Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220434

Linking EMR Data to REDCap: Implementation in the SOLKID Register

Tobias J. BRIX[a,1,2], Leonard GREULICH[a,1], Alice JANSSEN[a], Sarah RIEPENHAUSEN[a], Philipp NEUHAUS[a], Johannes OEHM[a], Jeannine WEGNER[b], Barbara SUWELACK[b], Michael STORCK[a,1] and Julian VARGHESE[a,1]
[a]*Institute of Medical Informatics, University of Münster, Germany*
[b]*Department of Medicine D, Transplantnephrology, University of Münster, Germany*

Abstract. Secondary use, the reuse of medical patient data stored during routine care in the hospital's electronic medical records (EMR) for research purpose is common, especially for registers and pragmatic trials. Often the medical data items are copied manually from the EMR into the used research database. This process is time consuming and error prone. In the "Safety of the Living Kidney Donor – The German National Register" (SOLKID-GNR), laboratory results gathered during control check-ups of the living donors before and after the transplantation are to be transferred from the EMR into the electronic data capture system REDCap of the register. In this work, we present our approach of realizing an automated transfer of time-dependent laboratory results from the EMR of the University Hospital of Münster to REDCap. A challenge lies in the multi-center structure of SOLKID-GNR. The participating transplant centers are using different EMR systems, which requires a flexible architecture design. In addition, we aimed to support reuse of the implementation for other research settings with other medical data items of interest.

Keywords. Secondary use, SOLKID-GNR, REDCap

1. Introduction

Medical patient data collected during routine care in hospitals is stored in electronic medical records (EMR). In contrast, medical patient data for scientific research purposes, e.g., clinical trials, is collected in specific electronic data capture systems (EDC). Many data items can be directly reused from routine care. This is a common approach known as secondary use and is often applied for registers and pragmatic trials. Thus, for instance, laboratory observations do not have to be performed twice, which saves costs and labor. However, a transcription from the EMR to an EDC is often time consuming and error prone, since manual copying of information always implies the risk of human mistakes. Therefore, an automated transfer is preferable.

The German living donor register "Safety of the Living Kidney Donor – The German National Register" (SOLKID-GNR) collects data of the medical and psychosocial outcome of living kidney donors [1]. As part of the register, donors are

[1] These Authors contributed equally.

[2] Corresponding Author, Tobias J. Brix, Institute of Medical Informatics, University of Münster, Albert-Schweitzer-Campus 1, Building A11, 48149 Münster, Germany; E-mail: tobias.brix@uni-muenster.de.

monitored before and after the transplantation in multiple check-ups. During these check-ups, laboratory observations are documented, e.g. creatinine. Since living donors must be monitored very strictly in the time after transplantation anyway, many data items, especially laboratory results, are already documented during routine care. Previously, this given data was transferred manually by study nurses for secondary use. SOLKID-GNR is a multi-center register consisting of multiple hospitals. While the destination of an automated data transfer is always the same EDC system, the source of the data transfer varies. Each hospital is using its own EMR, and even if the same EMR is used, these systems can be configured individually and the same data can be stored in different locations and formats.

Objective of this work is the implementation of a prototype, which allows an automated transfer of medical patient data from the EMRs of different healthcare centers to the EDC of SOLKID-GNR. The current prototype is focused on the EMR of the University Hospital of Münster (UKM) and the current use case is the SOLKID-GNR register. The implementation is aimed to be as generic as possible by utilizing technical and semantical standards. This allows reuse in future research projects and different EMR installations.

2. Methods

2.1. SOLKID-GNR Setting and Challenges

The German living donor register SOLKID-GNR consists of 38 German transplant centers. A single "Research Electronic Data Capture" (REDCap) instance is used as central EDC system storing the data of all 38 centers [2]. Data access groups (DAG) of REDCap are used to ensure that a center has only access to its own patients. The pseudonymization of the data records is carried out by a central pseudonymization service "Mainzelliste" [3], which is seamlessly connected to REDCap [4].

The first challenge is the diversity of EMRs used by each transplant center. This requires an individual customizable middleware to connect to the specific EMRs. If a connection to the EMR is established, the second challenge is the identification of the correct patient. REDCap only stores the pseudonym of a patient while the EMR identifies patients by their identifying data or special IDs. Third, the correct data within the EMR must be identified. Finally yet importantly, a time component must be considered. SOLKID-GNR has longitudinal follow-ups and the used data must be associated to the given time period.

2.2. Implementation Details

While REDCap can be used without cost under the REDCap license, it is no open-source software and does not allow complete customization. However, REDCap supports a powerful extension and plugin mechanism based on modules. These modules can be implemented in PHP for the back-end and HTML5, CSS and JavaScript in the front-end. In this paper, we have implemented a specific REDCap module for our use case.

