From EHR to EDC - The Experience at the Policlinico Hospital in Milan

Sara PIZZIMENTI [a,1] , Mauro BUCALO [b], Amedeo GUZZARDELLA [a],
Eleonora FERRETTI [a], Angelo CAROLI [a], Alberto ZANELLA [a],
Giacomo GRASSELLI [a], Nicola BARBARINI [b] and Silvano BOSARI [a]

[a] *Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy*
[b] *BIOMERIS (BIOMEdical Research Informatics Solutions), Pavia, Italy*

Keywords. data reuse, secondary use, i2b2, REDCap, eCRF, EDC

1. Introduction

As in many hospitals, especially in Italy, the Electronic Health Record (EHR) of IRCCS Ca' Granda Ospedale Maggiore Policlinico hospital in Milan is composed by a variety of proprietary software applications used to support clinical practice. For this reason, clinical data reuse or secondary use for research purposes is a difficult goal to achieve. Our hospital faced this problem by i2b2 (Informatics for Integrating Biology and the Bedside) [1], a data warehouse that aggregates EHR heterogenous data following the FAIR principles, making them easy to access and find by researchers but also reusable and interoperable with other systems. i2b2 system includes currently 4 million patients and 520 million observations, now available for research purposes, such as automatically filling an EDC (Electronic Data Capture) system for clinical studies [2]. The EDC software in use within our hospital is REDCap (Research Electronic Data Capture) [3], which is accessible through Application Programming Interfaces (APIs) that allow data import from external sources. The aim of this work is to evaluate a procedure for the automatic import of data from clinical practice to an EDC system for a clinical study.

2. Methods

i2b2 system within the Policlinico Hospital in Milan aggregates heterogenous data sources and the i2b2 table Patient Mapping contains the lookup between the patient identifier of each data source and the unique i2b2 patient id.

We developed an ETL (Extract, Transform, Load) procedure for importing data from i2b2 system to a REDCap study. The ETL procedure is based on JavaScript code, which runs on Mirth Server, from which HTTP POST requests containing data to be imported are sent using the REDCap API. Preliminary operation of the procedure includes the mapping of the REDCap variables into the i2b2 concepts. Then the mapping table between i2b2 and REDCap is created by joining i2b2 patient mapping table with

---

[1] Corresponding Author, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Via Francesco Sforza, 35 – 20122 Milan, Italy; E-mail: sara.pizzimenti@policlinico.mi.it.

the enrollment list containing the patient REDCap project identifier and the identity information. i2b2 queries are implemented for each data collection event in the REDCap project to extract data from i2b2 and then import into REDCap. The main query parameters are the patient identifier, the event date and the concept related to REDCap variables. Dates can be obtained directly from the value of a variable manually entered or automatically calculated starting from one or more reference dates within the project.

3. Results

The impact of automated import procedures has been evaluated on a multicenter study promoted by Policlinico Hospital and approved by Ethical Committee. The study aims to collect data from patient admitted to intensive care unit with a COVID19 infection. Our hospital has currently collected data about 279 patients. This study started in 2020 following the health emergency and patient data has been collected by hand ever since.

We decide to evaluate the import procedures by reusing structured data already stored in the hospital. In particular, we considered 31 laboratory tests collected in different events: on the first day of admission and the following hospitalization days, a total of 2482 REDCap events for the whole cohort. We compared the values entered manually with the automatically imported ones for a total number of 76942 (31 laboratory test values for each REDCap event). The comparison with a 5% of value tolerance shows that 93,5% of the manually entered exams in REDCap were found also by the automatic procedure; the remaining values could be due to manual entry errors on the exam's values or on the associated dates. Furthermore, 62,8% of the 36018 missing values were filled in by the automatic procedure. Missing values have different reasons including the onerous request for time for manual data entry. Finally, considering the total of 63530 of exams found by almost one of the two approaches, 62% come from both systems, while the 36% was entered only by automatic procedure.

4. Conclusions

Automatic laboratory tests results import led to improvements in terms of time, accuracy, and quality, showing that the automatic procedure should have applied from the beginning of data collection. For these reasons we plan to apply such a procedure on other studies and expanding the coverage of imported data types.

References

[1] Kohane IS Churchill SE Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. J Am Med Inform Assoc 2012; 19:181–5.
[2] 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.
[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.