The COVID-19 Data Exchange Platform of the German University Medicine

Hans-Ulrich PROKOSCH[a,1], Thomas BAHLS[b], Martin BIALKE[b], Jürgen EILS[c], Christian FEGELER[d], Julian GRUENDNER[a], Birger HAARBRANDT[e], Christopher Hampf[b], Wolfgang HOFFMANN[b], Hauke HUND[d], Marvin KAMPF[f], Lorenz A KAPSNER[f], Piotr KASPRZAK[n], Oliver KOHLBACHER[g,l,m], Dagmar KREFTING[h], Jonathan M MANG[f], Michael MARSCHOLLEK[e], Sebastian MATE[f], Armin MÜLLER[c], Fabian PRASSER[c], Julian SASS[c], Sebastian SEMLER[i], Holger STENZHORN[g,j], Sylvia THUN[c], Sven ZENKER[k] and Roland EILS[c]

[a]*Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[b]*Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany*
[c]*Berlin Institute of Health at Charité – Universitätsmedizin Berlin, Berlin, Germany*
[d]*GECKO Institute, Heilbronn University of Applied Sciences, Heilbronn, Germany*
[e]*Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*
[f]*Medical Center for Information and Communication Technology, Universitätsklinikum Erlangen, Erlangen, Germany*
[g]*Institute for Translational Bioinformatics, University Medical Center, Tübingen, Germany*
[h]*Department of Medical Informatics, University Medical Center Göttingen, Göttingen, Germany*
[i]*TMF - Technology, Methods, and Infrastructure for Networked Medical Research, Berlin, Germany*
[j]*Institute for Medical Biometry, Epidemiology und Medical Informatics, Saarland University Medical Center, Homburg, Germany*
[k]*Staff Unit for Scientific & Medical Technology Development & Coordination (MWTek), Commercial Directorate; Institute for Medical Biometry, Informatics & Epidemiology; Department of Anesthesiology and Intensive Care Medicine, University Hospital Bonn, Bonn, Germany*
[l]*Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany*
[m]*Department of Computer Science, University of Tübingen, Tübingen, Germany*
[n]*Gesellschaft für wissenschaftliche Datenverarbeitung mbH, Göttingen, Germany*

Abstract. COVID-19 has challenged the healthcare systems worldwide. To quickly identify successful diagnostic and therapeutic approaches large data sharing approaches are inevitable. Though organizational clinical data are abundant, many of

---

[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.

them are available only in isolated silos and largely inaccessible to external researchers. To overcome and tackle this challenge the university medicine network (comprising all 36 German university hospitals) has been founded in April 2020 to coordinate COVID-19 action plans, diagnostic and therapeutic strategies and collaborative research activities. 13 projects were initiated from which the CODEX project, aiming at the development of a Germany-wide Covid-19 Data Exchange Platform, is presented in this publication. We illustrate the conceptual design, the stepwise development and deployment, first results and the current status.

Keywords. Real world data, data sharing network, Covid-19, pandemic preparedness

1. Introduction

Real world data analysis in medicine today relies on the availability of clinical data collected during clinical care processes [1]. The harnessing and cross-consortial 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]. The collection, exchange and joint analysis of diverse data referring to COVID-19 patients has been an important method of the network university medicine (NUM) that has been founded in April 2020 to understand and manage the COVID-19 pandemic in Germany. As one NUM project, the COVID-19 Data Exchange Platform CODEX has been designed and implemented by the MII partners to enable all university hospitals in Germany to harmonize, share and analyze COVID-19 real world data.

The objective of this publication is to describe the project´s conceptual approach, the stepwise development and deployment, first results and the current CODEX status.

2. Methods

To quickly initiate the CODEX project [3] most of the German university hospitals built on the data integration centers, governance, and methods established within the MII. CODEX brings together the competencies and previous work of more than 15 university hospitals and industrial partners making the results of their work available to the scientific community as open-source software. Further design principles were vendor neutrality and open application programming interfaces.

