementation of the FDPG), which was set in routine use in October 2022 [6], and to further elaborate the challenges which had to be mastered, the current achievements and plans for future extensions.

2. Methods

The FDPG applies concepts and regulations developed by the MII working groups and approved by its national steering committee and functions as the central coordination and access point for. It consists of four main components: transparency registry, application portal, feasibility portal and the DIC-Dashboard.

All aspects from heterogeneous regulations, application processes and project implementation had to be considered simultaneously and agreed upon by all stakeholders: data protection, ethics, consent, professional law (doctors' duty of confidentiality), legal frameworks and contracts. Four main documents were developed: participation framework contract, data usage request form, usage agreement, terms of data/biospecimen use.

The legal framework is the standardized foundation for multi-site data sharing [7]. The FDPG assumes the role of a coordinating body in all processes. The access process is based on a data/biospecimen request form supported by an ethics approval and a feasibility analysis protocol. After a formal review by FDPG staff those documents are distributed to the requested DIC. Use and access committees (UAC) at the participating hospitals review this information and grant or deny access to the local data-pool. UAC decisions are collected centrally at the FDPG until all votes are given and an applicant receives the data provisioning information from all DIC. Next, data/biospecimen usage agreements are rolled out to DIC providing data for the respective project to receive legal signatures. After the reception of the usage agreement with all required signatures the FDPG staff initiates the data sharing process. There is no negotiation about the usage agreements since all terms have been previously agreed upon by all university hospitals´ legal departments.

Two basic modes of data sharing are being supported: central analysis (based on patients´ broad consent [8]) and federated analysis, e.g. supported by an adapted DataSHIELD framework.

The FDPG incorporates the terms of use within its application portal, which combines registration, application, administration, contract formation as well as monitoring of the provisioning of data and biospecimen.

The HL7 FHIR (Fast Healthcare Interoperability Resources) [9,10] standard is applied to define the MII core data set (CDS). Semantic interoperability is achieved by using international terminologies such as ICD10, SNOMED CT and LOINC. Data elements defined for the CDS are grouped in modules. Seven basic modules are required by all DIC (Person, Consent, Treatment Case, Medication, Diagnosis, Procedures and Lab results). Further, 9 extension modules are being provided or will be provided in the near future (Oncology, Symptoms, Pathology, Imaging, Microbiology, Molecular Genetics, Intensive Care, Bio Samples, Research Projects). All FDPG components and a simplified data sharing process is shown in Figure 1.

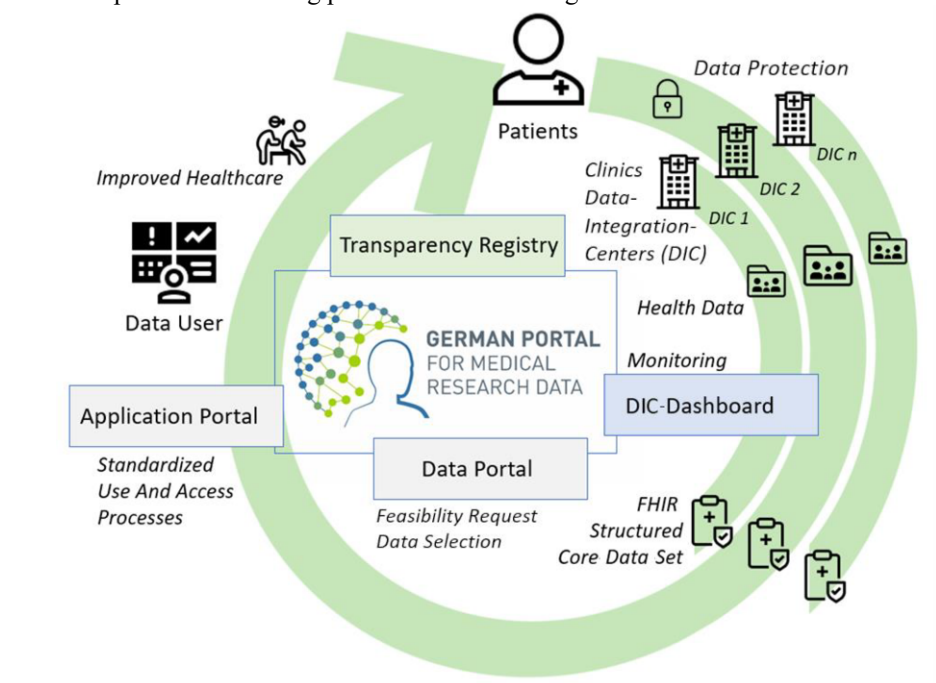

Figure 1. Overview of the FDPG components in the data flow from patient to data user. Transparency Registry is publicly available while Application and Data Portal require registration. The DIC-Dashboard is only available to staff members at the clinics.

3. Results

Based on these concepts the FDPG was developed and set into routine use in October 2022. All FDPG software is open source. Currently more than 500 researchers are registered.

The transparency portal provides standardized information of data use projects, incorporating also easy language summaries for lay persons. Currently 30 data use projects have been applied for, 19 projects are contracted and registered in the transparency portal, nine projects have already received datasets.

To reduce individual administration efforts at each participating data holder, all multi-site applications are being coordinated centrally by the FDPG staff and subjected to a strict formal review.

Federated feasibility requests are supported using inclusion and exclusion criteria based on specifications within the CDS [11]. Results are obfuscated for data protection. More than 10 million patient resources are currently available containing hundreds of millions of laboratory measurements, administered medications, procedures and biospecimen.

