A Seamless Pseudonymization and Randomization Workflow for REDCap

Leonard GREULICH [a,1], Tobias J. BRIX [a,1,2], Michael STORCK [a] and Martin DUGAS [a]

[a] *Institute of Medical Informatics, University of Münster, Germany*

Abstract. The interaction of multiple computer systems during multi-center randomized controlled trials (RCTs) is a hurdle for IT-specialists as well as medical staff. A common workflow for the initial registration of a patient requires the generation of a pseudonym by a pseudonymization service, a manual transmission of the pseudonym to a randomization service, and a manual transfer of the pseudonym and assigned study arm into an electronic data capture (EDC) system. This interaction is often time consuming and error prone due to multiple system changes. Objective of this work is to enhance a commonly used EDC system, Research Electronic Data Capture (REDCap), as a single source of interaction for multi-center RCTs. This is achieved by providing two modules for a seamless integration of a pseudonymization service, i.e., Mainzelliste, and a randomization service, i.e., RandIMI. Thus, no site-specific system changes are required, which increases time efficiency and reduces errors. From a technical perspective, only authentication credentials and firewall exposure for a single system must be managed. To evaluate the usability of our implementation, the system usability scale was employed. The increase of time efficiency was measured in laboratory conditions by a comparison of the time for patient registrations with and without our modules. An "excellent" usability was shown and an average time reduction by nearly 64 %. Both open-source modules are available from the REDCap Repository of External Modules.

Keywords. REDCap, Pseudonymization, Randomization, Open-Source

1. Introduction

Carrying out multi-center randomized controlled trials (RCTs) is an organizational challenge. Especially, the interaction of multiple computer systems, which includes the management of authentication credentials for each system and individual firewall settings are a hurdle for IT-specialists as well as medical staff. Main systems of a clinical trial are an electronic data capture (EDC) system, a pseudonymization service, and a randomization service. The collection of medical data should be done in EDC systems, which are specifically tailored for traceability and reproducibility. As the European Medicines Agency (EMA) and European General Data Protection Regulation (GDPR) stated, data collection should be performed in a pseudonymized way, i.e., replacing the identifying data of a subject with an arbitrary string [1,2]. While most EDCs support the automated generation of pseudonyms, for multi-center studies a centralized

---

[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.

pseudonymization service is required to ensure that identical patients at different sites always are assigned the identical pseudonyms. Finally, a randomization service is required for RCTs, which uses specific randomization algorithms to assign participants to different study arms by chance, e.g., control or intervention. Often these services require the patient's pseudonym for traceability and reproducibility. Thus, a common workflow for the registration of a patient requires the generation of a pseudonym in the pseudonymization service, a manual transmission of the pseudonym to the randomization service for randomization, and a manual transfer of the pseudonym and assigned study arm into the EDC system. This interaction requires multiple system changes and is thus often time consuming and error prone, since manual copying of information always implies the risk of human mistakes.

Objective of this work is to enhance a commonly used EDC system as a single source of interaction for multi-center RCTs. This is achieved by a seamless integration of a pseudonymization and randomization service. Thus, no site-specific system changes are required, which increases efficiency and reduces errors. From a technical point of view, only authentication credentials and firewall exposure for a single system must be managed. To evaluate the usability of our implementation, the system usability scale (SUS) was applied [3]. Furthermore, the increase of time efficiency was measured by comparison of the time that is required to register a patient with and without the newly developed system.

2. Methods

2.1. REDCap and its Modules

Research Electronic Data Capture (REDCap) is one of the most popular EDC systems [4]. Although it is no open-source software, it is still free to use under its REDCap license. Like all EDC systems, REDCap allows capturing trial data in electronic case report forms (eCRF), which can also be assigned to different study arms. During the registration process of a new patient, a pseudonym is required and, if applicable, a manual assignment to a study arm. By default, REDCap supports the automated generation of simple pseudonyms, but no record linkage functionality as required for multi-center or longitudinal studies. An automated randomization is not supported.

