ICT for Health Science Research
A. Shabo (Shvo) et al. (Eds.)

doi:10.3233/978-1-61499-959-1-21

An Extension of the i2b2 Data Warehouse to Support REDCap Dynamic Data Pull

Matteo GABETTA[a,*], Marco MIRABELLI[b], Catherine KLERSY[c], Valeria MUSELLA[c], Gianpiero RIZZO[c], Paolo PEDRAZZOLI[c], Nicola BARBARINI[a], Riccardo BELLAZZI[b] and Cristiana LARIZZA[b]
[a]*BIOMERIS, Pavia, Italy*
[b]*Dep. of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy*
[c]*Fondazione IRCCS Policlinico San Matteo, Pavia, Italy*

Abstract. i2b2 and REDCap are two widely adopted solutions respectively to facilitate data re-use for research purpose and to manage non-for-profit research studies. REDCap provides the design specifications to build a web service used to import data from an external source with a procedure called DDP. In this work we have developed a web service that implements these specifications in order to import data from i2b2. Our approach has been tested with a real REDCap study.

Keywords. i2b2, REDCap, Dynamic Data Pull, DDP, data integration.

1. Introduction

The i2b2 (informatics for integrating biology and the bedside) data warehouse has proven to be an effective and widely adopted solution to reuse clinical data for research purposes [1,2]. It allows to transform patient data aggregated from disparate sources among the clinical organization into a format optimized for various types and stages of research, including feasibility analysis, study design, eligibility criteria, cohort identification and recruitment and population health studies [3]. Currently, over 150 institutions worldwide use i2b2 to query patient data. An even more widespread software for supporting clinical research is REDCap (Research Electronic Data CAPture), a web-based eCRF (electronic Case Report Form) platform designed to support research studies, providing, among the many functions, also procedures for importing data from external sources [4]. REDCap is currently installed in more that 3000 clinical not-for-profit institutions [5].

The Dynamic Data Pull (DDP) is one of the procedures to import data in a REDCap study. DDP provides an adjudication process whereby REDCap users can approve all incoming data from an external source system before it is officially saved in their REDCap project. DDP assumes that a record identifier (e.g. medical record number) is stored within each REDCap record so that the it can be associated with data from an external source system. Once connected with REDCap, the external system can be fetched by the DDP both manually in real time and automatically at a regular interval.

Given the great variety of systems suitable to provide data to research studies, REDCap developers have only defined the interface to which an application must adhere in order to be used as an external data source by REDCap. In particular DDP requires

*Corresponding Author. E-mail: matteo.gabetta@biomeris.com.

the use of two web services which have to be developed on *ad-hoc* basis depending on which data source is intended to be integrated. In this work we have developed an extension of the i2b2 platform in order to make it compliant with the DDP interface.

2. State of the art

Since putting in place i2b2 in a large organization requires the integration of data from different systems (e.g. demographics, diagnoses, labs), it seemed natural to choose it as the ideal data source to be combined with REDCap through the DDP.

Another approach [6] has developed a DDP system that connects directly to the source systems via SQL queries. If this approach on the one hand is adaptable to any source with a relational database underneath, on the other hand it requires the service to be granted with direct access to all clinical sources and the comprehension of their data models. In our experience, such an effort, justified for deploying a data warehouse like i2b2, might not be acceptable only to provide a data import procedure to REDCap.

Another interesting approach is made possible by two recent updates of i2b2 and REDCap: 1) from version 8.5 of REDCap a new type of DDP has been introduced called "DDP on FHIR" which exploits HL7 Fast Healthcare Interoperability Resources (FHIR) standard to enable DDP with any compliant EHR without the need of developing an *ad hoc* service and 2) the recent work described in [7] which puts the basis to make i2b2 a FHIR compliant data source. Due, in particular, to the progress of the extension of i2b2 (still at 70% according to the authors) this integration solution, despite being extremely interesting, is not yet testable and then cannot be installed in a production environment.

3. Methods

The DDP extension of i2b2 consists of two webservices consumable by REDCap and having access to the i2b2 database. All the information about the importable fields and the rules for obtaining them from i2b2 are stored in a JSON file. The overall architecture of the developed application is represented in Figure 1.

The two web services necessary to execute the DDP are called *metadata* and *data*. The first exposes the catalog of the fields available from the source system (i.e. i2b2) and is used during the setup phase of the REDCap study. The second is the one that, given a specific patient, actually sends his/her data from i2b2 to REDCap and is called when the patient records are added/updated in the study.