As middleware, i.e., connection service between REDCap and the EMR, we use a RESTful web-service implemented in the Java Spring Framework. This middleware was already in use, but had to be extended to support the current workflow [5]. The current implementation is limited to the UKM, which is using Dedalus ORBIS as EMR.

Specifying the exact semantic meaning of a medical data item is a common problem in medical informatics. The current state of the art is to use codes of classifications and terminologies to be independent of language, spelling and formulation. In our use case of the transfer of laboratory results, the "Logical Observation Identifiers Names and Codes" (LOINC) is most relevant, a terminology for laboratory and clinical observations.

3. Results

To use the implemented REDCap module, it must be installed on the REDCap server and be activated for the project. Afterwards, the module must be configured by selecting data items on "electronic case report forms" (eCRF) that should be queried. On selected forms, a new button is rendered in the normal data entry mode, which triggers the data transfer. Since the currently opened form is associated with a patient record, REDCap knows the record's pseudonym and can send it to the middleware. Additionally, an individual mapping between LOINC codes and each selected data item of the eCRF must be configured. Thus, during data request, REDCap knows the associated codes of the opened form and can send the list to the middleware. To achieve the time dependency, an interval can be specified for each form. Since values like, one week before or two weeks after the follow up cannot be specified in advance, this configuration can be done relatively to a reference date element. In the case of SOLKID-GNR, an item in the eCRF contains the date of follow up and the interval sent to the middleware is calculated relatively to this reference date.

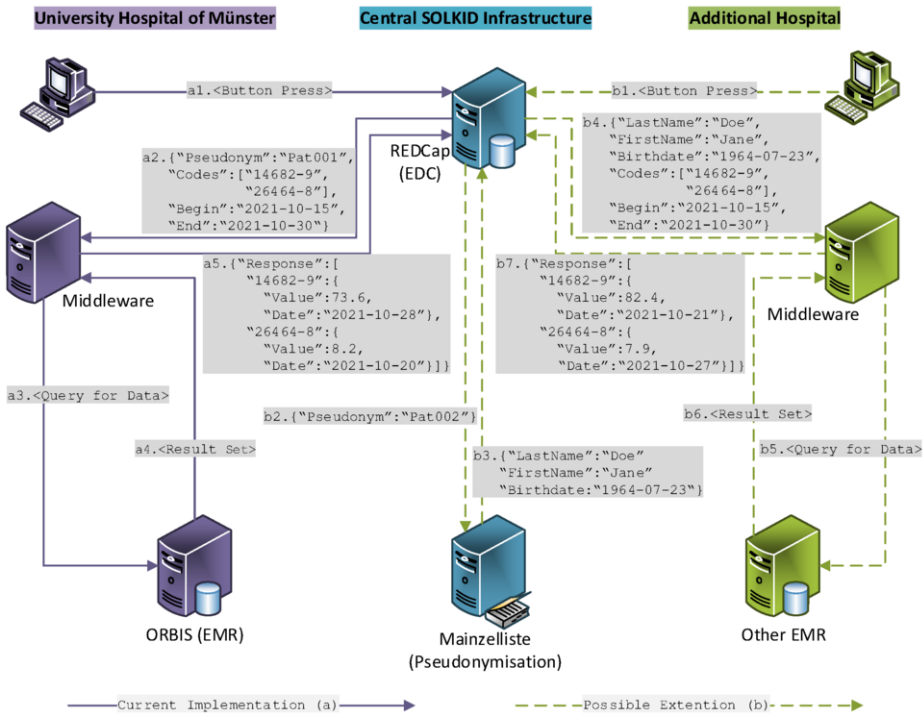

Figure 1. Architecture and workflow of the secondary use in SOLKID-GNR.

The currently implemented architecture and workflow can be seen on the left side of Figure 1. If a user triggers the data transfer in REDCap (1.), the EDC communicates with the middleware's REST-API via an authenticated HTTPS request and submits a patient's pseudonym, the data elements of interest coded as a list of LOINC codes and a time interval that constrains the laboratory observation's documentation date in the EMR (2.).

Once the request is received by the middleware, it checks whether the given pseudonym is registered in ORBIS for SOLKID-GNR. In Münster, an EMR-integrated system is in use, which allows the management of research projects and the registration of their participants [6]. The participant's pseudonym is part of the registered properties. If a patient is not yet registered, a HTTP-error code is returned and a meaningful error message is displayed in REDCap. If the pseudonym and the associated patient have been identified successfully, the data export from the EMR begins (3.). Based on the submitted time interval, the associated medical data is extracted. If multiple laboratory results are found in the given interval, the newest observation is returned (4.). At the moment, the mapping from LOINC codes to laboratory observation IDs in ORBIS are based on a manual mapping file provided by the local medical data integration center. A map, containing the LOINC codes as keys and the found laboratory results and associated observation dates, is sent back to REDCap (5.).

