Healthcare of the Future 2022
T. Bürkle et al. (Eds.)

doi:10.3233/SHTI220317

Aligning Biobanks and Data Integration Centers Efficiently (ABIDE_MI)

Hans-Ulrich PROKOSCH[a,1], Ronny BABER[b,c], Paula BOLLMANN[b], Marie GEBHARDT[d], Julian GRUENDNER[a] and Michael HUMMEL[e,f]
[a]*Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[b]*Leipzig Medical Biobank, Leipzig University, Leipzig, Germany*
[c]*Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Leipzig, Germany*
[d]*TMF - Technology, Methods, and Infrastructure for Networked Medical Research, Berlin, Germany*
[e]*German Biobank Node, Charité Universitätsmedizin Berlin, Berlin, Germany*
[f]*Central Biobank Charité (ZeBanC), Institute of Pathology, Charité Universitätsmedizin Berlin, Berlin, Germany*

Abstract. ABIDE_MI is a complementary funded 18 months project within the German Medical Informatics Initiative (MII), which aims to align IT infrastructures and regulatory/governance structures between biobanks/biobanking IT and the MII data integration centres (DIC) at German university hospitals. A major task in 2021 was the systematic collection of all documents describing rules, as well as proposal/contract templates for data and biosample use and access at each of the participating 24 university hospitals and their comparison with MII-wide consented data sharing principles, documents and governance structures. This comparison revealed large heterogeneity across the ABIDE_MI sites and further, redundant structures/regulations currently established at the German university hospitals. A second task was the design and stepwise development of an IT network infrastructure with central components (data and biosample query portal) and decentralized standardized FHIR servers to capture the standardized FHIR-based core data set modules (resources) defined within the MII working group "Interoperability". Subsequent steps in the project are the harmonization of the data and biosample sharing governance/regulation frameworks at each ABIDE_MI site, creating synergies for the research infrastructures at the German university hospitals and to link those resources to the German Portal for Medical Research Data and with the BBMRI-ERIC Directory and Negotiator tools.

Keywords. Biobank networks, real world data, medical informatics initiative, data and biosample sharing

1. Introduction

Real world data analysis in medicine today relies on the availability of clinical data as well as information about biospecimen collected during clinical care processes [1].

[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.

The harnessing and cross-consortial research use of such data is one of the major goals of the Federal Ministry of Education and Research (BMBF)-funded Medical Informatics Initiative (MII) and its four consortia [2]. Further, networking and collaboration of German biobanks to leverage high quality biobanking throughout Germany has been funded in Germany since 2011, starting with a program for the establishment of centralized biobanks (cBMB Program), the subsequent initiation of the German Biobank Node (GBN, 2013) and its German Biobank Alliance (GBA, 2017) [3, 4]. Both German initiatives (MII and GBN/GBA) aim at establishing federated data networks with distributed data repositories and central feasibility platforms. Unfortunately, until 2020, almost no major synergies between GBA and MII had been realized and partially parallel structures (on technical as well as organizational/regulation level) existed within the German university hospitals. Thus, it is the aim of the ABIDE_MI project (a complementary BMBF-funded project within the MII) to align such structures between biobanks and biobanking IT as well as the data integration centres (DIC), to create synergies for the research infrastructures at the German university hospitals and to link those resources to a central feasibility portal. The objective of this publication is to describe the results achieved after half of the project duration and to illustrate the plan for the second project phase.

2. Methods

ABIDE_MI started in May 2021 and comprises 24 German University Hospitals with 24 DIC and 25 biobanks (one of the University hospitals owns two biobanks at its separate locations in two cities). The project comprises central IT development tasks (e.g. developing a graphical user interface for cohort identification and biosample search (feasibility tool) linked via middleware components to distributed FHIR servers), deployment of such IT tools for all participating university hospitals, and decentralized work packages to be pursued within all university hospitals.