Although REDCap is no open-source software, it includes the External Module Framework to enhance its existing functionality. All published modules are available from the REDCap Repository of External Modules [5] and can easily be installed. In this work, we utilize the External Module Framework to achieve the seamless integration.

2.2. Used Pseudonymization and Randomization Service

As pseudonymization service, the so-called "Mainzelliste", developed at the University of Mainz, has been used [6]. It supports record linkage and provides a REST API for the pseudonymization [7,8]. In Germany, it is recommended by the Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. (TMF) and is used, for instance, in a 43-center register study for multiple sclerosis and by the German market leader for bio-banking software [9,10]. A module for the integration of the Mainzelliste

into REDCap has already been published and is publicly available [11]. This module uses the Mainzelliste's REST API for communication.

As randomization service, "RandIMI", a development of the Institute of Medical Informatics, has been used. RandIMI supports different randomization algorithms and stratifications, i.e., different randomizations based on patient attributes like gender or age. Stratums can be defined as numeric intervals or enumerations. For traceability, RandIMI requires credentials, the patient's pseudonym and the current study center. Like the Mainzelliste, it also provides a REST API for automated integrations. At the time of writing, RandIMI has not yet been published, but an open-source release is planned.

2.3. Evaluation Setup

The usability and time efficiency improvements of our work have been evaluated in laboratory conditions. To evaluate the usability, the system usability scale (SUS) has been applied. The time efficiency was evaluated by comparing the required time needed for a registration, if the pseudonymization and randomization are performed in separate systems (a) in comparison to our seamless integrated version in REDCap (b). Participants were asked to pseudonymize and randomize five patients for each workflow, while the time in seconds for each registration was measured. For each participant, the execution order of workflow (a) and (b) was randomized to mitigate potential learning biases. Before each workflow, a training took place, where the participants were able to practice the entire process with a test patient. The workflow (a) was the following: (1) Login to the Mainzelliste's web-service and enter first name, last name and date of birth to request a pseudonym, (2) login to RandIMI, copy the pseudonym and apply the randomization, (3) login to REDCap, copy the pseudonym and study arm and apply the registration. For workflow (b), the procedure was: (1) Login to REDCap and enter first name, last name and date of birth in the pseudonymization module and apply the registration including randomization.

3. Results

3.1. Randomization Module and its Configuration

The integration of the randomization service RandIMI has been implemented as REDCap module [12] and is available from REDCap's Repository of External Modules. As configuration, the URL of the RandIMI server, the associated Study ID in RandIMI, and the authentication credentials of RandIMI must be defined. Afterwards, the module can be used in REDCap. To perform a randomization, the patient's pseudonym and the current study center must be entered. It is worth noting, that the study center can be selected from a drop-down menu but is restricted to the sites a user has access to in REDCap. If an error occurs during the randomization, a localized error message is presented to the user. Otherwise, the new patient is registered in REDCap with the study arm and pseudonym. It is important to note that the study arms in RandIMI and REDCap must be named equally to enable an automated mapping.

The configuration of stratums is not needed in the module. During the initialization of the module, all required stratum parameters are requested from RandIMI and the input form is dynamically rendered. Thus, stratums like age can be entered in numeric inputs, while enumerated stratums can be selected via drop-down menus.

The seamless integration with the pseudonymization module has been achieved by extending its configuration. There is now an additional option to include the randomization module. If chosen, the pseudonymization module does not automatically create a new patient record after a successful pseudonymization, but transfers the generated pseudonym to the randomization module. Thus, the second module is opened automatically and only the study center and potential stratum parameters must be entered. The entire workflow is depicted in Figure 1.
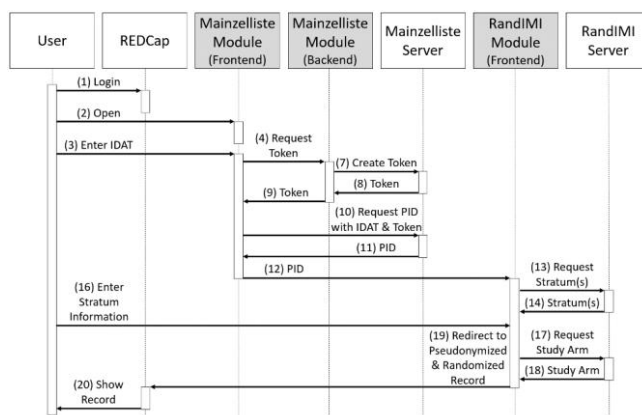

Figure 1. Workflow of the patient registration in REDCap.