A first coordinated step within NUM was the interdisciplinary definition and specification of the FHIR-based GECCO (German Corona Consensus Dataset) dataset [4], with data elements and response options being semantically mapped to e.g., SNOMED CT and LOINC for international comparability. To quickly provide a harmonized deployable data repository environment we started with distributed i2b2 repositories, which had already proven their capabilities within MIRACUM [5] and a cross-consortial MII demonstrator study [6]. It was also decided to further build on the already established MII concepts, with federated data collection within routine clinical care environments, thus providing federated analysis options, but extend this with a central platform (similar to the N3C collaborative [8]; based on results from the MII HiGHmed consortium [7]). This enabled linkage with for example COVID-19 data from citizen apps. The implementation of a federated Trusted Third Party (fTTP) [9] enables cross-institutional as well as privacy-preserving record linkage (PPRL) and assures

compliance with the requirements of the European General Data Protection Regulation (GDPR). Ethics approval for data transfer to the central CODEX platform was given amongst others by the Erlangen University Ethics Committee (No. 500_20 B; R. Maas).

3. Results

3.1 The decentral CODEX nodes

Local data integration centers (DICs) at each of the university hospitals provide core data infrastructure for routine clinical care data extraction, harmonization, and integration. Based on these, CODEX implemented 'CODEX nodes' – harmonized local software infrastructures for providing the GECCO data. The nodes rely on a set of centrally provided software components, which had to be combined and integrated into the local IT infrastructures. The core components comprise:

- an **i2b2 repository** with patient demographics, diagnosis, and procedures as well as ETL processes to fill it from standardized German billing data records
- a **FHIR server** to provide the GECCO dataset
- the **Trusted Third Party components** to support local pseudonymization (gPAS®), record linkage (E-PIX®), consent management (gICS®), local provisioning of standardized FHIR consent resources and support for a FHIR-based linkage to the federated Trusted Third Party (fTTP) [9]
- the **'feasibility triangle'**, named after its three components, providing (1) a FHIR Server, (2) FHIR Search, respectively CQL, execution environments and (3) a middleware for secure linkage with the central feasibility user interface
- the **FHIR and business process engine** (BPE) servers of the HiGHmed Data Sharing Framework (DSF) [10] enable secure transport of GECCO datasets between local nodes and the central platform.
- an **EDC system** (e.g., REDCap [11]) to capture GECCO data which were not yet part of the electronic documentation at the German University Hospitals
- the **ODM2FHIR converter** to transform standardized EDC system exports in ODM format to the GECCO FHIR profile.

These components were developed, adapted, and pre-configured by consortium members and provided to all partners within the network together with templates for the required documentation (e.g., data protection concepts, IT security documentation) to accelerate its use in the highly regulated clinical environment. The development and deployment of these components were tightly coordinated through a series of weekly video conferences.

3.2 The central CODEX platform components

The central CODEX platform extends the federated analysis capabilities of the CODEX nodes, with the possibility of linking patient data across institutions and providing data usage for specific research projects. This is implemented by the following components:

- **the** central **research data platform** is based on EHRbase, an open-source software backend for electronic health records based on OpenEHR [12]. Access to

the platform is provided through a central portal enabling cohort selection and data export requests to answer specific research questions.

- **the DSF-**based **GECCO Transfer Hub** acts as a gatekeeper and orchestrates the transfer of data from the CODEX nodes to the central research data platform including a pseudonym exchange using the fTTP.
- **a** federated **Trusted Third Party** (fTTP) performs privacy-preserving record linkage and pseudonymization for the central platform during the data transfer process. The functionality is implemented using the open-source tools of the Trusted Third Party Greifswald [9].
- **a CODEX Dashboard** continuously provides anonymized information on the total number of patients with COVID-19 infection, level of care, length of stay in intensive care as well as age and gender [13]. Most recently, a MII core dataset FHIR specification compliant open-source processing engine to feed the dashboard from local FHIR stores has been provided.

A central **Data Use & Access Committee** reviews requests for data usage.

