Procedure to integrate i2b2 and REDCap: a case study at ICSM

Valentina Tibollo[1], Mauro Bucalo[2], Danila Vella[1], Morena Stuppia[2], Nicola Barbarini[2] and Riccardo Bellazzi[1,3]
[1]ICSM (Istituti Clinici Scientifici Maugeri), Pavia, Italy, [2]BIOMERIS (BIOMEdical Research Informatics Solutions), Pavia, Italy, [3]University of Pavia, Pavia, Italy
Contact: valentina.tibollo@icsmaugeri.it

Introduction

REDCap (Research Electronic Data Capture) is one of the most popular web-based applications to support data capture for research studies and registries [1]; while i2b2 (Informatics for Integrating Biology and the Bedside) is a widely adopted data warehouse to re-use clinical data for research purposes [2]. For this reason, a general procedure able to integrate these solutions could facilitate research activities in several institutions.
A complete i2b2-REDCap integration should permit to synchronize an i2b2 project with data coming from some REDCap studies enabling secondary analyses, but also to exploit i2b2 data (coming from Health Information System HIS) to create REDCap studies and facilitate data collection. SEINE (Synthesizing EDC IDR Network Exchange) is one of the most successful approach designed to this aim [3].
Starting from the principles adopted by the SEINE approach, we proposed a general and flexible ETL (Extract Transform and Load) procedure for synchronizing an i2b2 project with a REDCap study. This procedure is designed to be applied to all the types of REDCap studies and registries.

Methods

The integration is based on the fact that REDCap and i2b2 database schema have a similar Entity Attribute Value (EAV) model with a central observation table. The observations are grouped by events for REDCap and by encounters/visits for i2b2. So the basic assumption is that a REDCap event corresponds to an i2b2 visit. The main difference is that each i2b2 observation and visit needs a date. Instead the concepts/variables are stored differently: grouped by forms and studies in REDCap, whereas they are organized according to a taxonomy/ontology of terms in i2b2.
Here we present a general ETL procedure, developed using KETTLE software (Pentaho suite), to synchronize an i2b2 project with data coming from some REDCap studies (Figure1).

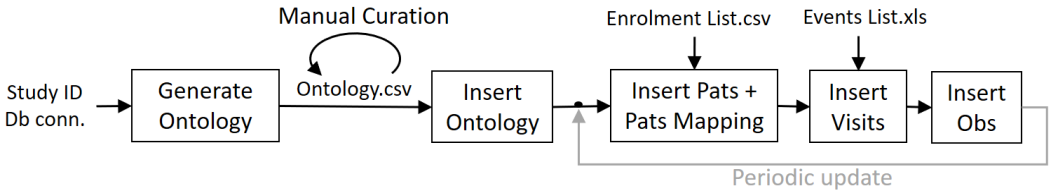

Figure 1. Data flow of the procedure to synchronize an i2b2 project with data from a REDCap study. *Pats* is an abbreviation for patients.

The initial configuration of such a procedure needs the connection parameters to databases and the ID of the study to synchronize with an i2b2 project.

- A first release of i2b2 ontology in .csv format is automatically generated as a taxonomy with the REDCap study name as main folder, then a folder for each form. A REDCap text box variable becomes a leaf of i2b2 ontology, whereas field types like multiple choice and check-box becomes structures with a folder containing leaves for the corresponding possible choices.
- Then a step of manual curation on the ontology configuration file permits to personalize some terms of the taxonomy and to specify patient information, including PHI (Protected health information) if available, and exceptions to the observation dates related to the concepts.
- A specific procedure permits to insert patients information into i2b2 exploiting the ontology file. The mapping between i2b2 and REDCap patients is based on PHI. An external enrolment list (containing REDCap ID and PHI) could be used if REDCap study is de-identified.
- The two final steps permit to insert visits/encounters and observations into i2b2 starting from REDCap events and observations. Each encounter date and observation date is assigned

according the following priorities: i) if specified in the ontology file; ii) the first date field of the form (or the forms designated for an event) if available; iii) using the log date from the log file.

Finally to make the procedure as general as possible, it is designed to manage one of the most used REDCap extension to collect concomitant therapies and adverse events as repeating fields.

