German Medical Data Sciences 2022 - Future Medicine
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI220804

CODEX Meets RACOON - A Concept for Collaborative Documentation of Clinical and Radiological COVID-19 Data

Marvin SCHMIDT[a,1], Sebastian GEBAUER[a], Annette BARTHOLMES[b], Dennis KADIOGLU[a,b], Jens KLEESIEK[c], Bernd HAMM[d], Thomas J. VOGL[e], Tobias PENZKOFER[d], Andreas Michael BUCHER[e] and Holger STORF[a,b]
[a] *Institute of Medical Informatics, Goethe University Frankfurt, University Hospital Frankfurt, Frankfurt am Main, Germany*
[b] *Data Integration Center, University Hospital Frankfurt, Frankfurt am Main, Germany*
[c] *Institute for AI in Medicine (IKIM), University Hospital Essen, Essen, Germany*
[d] *Department of Radiology, Charité Universitätsmedizin Berlin, Berlin, Germany*
[e] *Department of Diagnostic and Interventional Radiology, University Hospital Frankfurt, Frankfurt am Main, Germany*

Abstract. Within the scope of the two NUM projects CODEX and RACOON we developed a preliminary technical concept for documenting clinical and radiological COVID-19 data in a collaborative approach and its preceding findings of a requirement analysis. At first, we provide an overview of NUM and its two projects CODEX and RACOON including the GECCO data set. Furthermore, we demonstrate the foundation for the increased collaboration of both projects, which was additionally supported by a survey conducted at University Hospital Frankfurt. Based on the survey results mint Lesion™, developed by Mint Medical and used at all project sites within RACOON, was selected as the "Electronic Data Capture" (EDC) system for CODEX. Moreover, to avoid duplicate entry of GECCO data into both EDC systems, an early effort was made to consider a collaborative and efficient technical approach to reduce the workload for the medical documentalists. As a first effort we present a preliminary technical concept representing the current and possible future data workflow of CODEX and RACOON. This concept includes a software component to synchronize GECCO data sets between the two EDC systems using the HL7 FHIR standard. Our first approach of a collaborative use of an EDC system and its medical documentalists could be beneficial in combination with the presented synchronization component for all participating project sites of CODEX and RACOON with regard to an overall reduced documentation workload.

Keywords. COVID-19, CODEX, RACOON, EDC, clinical documentation

1. Introduction

As part of an initiative of the "German Federal Ministry of Education and Research" (BMBF), the German COVID-19 Research "Network of University Medicine" (NUM) was founded. The goal of this network of all German university hospitals is to improve

[1]Corresponding Author, Marvin Schmidt, Institute of Medical Informatics, Goethe University Frankfurt, University Hospital Frankfurt, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany; E-mail: marvin.schmidt@kgu.de.

the general availability of relevant routine and research data as needed for thirteen innovative research projects which should contribute to the management of the pandemic and thus enable the best possible treatment of COVID-19 patients. Accordingly, the network also takes into account the necessary digitalization and finally promotes "Pandemic Preparedness" [1]. In addition to the network, the “German Corona Consensus Data Set” (GECCO) provides another foundation for collaboration in the wake of the COVID-19 pandemic. The GECCO data set describes 83 data items grouped by different categories such as anamnesis/risk factors, imaging and demographics. This consensus data set provides scientists with a research foundation, related to key clinical parameters of COVID-19 [2].

1.1. CODEX and RACOON

The “COVID-19 Data Exchange Platform” (CODEX) is one of the thirteen NUM projects. It is mainly focused on the development of a secure nationwide, interoperable and privacy-compliant research infrastructure for the storage and provision of COVID-19 research datasets [3]. The architecture of CODEX is structured into decentralized nodes (NUM nodes), which are mainly operated by the “Data Integration Centers” (DIZ). These decentralized parts of the platform use existing infrastructures of the “Medical Informatics Initiative” (MII). In addition, there is also a central component (CODEX platform) which supports data provision and use. The platform processes the collected data in the format of the nationwide coordinated GECCO data set [4]. An integral requirement for the storage of the data in the CODEX platform, besides the local pseudonymization, is the provision of the MII Broad Consent, which is ultimately the corresponding consent of the patients for the further processing of their data. The Broad Consent is also the current, coordinated legal basis for the decentralized collection, aggregation and centralized release of health data for research in NUM [5]. If both of these conditions are met, the data is passed on with the cooperation of several components and organizations such as the “GECCO Transfer HUB” (GTH) and the “federated Trusted Third Party” (fTTP). The GTH is responsible for transmitting the medical data (MDAT) distributed by the DIZ to the CODEX Platform. It exchanges the local DIZ pseudonym (DIZ PSN) for the central CODEX pseudonym (CODEX PSN) provided by the fTTP. This ensures a consistent separation of medical data and patient-identifying data (IDAT) [5].