The latter are split into

- organizational tasks, like analyzing all governance and regulatory documents as well as committees associated with data and biosample sharing established at each hospital locally (typically based on previous biobank establishments as well as DIC establishments), and to compare them with the respective documents and regulations agreed upon in the MII working groups consent and data sharing.
- technical developments, such as developing the ETL processes to provide clinical data formatted according to the FHIR implementation guidelines of the basic modules of the MII core data set (person, encounter, diagnosis, procedure, laboratory, medication; see https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set for details), the MII consent information module and biosample data, based on the MII biosample FHIR module.

Further, the IT framework should be designed in such a way, that the central feasibility tool should smoothly fit into the German Portal for Medical Research Data (Deutsches Forschungsdatenportal für Gesundheit = FDPG), currently being implemented by the MII coordination centre, and it should also provide an integration

pathway into the European biobank network BBMRI ERIC, with its federated search tools, the BBMRI Directory and the Sample Negotiator [5, 6].

For the feasibility tool development we built on previous project experiences and developments, e.g. the German Biobank Alliance, the German AKTIN emergency ward registry [7], the MIRACUM project [8] and the development of a national Covid-19 data exchange platform (CODEX) within the German network university medicine (NUM) [9]. The software development is pursued in an agile development process by a team of developers spread over six locations. The comprehensive software architecture relies on a set of micro services interacting with each other to translate the feasibility query user input in a predefined structured query syntax, transfer the queries securely to the local implementations of the network partners and execute the queries on a locally installed FHIR server.

The complete project is coordinated using the Atlassian® confluence collaboration platform and biweekly web conferences. In such regular web conferences biobanking representatives, DIC representatives of all partner sites and the coordination team met to discuss the status of project deliverables/milestones and particularly focus on the alignment of regulations/governance structures established on one side in the local biobank environments and on the other side within the data integration centre environments. Those meetings were especially helpful to present the arguments for and reasons behind all such regulations and structures and achieve a common understanding of the respective historical development within both communities. All preexisting documents from each of the partners (DICs and biobanks) describing e.g. consenting procedures, governance structures, application for the use and material transfer agreements for the exchange of biospecimen and associated clinical data were collected in the collaboration platform and then systematically compared to the central documents/regulations established within the MII, to identify gaps between the needs of the biobanking community and the regulations defined within the MII.

3. Results

3.1. Organizational Tasks

The major result of the first project phase is a technical report describing the current status of all regulations and governance structures established at the 24 participating university hospitals, and their comparison with their counterparts centrally agreed upon in the MII working groups. The documents identified to be important in this context are summarized in table 1.

The gap analysis depicts a heterogeneous situation, with some university hospitals having established biobanks (and the respective governance structures and use/access rules) many years ago and others, which had just recently established a centralized biobank at their university hospital, which enables to directly rely on the structures and regulations defined within the MII. One of the ABIDE_MI partners did neither have a central biobank nor a data integration centre (because up until now it was only a networking partner without DIC in a MII consortium) established until 2021 but uses the ABIDE_MI project to establish both such institutions at its location (this partner was not included in the following statistics).

With respect to the local adoption of the MII overall standardized use and access rules, 16 DIC reported, that they have already implemented this governance instrument,

6 were currently in the process of implementing it and one did not have any data/biosample use and access rules implemented yet. Compared to this, only 8 biobanks had already implemented use and access rules for biosample use in research projects, 3 were in the process of implementing local use and access rules. Only 3 biobanks had adapted the MII overall standardized use and access rules and 8 were in the process of implementing this standardized MII governance instrument. One biobank has just started to work on defining its use and access rules. 3 biobanks were not implementing any use and access rules yet (November 2021). In summary at 5 university hospitals DIC and biobank had already adapted MII overall standardized use and access rules. Use and access committees with already joined boards for biosamples and data existed at 8 university hospitals whereas in 15 university hospitals the data use and access committee and the biosample use and access committee were separate boards.