3.2. Evaluation Results

The evaluation was conducted with 12 participants (6 female, 6 male). The average age of all participants was 29.2 ± 3.4. They all had a background in medical informatics and were familiar with the procedure of patient registration and EDC systems. Regarding the time efficiency, an average time of 74.5 ± 10.8 seconds was conducted for workflow (a) and an average time of 26.5 ± 2.8 seconds for workflow (b). This results in an average time reduction of 63.9 ± 4.8 %. The SUS showed an average value of 94.4 ± 6.6 which equals an "excellent" usability [13].

4. Discussion

We developed and published modules for the popular EDC system REDCap that seamlessly integrate the process of pseudonymization and randomization. Both modules are freely available in REDCap's Repository of External Modules under the MIT open-source license. The evaluation has shown an increase in time efficiency by nearly 64 %. The increase is mainly caused by the used single sign-on mechanism, which avoids switching systems and re-authentication. Moreover, a generated pseudonym is automatically passed from the pseudonymization module to the randomization module, which then automatically redirects the user to the respective patient record.

An "excellent" usability has been stated. However, these results must be seen with caution. The evaluation was performed in laboratory conditions with participants being computer affine. Thus, long-term results in real setups have to be evaluated. At the time of writing, both modules are used in real clinical trials and first feedback is gathered.

This work has limitations. While Mainzelliste is openly available and widely established in Germany, RandIMI is currently being internally developed. However, RandIMI will be documented and provided as open-source project in the near future. A current version can be obtained on request by the authors.

5. Conclusion

In this work, we have presented the combined usage of two REDCap modules to allow a seamless pseudonymization and randomization directly within REDCap. Thus, only a single system is needed as point of interaction. An evaluation under laboratory conditions have shown an "excellent" usability of this approach and a time efficiency improvement of nearly 64 %. Both modules are freely available from the REDCap repository.

Acknowledgements

References

[1]   European Medicines Agency. ICH E6 (R2) Good clinical practice, Available at https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice.
[2]   The European Parliament and the Council of the European Union. Regulation (EU) 2016/679. General Data Protection Regulation (GDPR), Available at https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN.
[3]   Brooke J. SUS: a quick and dirtyusability scale. Usability evaluation in industry 1996:189.
[4]   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.
[5]   REDCap - Vanderbilt University, Available at https://redcap.vanderbilt.edu/consortium/modules/index.php.
[6]   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.
[7]   Warnecke T, Borg A, Ückert F, Lablans M (eds.). Fehlertolerantes Record Linkage von Patientendaten durch den Phonet-Algorithmus. German Medical Science GMS Publishing House; 2013.
[8]   Lablans M, Borg A, Tremper G. Schnittstelle der Mainzelliste, Available at https://www.unimedizin-mainz.de/fileadmin/kliniken/imbei/Dokumente/MI/Mainzelliste/Dokumente/Mainzelliste__Schnittstelle.pdf.
[9]   Simbrich A, Thibaut J, Khil L, Maximov S, Wiendl H, Berger K. Chances and Challenges of Registry-Based Pharmacovigilance in Multiple Sclerosis: Lessons Learnt from the Implementation of the Multicenter REGIMS Registry. Drug safety 2020.
[10]   KAIROS GmbH. CentraXX, Available at https://www.kairos.de/produkte/centraxx/.
[11]   REDCap Mainzelliste Integration, Available at https://imigitlab.uni-muenster.de/published/redcapmainzellisteintegration.
[12]   REDCap RandIMI Integration, Available at https://imigitlab.uni-muenster.de/published/redcaprandimiintegration
[13]   Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008;24(6):574–94.