A nationwide infrastructure is also being established within the “Radiological Cooperative Network” (RACOON) for the structured collection of radiological data from COVID-19 cases [6]. For this purpose, all 36 university radiology departments from Germany have joined forces to face the challenges of the COVID-19 pandemic from a radiological perspective [7]. The infrastructure in the RACOON project is based in particular on a nationwide server system with network nodes (RACOON-NODE) at all German university hospitals. The preliminary work, project structure and design were in large parts contributed by the project partner now affiliated with IKIM. The sites in Frankfurt and Berlin are in the lead for project organization and coordination across all partner sites. The RACOON infrastructure represents an integration of the decentralized and centralized (RACOON-CENTRAL) components into a powerful overall concept. RACOON enables a consolidation of the collected diagnostic data within a standardized data model based on the GECCO data set for collaborative research projects. Within RACOON-CENTRAL, both anonymized data sets as well as structured reports of

findings can be generated. The data protection description for RACOON specifies that personal data are only processed locally within the respective clinic network under the sole control of the responsible authority by an approved medical device. Outside the clinic network (especially on RACOON Central), only anonymized data is transmitted and processed [8].

1.2. Collaboration between CODEX and RACOON at UKF

The University Hospital Frankfurt (UKF) is involved in seven of the 13 NUM research projects, three of which are coordinated by Frankfurt in the overall network. Since the start of NUM it has always been a primary goal to foster synergies and increase cooperation between the projects at the UKF [9]. Since both in CODEX and RACOON "Electronic Data Capture" (EDC) systems are used for follow-up documentation of clinical parameters, the question arose early on how to approach this matter as efficiently and collaboratively as possible for both projects. The primary goal of this paper was to provide a first preliminary technical concept for documenting clinical and radiological COVID-19 data for the two NUM projects in a collaborative and efficient approach, to ultimately minimize redundant documentation and therefore provide an additional technical and organizational effort for them. We present the main outcomes of the preceding requirement analysis, the resulting preliminary concept and its findings which could be relevant for other participating sites.

2. Methods

2.1. Software used within CODEX and RACOON

"Extract, Transform, Load" (ETL) processes are used within CODEX to transform the collected data into a standardized format. In addition, medical documentalists use an EDC system for follow-up documentation of missing parameters [10]. For CODEX, the EDC systems "Research Electronic Data Capture" (REDCap) and "Data Integration System" (DIS) were provided by the national project management with the further option to use other systems [11,12]. The documented data in the EDC system are in "Clinical Data Interchange Standards Consortium -Operational Data Model CDISC ODM" format. This format is primarily used for EDC systems. It consists of a structured XML file that supports the platform-independent exchange and archiving of research data [13]. The CDISC ODM data are then transformed using the ODM2FHIR application, which is part of a software node (NUM Node v2) that was implemented by the various CODEX project partners and made available to all sites [14]. During this process, they are transformed into HL7 "Fast Healthcare Interoperability Resources" (FHIR) resources. This standard offers high interoperability as an exchange format [15]. In this node, important adjustments such as transformation into standardized formats and the general pseudonymization of the data are performed [16]. After complete transformation, the research data and their corresponding pseudonyms are then transmitted to the central CODEX platform via the GTH and the fTTP in a privacy-compliant and pseudonymized manner and thus made available to the researchers [10]. Within RACOON, the "mint Lesion™" software from industrial partner "Mint Medical GmbH" (Heidelberg) is used at each project site. mint Lesion™ is a certified medical software for the assessment of radiological images. The software is primarily aimed at radiologists, physicians and

researchers for context-supported reporting of findings using the integrated EDC component and for structured reporting in radiology [7].

2.2. Requirements analysis

In order to promote synergies between the different NUM projects, a requirements analysis with the involved partners at the Frankfurt site was conducted at the beginning of the projects. For this purpose, the scope of the required clinical parameters was relevant, in order to identify the greatest possible overlap between the local NUM projects, to then start a more in-depth collaboration. Accordingly, a survey was conducted in late 2020 together with the DIZ, which aimed to capture the relevant data elements of the GECCO data and the considered cohorts of the corresponding NUM projects at the Frankfurt site. The survey was conducted as a multi-page questionnaire which was send to the participating NUM projects. Additionally, for CODEX a decision had to be made in favor of an EDC system, which would be used to complete the documentation and follow-up documentation of the GECCO data in the further course of the project. During initial discussions regarding the decision about the EDC system it became clear, that in addition to REDCap and DIS the EDC component of mint Lesion™ could also be used for CODEX. This consideration emerged during previous discussions regarding the recruitment of medical documentalists, which are essential in both projects to perform follow-up documentation of clinical and radiological parameters. Moreover, to avoid duplicate entry of GECCO data into both EDC systems, an early effort was made to consider a collaborative and efficient technical approach to reduce the workload of medical documentalists.