3.3 First CODEX-based research results

Based on the i2b2 deployments provided in early summer 2020 first research questions could be answered. These were for example:

- How did university hospital patient numbers change during the first German COVID-19 lockdown (data from 14 German university hospitals)? [14]
- What was the lethality of COVID-19 patients in the first half year of the pandemic in association with clinical risk factors (data from 18 German university hospitals)? [15]
- Was there an observable shift of radiotherapy use during the first wave of the COVID-19 pandemic (data from 14 German university hospitals) [16]?

Some of the university hospitals leveraged their i2b2 implementations and participated in the international consortium (4CE) of 96 hospitals across five countries (www.covidclinical.net) which aimed to use the largely untapped resource of EHR data to address critical clinical questions about COVID-19 [17].

4. Discussion and outlook

During a short project period a very comprehensive data sharing platform was established across all German university hospitals. The technical implementation, despite its complexity and the widely distributed developer team, could be smoothly established. Towards this end, the project nicely illustrates how integrating the different competencies from the whole university hospital landscape of Germany leads to synergies for an efficient joint data sharing platform development. However, it also illustrated the socio-technical hurdles within bureaucratic environments. The regulatory aspects (e.g., data provisioning contracts between the central platform and all university hospitals, obtaining positive ethics votes for the MII broad consent text extended with the data transfer aspects to the CODEX central platform), proved to be much more challenging. Thus, we could, by the end of 2021, provide more than 17,000 GECCO records within the distributed DIC repositories. To enable their research use local data

use and access committees had authorized provision of their data for the above mentioned federated data analysis projects. Positive ethics votes for transferring data to the central platform, in many hospitals however, could only be obtained very late in the project. Thus, the consent for central data transfer currently only exists for about 350 patients.

Based on the collaboration culture and joint infrastructure established in CODEX in 2021, two follow up projects have been initiated in 2022. The CODEX Routine Data Platform will ascertain sustainability of the established infrastructure by addressing both governance and maintenance tasks and providing deeper integration with MII structures. Starting in 2023 it will be extended to cover the complete MII core dataset. The CODEX+ project extends the achieved results in various directions, including interfacing more deeply with infrastructures previously developed in MII use cases, broadening the data to be included in the platform and linking with international projects.

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.

[3]   CODEX | COVID-19 Data Exchange Platform, Netzwerk Universitätsmedizin. https://www.netzwerk-universitaetsmedizin.de/projekte/codex [last visited on January 3rd, 2022]

[4]   Sass J, Bartschke A, Lehne M, et al. 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.

[5]   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.

[6]   Ganslandt C, Schaaf J, Schepers J, et al. Experiences from the National Demonstrator Study within the German Medical Informatics Initiative. AMIA Conference, Washington, 2019.

[7]   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.

[8]   Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021;28(3):427-443.

[9]   Bialke M, Penndorf P, Wegner T, et al. A workflow-driven approach to integrate generic software modules in a Trusted Third Party. J Transl Med. 2015;13:176.

[10] Hund H, Wettstein R, Heidt CM, Fegeler C. Executing Distributed Healthcare and Research Processes - The HiGHmed Data Sharing Framework. Stud Health Technol Inform. 2021 May 24;278:126-133.

[11] https://projectredcap.org/software/ [last visited on January 19th, 2022].

[12] https://openehr.org/ [last visited on January 19th, 2022].

[13] https://coronadashboard.ukbonn.de/ [last visited on January 19th, 2022].

[14] Kapsner LA, Kampf MO, Seuchter SA, et al. Reduced Rate of Inpatient Hospital Admissions in 18 German University Hospitals During the COVID-19 Lockdown.Front. Public Health. 2020;8:59411.

[15] Schüttler J, Mang JM, Kapsner LA, et al. Letalität von Patienten mit COVID-19: Untersuchungen zu Ursachen und Dynamik an deutschen Universitätsklinika. Anästh. Intensivmed 2021; 62:244–257.

[16] Medenwald D, Brunner T, Christiansen H, et al. Shift of radiotherapy use during the first wave of the COVID-19 pandemic? An analysis of German inpatient data. Strahlenther Onkol (accepted).

[17] Brat GA, Weber GM, Gehlenborg N et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. npj Digi. Med 3, 109, 2020.