During data retrieval, the previous module configuration is used to populate the results into the correct eCRF data item in REDCap. Although optional, we have extended the eCRF by a laboratory date item for each laboratory result. This is needed since laboratory observations may have been determined on different dates. To ensure reproducibility and traceability, these exact dates have to be documented. As optimization, the most common date is calculated by the module and stored as primary date, and only deviating dates are documented or hidden otherwise. This keeps the eCRF clear if all values are from a single observation, but also provides the flexibility of multiple observations.

4. Discussion

The presented approach is working successfully for the site Münster of the SOLKID-GNR. However, it has its limitations regarding the transition to other participating hospitals and to other research projects. While the REDCap module can be configured individually for each project and can be adapted for each new eCRF, the middleware, i.e., the export process itself, is highly EMR and medical concept dependent. Each change requires an individual adaptation. Even hospitals also using ORBIS may not be able to use the implementation without adaptation, since ORBIS internal structures are customizable. It is worth mentioning that a communication server may replace the middleware at some hospitals. We used the existing middleware since it was easier to configure and adapt.

Currently, we only applied LOINC in our approach. If more data besides laboratory or clinical observations should be identified, possible options would be "International Statistical Classification of Diseases and Related Health Problems" (ICD) for diagnoses, "Systematized Nomenclature of Medicine – Clinical Terms" (SNOMED-CT) as the reference terminology for medical concepts, or the "Unified Medical Language System" (UMLS), a metathesaurus combining all the previously mentioned ones. All these systems are using fixed codes, i.e., strings, to encode medical concepts. The APIs of both

REDCap and middleware can handle any string-based codes, and therefore any code system best suited for the given application case can be used. Only the mapping from codes to EMR elements must be adapted in the middleware. This mapping is EMR and data item dependent.

Another limitation is the identification of a patient in the EMR only knowing the pseudonym of the register. The beneficial situation of having the pseudonyms present in the EMR is often not given in other hospitals. A possible solution is illustrated on the right side of Figure 1. A patient in the EMR can only be identified by his or her identifying data like first name, last name and birthdate. The Mainzelliste allows a re-identification based on the pseudonym. Thus, the identifying data can be automatically fetched by depseudonymization and sent to the middleware to find and identify the associated patient. This approach rises additional data protection and ethical concerns, which must be addressed. Furthermore, the general issue about how to uniquely identify a person has to be discussed, since these three attributes do not have to be distinctive. The Mainzelliste also allows custom attributes. Storing the EMR-ID or social security number during the pseudonymization process could be a solution.

5. Conclusions

In this work, we presented our implementation of the SOLKID-GNR register at the site of Münster to support the automated transfer of time dependent laboratory results from the EMR to the research database REDCap. While the approach is feasible for a single site, its reuse for other sites and research projects requires additional work to establish site-specific interfaces for querying the respective EMRs. The challenges and possible solutions for this adaptation have been discussed.

6. Acknowledgements

Supported by BMBF grant Ns. 01ZZ1802V (HiGHmed/Münster) and 01GY1906 (SOLKID-GNR/Münster). We would like to thank the entire SOLKID team.

References

[1] Suwelack B, Dugas M, Koch M, Sommerer C, Urban M, Gerß J et al. Die Sicherheit des Lebendnierenspenders – Das Deutsche Lebendspende Register SOLKID-GNR – Entstehung und Struktur eines nationalen Registers in der Versorgungsforschung. Gesundheitswesen (Bundesverband der Arzte des Offentlichen Gesundheitsdienstes (Germany)) 2021;83(S 01):S33-S38.

[2] 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;42(2):377–81.

[3] Lablans M, Borg A, Ückert F. A RESTful interface to pseudonymization services in modern web applications. BMC medical informatics and decision making 2015;15:2.

[4] Greulich L, Brix TJ, Storck M, Dugas M. A Seamless Pseudonymization and Randomization Workflow for REDCap. Studies in health technology and informatics 2021;281:952–6.

[5] Bruland P, Doods J, Brix T, Dugas M, Storck M. Connecting healthcare and clinical research: Workflow optimizations through seamless integration of EHR, pseudonymization services and EDC systems. International journal of medical informatics 2018;119:103–8.

[6] Brix T, Gessner S, Bruland P, Dugas M. CIS-integrated study patient management: registration, pseudonymization and randomization: German Medical Science GMS Publishing House; 2018.