3. Results

3.1. Requirements analysis

The evaluated results of this multi-page questionnaire revealed at a broad level, that the data required by all projects were nearly consistent. However, on finer inspection, heterogeneity was nevertheless apparent with regard to the data required as well as the definition of the patient cohorts required in each case. The highest overlap of all projects here was observed for RACOON and CODEX. While CODEX covers the entire GECCO data set, only an intersection of the data set is relevant in RACOON. It is worth mentioning, that RACOON is one of the three NUM projects with the project lead in Frankfurt, together with the Berlin site. Based on these findings, increased cooperation between CODEX and RACOON at the Frankfurt site for the further course of the projects appeared to be highly reasonable. Therefore, a “pool” of medical documentalists, consisting of several employees working for CODEX and RACOON was established.
In addition, after reviewing all available EDC systems presented in CODEX, the decision was made in favor of mint Lesion™. Mint Medical has made the necessary additions to mint Lesion™ for its intended use in CODEX. This included support for a CDISC ODM importer for uploading the various patient data, which was adapted to the CODEX data dictionary. A CDISC ODM exporter was integrated for further processing of the data in the software pipeline.

3.2. Preliminary technical concept

As a first preliminary concept, a draft software architecture (see Figure 1.) was developed, that depicts the currently planned data workflow for CODEX (upper part). Integrated here is a future workflow for RACOON (lower part), which includes a future software component for synchronizing GECCO data sets between the two EDC systems using the HL7 FHIR standard and a HAPI FHIR server.

Within CODEX, the clinical parameters are first extracted from a mirror system of the Frankfurt hospital information system (KIS) ORBIS, an Oracle database, using special SQL queries. Subsequently, these unprocessed raw data are converted from a CSV file into the required CDISC ODM format using a self-implemented ETL process (SQL2ODM). Once the data is in this format, it can be loaded into the mint Lesion™ platform. Within the EDC component of mint Lesion™, medical documentalists can then perform a follow-up documentation of missing GECCO dataset parameters. The fully documented dataset can then be exported via the mint Lesion™ exporter in the CDISC ODM format. Afterwards this exported dataset is transformed into HL7 FHIR format through another ETL process (ODM2FHIR) [14]. Within NUM Node v2, a number of modifications are performed including complete pseudonymization of the data. After the ETL process, the data is available in the HL7 FHIR format which can then be loaded into a deployed HAPI FHIR server. Subsequently, the pseudonymized data is transferred to the CODEX platform via the GTH in a privacy-compliant manner and thus made available to the user. In addition, the GTH exchanges the local DIZ PSN for the central CODEX PSN provided by the fTTP. This ensures consistent separation of MDAT and IDAT [5].

If a patient is registered in both systems, parameters and data already collected in CODEX mint Lesion™ can then be synchronized bidirectionally between RACOON and CODEX. The first approach for this is currently the synchronization of FHIR resources within the HAPI FHIR server. Here, a corresponding FHIR component would have to be integrated into mint Lesion™. After a synchronization of the GECCO data set from CODEX, an exchange of the RACOON relevant data sets can take place. The relevant data can then be loaded into the RACOON mint Lesion™. After the follow-up documentation, the anonymized data can be further processed to be ultimately loaded into RACOON-CENTRAL.