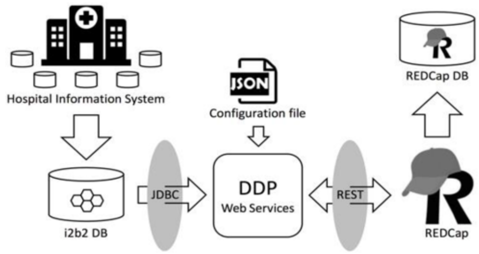

Figure 1. Overall architecture of the DDP extension for i2b2. Data coming from the Hospital Information System is first gathered in the i2b2 data warehouse. The DDP Web Services and their interfaces towards the i2b2 DB and REDCap have been developed in order to allow the use of DDP with the i2b2 data. All the information necessary to perform this operation are contained in the JSON configuration file.

The DDP setup starts with the REDCap server receiving the list of available fields. The study administrator then chooses which fields to import via DDP and associates them with specific fields of the REDCap study; for the so-called "temporal" data (i.e. those data that can be collected many times, like the labs) it is also necessary to specify which date, among those in the REDCap study, is associated with each one of them.

Both services rely on the JSON configuration file. The DDP extension can serve multiple projects from multiple REDCap installations (this is the case of institutions with a single i2b2 and more than one REDCap). For each project it is defined one “Patient Identifier” section and many “Item” sections, one for each field available from i2b2.

Each section has a sub-section called “Metadata”, which contains the data needed by the *metadata* service (e.g. name, description, category and, for an “Item”, if it is temporal or not). When the *metadata* service is called, the system aggregates the information from these sub-sections and simply returns them to REDCap.

When the *data* service is called, the system receives from REDCap an identification code; the first operation to be performed (Figure 2) is matching this code with a *patient id* in i2b2. This process relies on the information contained in the “Identification” sub-section of “Patient Identifier”. The incoming identification code is first, optionally, transformed by a JavaScript function specified in the sub-section; this is necessary when the incoming code needs to be manipulated before trying to retrieve it in i2b2. Finally the service, depending on how it is configured, tries to match the transformed code with: 1) an i2b2 patient id - this case is uncommon because the i2b2 id is typically auto-generated and is not shared with other software; 2) an identification code from the “patient_mapping” table, whose purpose is actually to store ids from external systems - this is the most common case; 3) a value stored as an observation.

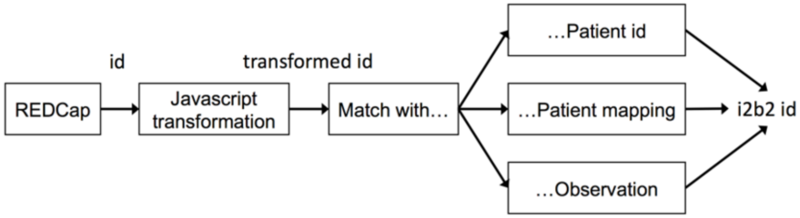

Figure 2. The id matching process.

Once the i2b2 patient id has been matched, the application retrieves, among the available items, those that are requested by REDCap and transforms them according to how it is specified in the “Data” sub-section of each “Item” in the configuration file. The process of retrieving data in i2b2 and manipulating it for REDCap is represented in Figure 3.

We have foreseen two data sources in i2b2: the table called “patient_dimension”, which contains the demographics data, and the “observation_fact” table, the central table of the data warehouse storing all the facts associated to the individuals (e.g. diagnoses, procedures, labs, medications). The retrieval process is different depending on which source is associated with the specific item.

When the item is extracted from the “patient_dimension”, the first operation performed is to select one of the demographics values associated to the patient (e.g. date of birth, date of death, sex, race) according to the configurations; this value can then be transformed to be compliant with its destination in the REDCap study. It is possible to transform the value with a JavaScript function or with a simple lookup table that specifies, given the value extracted from the “patient_dimension”, the value to send to REDCap.

If the item is based on the "observation_fact" table, the extraction process is more complex. All the rows belonging to a specific concept for the patient (e.g. "Cardiovascular diagnoses" or "Measures of Hematocrit") are first considered. The next operation aims at reducing this set of rows to the ones that contain the output information. For temporal items, only observations within the requested time range are maintained. For non-temporal items, a value (i.e. column) for each row is considered; the value can then be transformed with a lookup table or a JavaScript function and finally a selection logic is applied to these values. Available selection logics are: greatest and lowest value, oldest or most recent date and "no selection", to pass all rows to the last phase; if no logic selection is applied, a set of rows will be passed, otherwise a single row will be considered.

