

Automatic Data Transfer from OMOP-CDM to REDCap: A Semantically-Enriched Framework

Emanuele GIRANI^a, Matteo GABETTA^{b,1}, Anna ALLONI^b, Morena STUPPIA^b, Lucia SACCHI^a and Nicola BARBARINI^b

^a*Department of Electrical, Computer and Biomedical Engineering,
University of Pavia, Pavia, Italy*

^b*BIOMERIS (BIOMedical Research Informatics Solutions), Pavia, Italy*

Keywords. OMOP, OHDSI, REDCap, eCRF, EDC, Real World Data

1. Introduction and Background

Clinical research is currently limited by the need to manually enter data related to clinical trials through EDC (Electronic Data Capture) or eCRF (electronic Case Report Form). This implies replication of data between Electronic Health Record systems (EHR) and eCRF, the employment of a considerable amount of time and resources and easily leads to errors. Thus, the systematic reuse of Real World Data (mainly EHR data) to automatically fill in eCRF may represent a turning point in clinical trials [1].

OMOP/OHDSI (Observational Medical Outcomes Partnership - Observational Health Data Sciences and Informatics) is an ideal middleware to be interposed between EHR and eCRF, in order to decouple the complexity of the clinical sources from the target eCRF [2]. Its use of both a standardized data model and the main standardized terminologies, makes it a particularly suitable candidate.

REDCap (Research Electronic Data CAPture) is a widely adopted web-based eCRF system for non-profit studies [3]. There are different solutions to automatically import data into a REDCap study (direct ETL through REDCap API, Dynamic Data Pull (DDP), Clinical Data Pull (CDP)), but to the best of our knowledge no specific project is focusing on automatic data transfer from OMOP Common Data Model (CDM) to REDCap.

The aim of this work is proposing a semantically-enriched framework to support automatic data transfer from OMOP CDM to REDCap eCRF.

2. Method

The main components of the system are: (i) a semantic annotation framework to extend a REDCap study's metadata and, (ii) a software component which actually operates the data transfer accordingly to the semantic annotations. The semantic annotation is added

¹ Corresponding Author, BIOMERIS, Via Ferrata 1, 27100 Pavia, Italy; E-mail: matteo.gabetta@biomeris.it.

to the REDCap Data Dictionary, a specific metadata that annotates the study's elements (i.e., variables); these annotations are written in JSON format with different sections. The first section identifies, among the different terminologies included in the OMOP vocabularies, which concept better describes the element itself (e.g., SNOMED code for "Gender") and in some cases the possible values (e.g., SNOMED concepts for "Female"). Other sections of the semantic annotation cover aspects related to defining how the data transfer process shall occur considering several aspects: (i) the reference date, in the REDCap study, associated to a specific time-dependent element (e.g., a laboratory value), to accurately fetch the OMOP database; (ii) a time tolerance to relax the OMOP queries (e.g. reference date \pm 1 day); (iii) how dependent elements (e.g. because of a branching logic rule) have to be treated by the system; (iv) the actual order in which the data transfer attempts for a single or groups of variables have to be executed.

Besides the semantic annotation of the study, we developed a tool in Java language which: (i) reads the semantic annotation through the REDCap standard API, (ii) has access to an external lookup table to match REDCap record IDs with OMOP IDs, (iii) allows to execute the data transfer according to different policies/schedules, for example, on a periodic base (e.g., once a day, every night), whenever a new patient is created in REDCap (leveraging the trigger functionality of REDCap) or on-demand (i.e., the user of the tool decides when to fetch data).

3. Preliminary Results and Discussion

A first prototype of the framework has been developed in Java to test the proposed approach. We decided to test the framework on the Italian Registry for Severe Asthma Patients (SANI). SANI is an ideal test case for the system because it is managed in REDCap and an updated OMOP version of its data is maintained by the ERS SHARP initiative, where a particular subset of 198 OMOP concepts has been chosen as the CDM for several European registries for severe asthma in order to build a federated network.

In particular we are testing the framework on 148 REDCap variables representing different input format (numbers, dropdown, etc.) and data types (demographics, anamnesis, tests and therapies). Some preliminary results show a promising ability of the framework to automatic fill in SANI REDCap eCRF from the current SHARP-SANI OMOP database. These results foreshadow a great saving in terms of time and resources for data collection activity in REDCap eCRF starting from a OMOP CDM.

References

- [1] Griffon N, Pereira H, Djadi-Prat J, García MT, Testoni S, Cariou M, Hilbey J, N'Dja A, Navarro G, Gentili N, Nanni O, Raineri M, Chatellier G, Gómez De La Camara A, Lewi M, Sundgren M, Daniel C, Garvey A, Todorovic M, Ammour N. Performances of a Solution to Semi-Automatically Fill eCRF with Data from the Electronic Health Record: Protocol for a Prospective Individual Participant Data Meta-Analysis. *Stud Health Technol Inform.* 2020 Jun 16;270:367-371.
- [2] Hripesak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, Suchard MA, Park RW, Wong IC, Rijnbeek PR, van der Lei J, Pratt N, Norén GN, Li YC, Stang PE, Madigan D, Ryan PB. Observational Health Data Sciences and Informatics (OHDSI): Opportunities for Observational Researchers. *Stud Health Technol Inform.* 2015;216:574-8.
- [3] 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; 42(2):377-381.