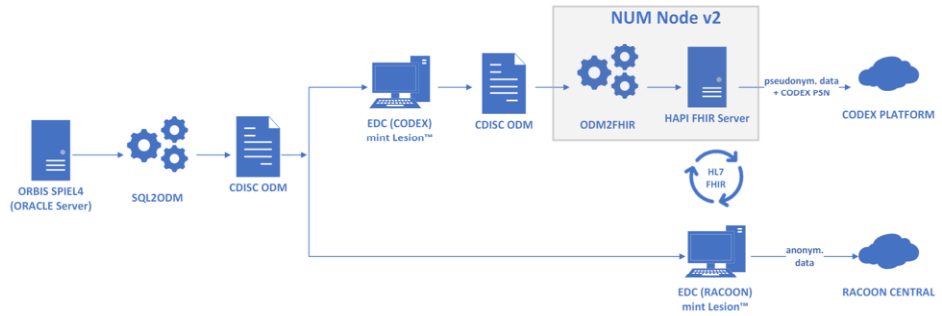

Figure 1. Preliminary technical concept

4. Discussion

One of the main arguments for choosing mint Lesion™ as the EDC system for CODEX was, that in comparison to the other available EDC systems REDCap and DIS, mint Lesion™ offers the advantage that the medical documentalists from the established pool only have to be trained in one system and not in multiple ones, which could ultimately save time and resources. In addition, the documentalists can exchange valuable experience with each other when using the system. Accordingly, synergies were used to efficiently utilize the resources of the medical documentalists for both projects. On the other hand, it must be mentioned here that if REDCap or DIS had been chosen, the general readiness and availability in CODEX would have been faster, since these systems are officially supported and selected in advance by the national project management [10]. mint Lesion™ had to be extended (CDISC ODM importer/exporter) for full use in CODEX in order to meet all project requirements. Another organizational challenge in the preliminary stages of the project was finding suitable and qualified medical documentalists.

Among the technical challenges of the presented concept is, of course, the technical feasibility. Since these projects are nationwide, a large number of stakeholders and locations have to be considered [1]. After a potential implementation, this technical concept could be a useful extension between the two projects in the future. Above all, it should be noted that it has both organizational and technical benefits. The transfer of unprocessed raw data from the KIS is therefore also handled in the same manner for both systems. This could prevent redundant data documentation, which could ultimately reduce the workload and also prevent potential errors while documenting. It is intended that the presented concept prospectively will be implemented by industrial partner Mint Medical. Initial consultations on such a synchronization have already been held with technical managers at Mint Medical. It would thus be an extension of the close collaboration that already existed prior to this paper. The mentioned first possible solution using the HL7 FHIR standard and the HAPI FHIR Server is only one of the suggestions how it could be solved technically.

Despite all the advantages mentioned, the potential difficulties and barriers of such an implementation should not be disregarded because in this context data privacy is crucial. Since these are two projects with different patient cohorts and data scopes, it is necessary that the adopted data privacy concepts in both projects must also be thoroughly reviewed for possible compliance. Prior to such a consolidation, these would naturally have to be reapproved by the corresponding data protection authority because both projects involve sensitive patient data.

The potential developers of such a synchronization component would therefore have to be in close contact with the project managers of both projects in order to coordinate the entire project adequately, which could be challenging due to the size of both projects. Overall, it became apparent during the initial implementation phase that the above-mentioned organizational and technical barriers ultimately prevented the concept from being implemented. The actual benefit of the technical concept presented can only be fully assessed in the future after its first prototypical implementation. In the further course, a more in-depth discussion on the feasibility of such a collaborative data documentation with regard to the data protection admissibility of both projects is inevitable. However, the benefits of the concept could outweigh the expressed concerns, so it is recommended to continue working actively on its implementation. This concept

is also expandable, so that other NUM projects can be integrated into the infrastructure and synergies can be used more intensively.

5. Conclusion

All in all, the results of the requirement analysis including the resulting preliminary technical concept presented here show large potential for improved collaboration between CODEX and RACOON in the future. The further findings, such as the adoption of Mint Medical's mint Lesion™ as the EDC system for CODEX and, the creation of a medical documentalists pool for both projects, show that beneficial collaboration is possible despite divergent project aims. The presented findings could therefore be helpful for other participating project sites. Within the follow-up project of CODEX called "CODEX+" are important efforts for deeper cooperation between all NUM projects, such as "Work Package 3.3 Integration with other NUM projects". In Conclusion, the results of this paper can serve as a first good approach to avoid duplication of effort and data entry to ultimately bring both projects closer together. The presented concept of such a data synchronization between the two projects could have great potential due to the reduced documentation workload for the medical documentalists. Furthermore, this solution could be rolled out to other participating sites of CODEX and RACOON to adequately support both projects.

Acknowledgements

CODEX and RACOON are funded in context of NUM by the German Federal Ministry of Education and Research (BMBF). Funding reference number: 01KX2021 (BMBF)

We thank both teams of CODEX and RACOON as well as Mint Medical GmbH (Heidelberg), especially Johannes Kast, Thomas Massier and Dominik Golz for their past and ongoing effort in regards to the support and further development of mint Lesion™.

Declarations

Conflict of Interest: The authors declare, that there is no conflict of interest.