Results

Istituti Clinici Scientifici Maugeri (ICSM) is a IRCCS (Institute for Research and Health Care) hospital network with 18 centers located in six different regions in Italy and is a reference point in the Italian rehabilitative medicine field. Since 2010 ICSM developed i2b2 instances with data from HIS, Biobank and registries. Currently the i2b2 framework includes an horizontal project, regarding the main chronic disease and data of over 55 thousand patients, and other vertical projects on Diabetes [4], Cardiology [5] and Oncology [6].

According to the new dispositions for pathology registries recommended by the Italian health ministry, since 2016 ICSM starts to create pathology specific registries based on the REDCap infrastructure for heart failure (HF), stroke, respiratory diseases and Parkinson syndrome.

The HF Registry has been used to test the procedure for REDCap-i2b2 integration. The REDCap registry collects demographics, hospitalization data, laboratory tests, risk factors and comorbidity, rehabilitation data and therapy prescriptions during hospitalization. The HF Registry currently includes data of 1614 patients since 2013 from 2 ICSM centers.

Figure 2 shows the i2b2 taxonomy obtained by HF Registry, focusing on the form *Contatto*. The field *Data Ricovero* (admission date) is assigned as i2b2 visit/encounter date thanks to the ontology configuration file.

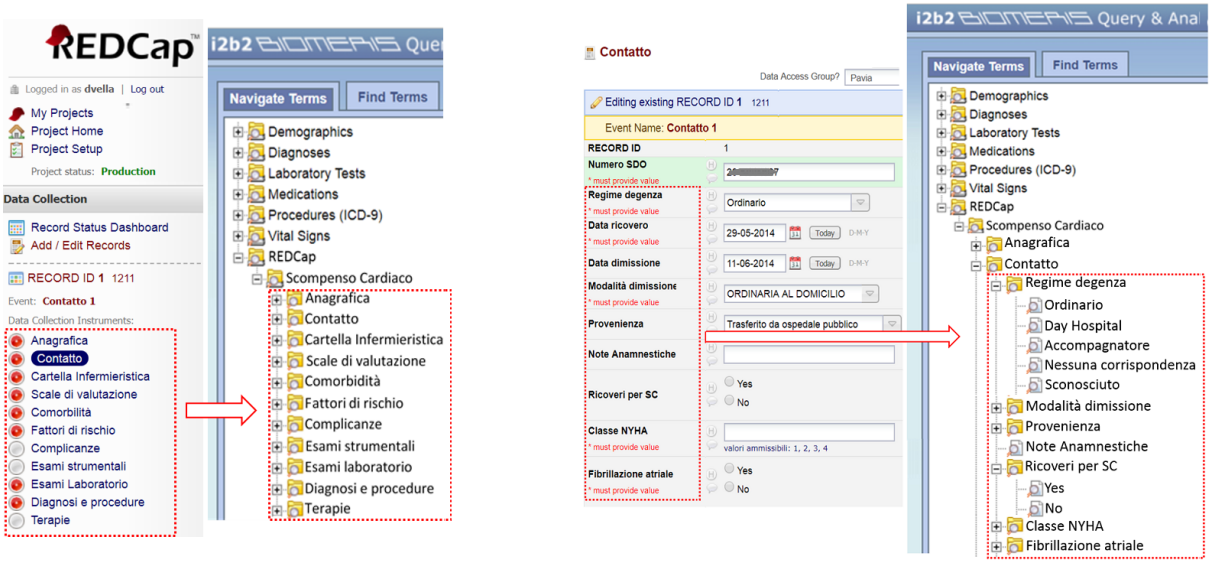

Figure 2. The HF Registry REDCap forms and the related i2b2 ontology, focusing on the form *Contatto*

Thanks to a suitable configuration of the ontology file, specific date fields have been assigned to each laboratory exam in the form *Esami Laboratorio*. Moreover the therapies collected by repeating fields in the form *Terapie* have been correctly inserted into i2b2.

The procedure is successfully applied to periodically synchronized i2b2 project with REDCap HF Registry.

References

1. P.A. Harris et al. J. Biomed. Inform. 2009; 42(2) 377–381
2. I.S. Kohane et al. J Am Med Inform Assoc. 2012;19(2)181-5
3. B. Adagarla et al. University of Kansas Medical Center. 2015; 2271(1303)
4. A. Dagliati et al. IEEE-EMBS International Conference on BHI, 2014, 244–247
5. D. Segagni et al. Studies in Health Technology and Informatics, 2012; 180, 1126–1128
6. D. Segagni et al. BMC Bioinformatics, 2012;13(Suppl 4) S5