4. Discussion and Conclusions

The FDPG is set to provide access to Germanys abundant resources of FHIR structured hospital-based real-world health data by applying standardized processes, data protection regulations as well as tools to automate and accelerate all data sharing processes.

While privacy preserving measurements are a main focus of this framework it comes at a cost. The distributed data sources as well as local use and access votes and contract procedures lead to time consuming consecutive process steps. Establishing process improvements for faster application pipelines are challenges lying ahead.

Constant effort must be invested to extend data sources within the DIC. Regular community events (e.g. projectathons, symposia) and complete transparency in all development steps motivate progress.

Another challenge is to train researchers to plan and propose high quality research projects while applying the available FDPG tools to reduce the consulting effort for FDPG staff and the effort at the local sites' UACs for review. Overarching improvements (such as the implementation of new modules and overarching software updates) are bound to a comprehensive yearly release cycle.

The FDPG is a crucial part of the MII infrastructure, which makes real world clinical health data available for research. Health care data can now, for the first time, be accessed for research across 34 German university hospitals which will lead to rapid development of tools aiding decision making for clinicians and improve healthcare for patients. The MII is also receiving a boost from current national legislation in Germany, the Health Data Usage Act (GDNG), which came into force at the end of March 2024 allows medical institutions to analyze patient data for medical research in the public interest, even without consent. The established infrastructures can now also be used at an early stage to answer questions that would require case numbers for which sufficient consent has not yet been obtained, and thus provide benefits for patient care at an earlier stage. Furthermore, the infrastructures set up by the MII - in particular the distributed, harmonized DIC and the FDPG as a central access point - are predestined to take on important task in the upcoming architecture for connecting Germany to the European Health Data Space (EHDS).

References

[1] Fleurence RL, Curtis LH, Califf RM, Platt R, Selby JV, Brown JS. Launching PCORnet, a national patient-centered clinical research network. Journal of the American Medical Informatics Association. 1. Juli 2014;21(4):578–82.

[2] Touré V, Krauss P, Gnodtke K, Buchhorn J, Unni D, Horki P, Raisaro JL, Kalt K, Teixeira D, Crameri K, Österle S. FAIRification of health-related data using semantic web technologies in the Swiss Personalized Health Network. Sci Data. 10. März 2023;10(1):127.

[3] Lawrence AK, Selter L, Frey U. SPHN - The Swiss Personalized Health Network Initiative. Stud Health Technol Inform. 16. Juni 2020;270:1156–60.

[4] Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong ICK, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. Stud Health Technol Inform. 2015;216:574–8.

[5] Semler S, Wissing F, Heyder R. German Medical Informatics Initiative: A National Approach to Integrating Health Data from Patient Care and Medical Research. Methods Inf Med. Juli 2018;57(S 01):e50–6.

[6] Prokosch HU, Gebhardt M, Gruendner J, Kleinert P, Buckow K, Rosenau L, Semler SC. Towards a National Portal for Medical Research Data (FDPG): Vision, Status, and Lessons Learned. In: Hägglund M, Blusi M, Bonacina S, Nilsson L, Cort Madsen I, Pelayo S, Moen A, Benis A, Lindsköld L, Gallos P, Herausgeber. Studies in Health Technology and Informatics [Internet]. IOS Press; 2023 [zitiert 12. Dezember 2023]. Verfügbar unter: https://ebooks.iospress.nl/doi/10.3233/SHTI230124

[7] Kirsten T, Kleinert P, Gebhardt M, Drepper J, Andreeff AK, Prasser F, Kohlbacher O. Grundlagen für die wissenschaftliche Nutzung umfangreicher Versorgungsdaten in Deutschland – Ergebnisse der AG Data Sharing der Medizininformatik-Initiative. Bundesgesundheitsbl [Internet]. 29. April 2024 [zitiert 6. Juni 2024]; Verfügbar unter: https://link.springer.com/10.1007/s00103-024-03880-y

[8] Zenker S, Strech D, Ihrig K, Jahns R, Müller G, Schickhardt C, Schmidt G, Speer R, Winkler E, Von Kielmansegg SG, Drepper J. 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. Journal of Biomedical Informatics. Juli 2022;131:104096.

[9] Sass J, Zabka S, Essenwanger A, Schepers J, Boeker M, Thun S. Fast Healthcare Interoperability Resources (FHIR®) Representation of Medication Data Derived from German Procedure Classification Codes (OPS) Using Identification of Medicinal Products (IDMP) Compliant Terminology. In: Röhrig R, Beißbarth T, Brannath W, Prokosch HU, Schmidtmann I, Stolpe S, Zapf A, Herausgeber. Studies in Health Technology and Informatics [Internet]. IOS Press; 2021 [zitiert 17. März 2024]. Verfügbar unter: https://ebooks.iospress.nl/doi/10.3233/SHTI210074

[10] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. In: Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems [Internet]. Porto, Portugal: IEEE; 2013 [zitiert 17. März 2024]. S. 326–31. Verfügbar unter: http://ieeexplore.ieee.org/document/6627810/

[11] Gruendner J, Deppenwiese N, Folz M, Köhler T, Kroll B, Prokosch HU, Rosenau L, Rühle M, Scheidl MA, Schüttler C, Sedlmayr B, Twrdik A, Kiel A, Majeed RW. 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. 25. Mai 2022;10(5):e36709.