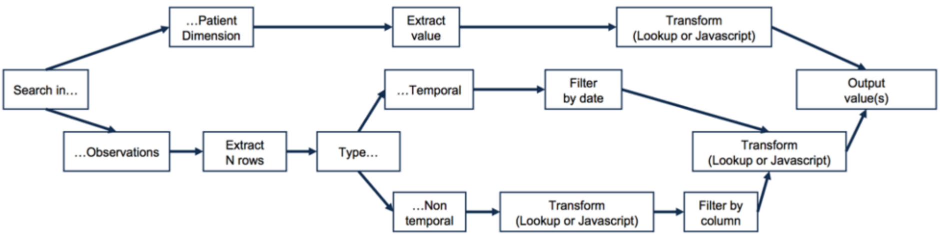

Figure 3. The data retrieve process

In the last phase it is possible to choose if one or more values will be returned to REDCap. The result set is obtained by considering a value (i.e. column of the observation_fact, not necessarily the same as the previous phase) for each row; this set of values is transformed (with a lookup or a JavaScript function, mandatory if two or more values have to produce a single output) and sent to REDCap.

4. Results

Using the approach described above we have deployed the DDP extension of i2b2 at the *Fondazione IRCCS Policlinico San Matteo* in Pavia (Italy), a Research General Hospital with more than 1000 accredited beds and a catchment area of 700 thousand inhabitants. Since 2015 *San Matteo* hospital has a REDCap installation with more than 20 active studies and more than 13000 records managed by 326 distinct users. More recently data coming from disparate clinical applications within the hospital (e.g. ADT system, laboratory tests), belonging to both inpatients and outpatients, are loaded periodically into an i2b2 data warehouse. In 2016 in this hospital: 36500 admissions and 99000 accesses to the emergency unit have been performed and more than 2.1 million outpatients have been treated.

The first REDCap study on which the technology described in the present work was applied is the "Cardioncologia" patient registry, active since 2015 for monitoring cardiotoxicity in patients treated with adjuvant anthracycline-based chemotherapy for early breast cancer; the registry counts 51 active records with 15 forms per patient. Several data that characterize the patients can be imported via DDP from the i2b2 data warehouse; below we will describe some of them.

The identification matching process is performed as follows: when a new record is created, REDCap assigns a unique identification code to it; this id, together with some identifying data, is manually stored by the data manager in the enrollment list. This list, which does not reside within REDCap database since it contains only pseudonymized data, is one of the sources that are periodically synchronized with i2b2: thus, the REDCap identification code is associated with the one of i2b2 in the “patient_mapping” table. Among the fields imported via DDP in the “Cardioncologia” registry we can find: (i) Patient’s gender, which comes from the demographic system of the hospital and is stored in i2b2, therefore it is importable via DDP; it was necessary to transform the values of the gender from those in the i2b2 “patient_dimension” table ("M" or "F") into those valid for the REDCap study ("0" or "1"); this operation was performed through a lookup table. (ii) The “Medical History” section: here it is possible to specify, for several conditions/diseases, whether the patient was affected or not. Since ICD9-CM diagnoses are among the data available in i2b2, we have used them to fill this section. It was also necessary to transform the presence/absence of certain diagnoses into a true/false value through a JavaScript transformation. (iii) Demographics values like age at enrollment. (iv) Values and units of measure of laboratory exams. (v) The TNM Classification of Malignant Tumors. (vi) Procedures like surgeries.

5. Conclusion

In this work we have implemented a web service to allow REDCap to import data from i2b2 through the DDP. The choice of i2b2 seemed ideal because, wherever this system is installed, the data coming from disparate hospital systems are periodically gathered in a single database after quality controls and appropriate transformations. Our approach has been successfully tested in a research hospital where both i2b2 and REDCap are used.

As possible future development we will test the system more thoroughly and with new studies. We will also consider the opportunity to import data from other i2b2 services, such as the Identity Management Cell, which is responsible for containing PID.

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] Prokosch HU, Ries M, Beyer A, Schwenk M, Seggewies C, et al. IT Infrastructure Components to Support Clinical Care and Translational Research Projects in a Comprehensive Cancer Center. Stud Health Technol Inform 2011;169: 892–896.
[3] Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2) J Am Med Inform Assoc. 2010;17(2):124–130.
[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 Apr;42(2):377–381.
[5] https://projectredcap.org (accessed October 23, 2018).
[6] Campion TR, Sholle ET, Davila MA. Generalizable Middleware to Support Use of REDCap Dynamic Data Pull for Integrating Clinical and Research Data. AMIA Summits on Translational Science Proceedings. 2017;2017:76-81.
[7] Paris N, Mendis M, Daniel C, Murphy S, Tannier X, Zweigenbaum P. i2b2 implemented over SMART-on-FHIR. AMIA Summits on Translational Science Proceedings. 2018;2017:369-378.