Table 1. Data/biosample use and access documents analyzed locally and compared with MII consented documents

local documents	MII consented documents
local "broad consent" based on the biobanking broad consent template (template from Arbeitskreis Medizinischer Ethikkommissionen e.V.)	MII "broad consent" template see https://www.medizininformatik-initiative.de/en/template-text-patient-consent-forms for details
local "broad consent" based on the MII broad consent template	
local biobank/DIC by-laws and/or statutes	MII data sharing process model
local biobank/DIC use and access rules	MII overall standardized use and access rules
local biobank/DIC templates for data/biosample use proposals	MII template for data use proposals
local biobank/DIC templates for data/biosample use contracts	contract template governing the use of data and biosamples in the MII

The local board of director approval of the MII broad consent template (which also includes a module for biosample use) was achieved at 20 university hospitals, nevertheless only at 7 of those hospitals the respective biobanks had started to use this broad consent template in their routine processes. 15 biobanks were still using a local biobanking consent form and in 1 biobank the consent process implementation was work in progress.

3.2. Technical Development

The IT development could benefit from earlier developments already pursued within the CODEX project and established a very similar network architecture. While in CODEX the focus was only on Covid-19 patients and thus the data items to be provided were reduced to the GECCO (German Corona Consent) dataset [10], the ABIDE_MI projects aims at providing access to clinical data and biosample information from all hospital patients with a much larger dataset, defined by the six basic modules of the MII core dataset, the consent module and the biosample core dataset module. Therefore, also the ontology used to build queries for ABIDE_MI was changed and created from the FHIR implementation guides of the above mentioned MII dataset modules. Further, the query

UI was enhanced with new features such as combination logic for linked data items and temporal restrictions.

The ABIDE_MI IT framework currently comprises

- a user interface (feasibility UI),
- a backend service which translates the user input into a standardized format (structured query) based on an ontology service [11] and
- an execution service, which can process the standardized format, convert it to queries for a FHIR server and execute the query (this execution service is distributed to all partners in the network).
- middleware components to provide a secure transportation of queries and query results between the central component and the decentral execution services.

The developed tools were built to support any FHIR server, which provides either a FHIR search interface or the ability to execute CQL queries, allowing the participating sites to choose which FHIR server to use.

4. Discussion and outlook

Health care integrated biobanks are usually closely cooperating with the institutional departments of pathology and clinical laboratory. Data governance of the biobanks is typically limited to the direct biosample information (managed e.g. in a biobank management system) and a small set of clinical data arising from the respective departments "own" departmental IT system and other data sources. Therefore, sample search tools, such as e.g., the GBA sample locator [12] are restricted to only a few clinical items. However, as illustrated e.g., by Castro et al. [13], Geiger et al. [14] and Lawrence et al. [15] precision medicine research in the future will require high quality assembled biosamples annotated with a comprehensive spectrum of clinical and molecular data. The integration of those data from historically separated data silos is one of the major goals of the MII. Thus, close coordination between biobanks and data integration centers is inevitable for future innovative medical research. In times of limited resources there is an utmost need to eliminate redundancies at the organizational and technical level.

This process was successfully initiated within ABIDE_MI with the described gap analysis concerning the organizational regulations and structures within the German university hospitals. It is the challenge now for the ABIDE_MI partners to synergistically align their data and biosample frameworks locally and to integrate them into the overall MII data/biosample sharing framework, especially the future German Portal for Medical Research Data. On the other hand, the review of the documents and intense discussions between the biobanking and MII/DIC communities have identified particular needs arising from the more complex process of sharing biosamples (as a limited resource) which are now brought into the MII working groups "consent" and "data sharing" to implement them in the MII framework in order to fully satisfy also the respective biobanking needs. As an example, an ABIDE_MI taskforce was initiated to develop a template for a material transfer agreement (MTA) to be added to the contract template governing the use of data and biosamples in the MII.

In February 2022 the first release deployed for the IT infrastructure was implemented at all ABIDE_MI sites and filled within a small projectathon with the respective MII core data set modules extracted and harmonized at each university hospital. At this point in time six DIC were capable to connect to the feasibility tool with their routine data FHIR server. In parallel we also pursued a usability evaluation of the new feasibility tool in order to receive user feedback and then further enhance the UI and the implemented search ontology during the second half of the project.