Contributions of the authors: MS and SG contributed to the development of the presented preliminary technical concept. MS, SG, AB, DK, JK, TP, AMB and HS contributed to the overall workflow and the project organization. All authors contributed to drafting and revising this manuscript.

References

[1] Netzwerk Universitätsmedizin. Netzwerk Universitätsmedizin, Home | Netzwerk Universitätsmedizin [Internet]. [cited 2022 Feb 21]. Available from: https://www.netzwerk-universitaetsmedizin.de/

[2] Sass J, Bartschke A, Lehne M, Essenwanger A, Rinaldi E, Rudolph S, 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. DOI: 10.1186/s12911-020-01374-w.

[3] Netzwerk Universitätsmedizin. CODEX | Netzwerk Universitätsmedizin [Internet]. [cited 2022 Feb 18]. Available from: https://www.netzwerk-universitaetsmedizin.de/projekte/codex

[4] Ingenerf J. Netzwerk Universitätsmedizin Umsetzungskonzept für den Aufbau einer COVID-19 Data Exchange Plattform (CODEX) [Internet]. [cited 2022 Feb 25]. Available from: https://www.imi.uni-luebeck.de/forschung/p51-netzwerk-uni-medizin.html

[5] Bahls T, Hampf C, Bialke M, Hoffmann W, Universitätsmedizin Greifswald. Lösungsbaustein fTTP (federated Trusted Third Party) als ein Enabler für vernetzte medizinische Forschung mit dezentraler Datenhaltung [Internet]. [cited 2022 May 11]. Available from: https://www.ths-greifswald.de/wp-content/uploads/2021/11/2021-09-10-fTTP-fact-sheet.pdf

[6] Netzwerk Universitätsmedizin. RACOON | Netzwerk Universitätsmedizin [Internet]. [cited 2022 Feb 14]. Available from: https://www.netzwerk-universitaetsmedizin.de/projekte/racoon

[7] Salg GA, Ganten MK, Bucher AM, Kenngott HG, Fink MA, Seibold C, et al. A reporting and analysis framework for structured evaluation of COVID-19 clinical and imaging data. npj Digit Med. 2021 Apr 12;4(1):1-9. DOI: 10.1038/s41746-021-00439-y.

[8] RACOON Network. RACOON Base – RACOON Infrastruktur [Internet]. [cited 2022 Feb 25]. Available from: https://racoon.network/?page_id=1638

[9] Universitätsklinikum Frankfurt. Netzwerk Universitätsmedizin (NUM): Universitätsklinikum Frankfurt am Main [Internet]. [cited 2022 Feb 22]. Available from: https://www.kgu.de/ueber-uns/vorstand-des-universitaetsklinikums/aerztliche-direktion/stabsstelle-medizinische-informationssysteme-und-digitalisierung/netzwerk-universitaetsmedizin-num

[10] Krefting D. Das Projekt CODEX des Netzwerks der Universitätsmedizin (NUM) [Internet]. 2021 [cited 2022 Feb 25]. Available from: https://www.miracum.org/wp-content/uploads/2021/07/2021_07_16_0930_Krefting_Das-Projekt-CODEX-des-Netzwerks-der-Universitaetsmedizin-NUM.pdf

[11] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics. 2009 Apr 1;42(2):377-381. DOI: 10.1016/j.jbi.2008.08.010.

[12] Lautenschläger R, Kohlmayer F, Prasser F, Kuhn KA. A generic solution for web-based management of pseudonymized data. BMC Medical Informatics and Decision Making. 2015 Nov 30;15(1):100. DOI: 10.1186/s12911-015-0222-y.

[13] Clinical Data Interchange Standards Consortium. ODM-XML | CDISC [Internet]. [cited 2022 Feb 25]. Available from: https://www.cdisc.org/standards/data-exchange/odm

[14] Erbelding C, Sailer B, Stenzhorn H, Biergans S, Kohlmayer F, Kobak EM, et al. GECCO on FHIR – Towards Interoperable Data on COVID-19. In German Medical Science GMS Publishing House; 2021. p. DocAbstr. 38. DOI: 10.3205/21gmds012.

[15] HL7 International. HL7 FHIR [Internet]. [cited 2022 Feb 27]. Available from: https://www.hl7.org/fhir/

[16] Kapsner LA, Kampf MO, Seuchter SA, Gruendner J, Gulden C, Mate S, et al. Reduced Rate of Inpatient Hospital Admissions in 18 German University Hospitals During the COVID-19 Lockdown. Frontiers in Public Health. 2021;8. DOI: 10.3389/fpubh.2020.594117.