References

[1] Maissenhaelter BE, Woolmore AL, Schlag PM. Real-world evidence research based on big data: Motivation-challenges-success factors. Onkologe (Berl). 2018;24 (Suppl 2): 91-98.
[2] 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.
[3] Schüttler C, Prokosch HU, Hummel M, Lablans M, Kroll B, Engels C; German Biobank Alliance IT development team. The journey to establishing an IT-infrastructure within the German Biobank Alliance. PLoS One. 2021 Sep 22;16(9):e0257632. doi: 10.1371/journal.pone.0257632.
[4] Baber R, Hummel M, Jahns R, et al. Position Statement from the German Biobank Alliance on the Cooperation Between Academic Biobanks and Industry Partners. Biopreserv Biobank. 2019;17(4):372-374.
[5] Holub P, Swertz M, Reihs R, van Enckevort D, Müller H, Litton JE. BBMRI-ERIC Directory: 515 Biobanks with Over 60 Million Biological Samples. Biopreserv Biobank. 2016;14(6):559-562. doi:10.1089/bio.2016.0088.
[6] Reihs R, Proynova R, Maqsood S, et al. BBMRI-ERIC Negotiator: Implementing Efficient Access to Biobanks. Biopreserv Biobank. 2021;19(5):414-421. doi:10.1089/bio.2020.0144.
[7] Brammen D, Greiner F, Kulla M, Otto R, Schirrmeister W, Thun S, Drösler SE, Pollmanns J, Semler SC, Lefering R, Thiemann VS, Majeed RW, Heitmann KU, Röhrig R, Walcher F; [AKTIN - The German Emergency Department Data Registry - real-time data from emergency medicine : Implementation and first results from 15 emergency departments with focus on Federal Joint Committee's guidelines on acuity assessment]. Med Klin Intensivmed Notfmed. 2020 Dec 21:1–10. German. doi: 10.1007/s00063-020-00764-2. Epub ahead of print. PMID: 33346852; PMCID: PMC7750913.
[8] 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.
[9] CODEX | COVID-19 Data Exchange Platform, Netzwerk Universitätsmedizin. https://www.netzwerk-universitaetsmedizin.de/projekte/codex [last visited on January 3rd, 2022]
[10] Sass J, Bartschke A, Lehne M, Essenwanger A, Rinaldi E, Rudolph S, Heitmann KU, Vehreschild JJ, von Kalle C, Thun S. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. BMC Med Inform Decis Mak. 2020 Dec 21;20(1):341.
[11] Rosenau L, Kroll B, Kiel A, Majeed RW, Szimtenings L, Köhler T, Ingenerf J, Prokosch HU, Gruendner J. Generation of a Fast Healthcare Interoperability Resources (FHIR)-based Ontology for federated Feasibility Queries in the context of COVID-19. JMIR Medical Informatics accepted for publication.
[12] Schüttler C, Prokosch HU, Sedlmayr M, Sedlmayr B. Evaluation of Three Feasibility Tools for Identifying Patient Data and Biospecimen Availability: Comparative Usability Study. JMIR Med Inform. 2021 Jul 21;9(7):e25531. doi: 10.2196/25531. Erratum in: JMIR Med Inform. 2021 Oct 8;9(10):e33105.
[13] Castro VM, Gainer V, Wattanasin N, Benoit B, Cagan A, Ghosh B, Goryachev S, Metta R, Park H, Wang D, Mendis M, Rees M, Herrick C, Murphy SN. The Mass General Brigham Biobank Portal: an i2b2-based data repository linking disparate and high-dimensional patient data to support multimodal analytics. J Am Med Inform Assoc. 2021 Nov 28:ocab264. doi: 10.1093/jamia/ocab264. Epub ahead of print.
[14] Geiger, J, et al. 2018 Hospital-integrated Biobanking as a Service – The Interdisciplinary Bank of Biomaterials and Data Wuerzburg (ibdw). Open Journal of Bioresources, 5: 6.
[15] Lawrence E, Sims J, Gander A, Garibaldi JM, Fuller B, Davidson B, Quinlan PR. The Barriers and Motivators to Using Human Tissues for Research: The Views of UK-Based Biomedical Researchers. Biopreserv Biobank. 2020 Aug;18(4):266-273. doi: 10.1089/bio.2019.0138. Epub 2020 Jun 17. PMID: 32551838; PMCID: PMC7415886.