

# Practical Implementation of a Bridge between Legacy EHR System and a Clinical Research Environment

Dina VISHNYAKOVA<sup>a,1</sup>, Sylviane BOTTONE<sup>b</sup>, Emilie PASCHE<sup>a</sup> and Christian LOVIS<sup>a</sup>

<sup>a</sup>*Division of Medical Information Sciences, University Hospitals and University of Geneva, Switzerland*

<sup>b</sup>*Division of Medical Information Sciences, University Hospitals of Geneva, Switzerland*

**Abstract.** Employing the bridge between Clinical Information System (CIS) and Clinical Research Environment (CRE) can provide functionality, which is not easily, implemented by traditional legacy EHR system. In this paper, the experience of such implementation at the University Hospitals of Geneva is described. General overview of the mapping of extracted from CIS data to the i2b2 Clinical Data Warehouse is provided. The defined implementation manages to provide the interoperability for the CRE.

**Keywords.** Data Integration, Clinical Information System, Electronic Health Records, Big Data, Clinical Data Warehouse.

## Introduction

The global data double their volume every two years. Furthermore, the increasing of the healthcare data digitization means that healthcare organizations, often, add terabytes of patient records to data centres annually. At present, most of this unstructured data is not used, partly, due to the legacy systems (LS) and mainly they are in sight of regulatory purposes.

A better understanding of data such as to know more about an individual patient or large groups of them and how to use this information provides better, more efficient and cheaper healthcare. Additionally, arose the challenge to reuse clinical information for real-time care or clinical research [1-3]. Furthermore, the process of health information exchange, necessary for getting the patient data is neither easy nor simple. For example, one electronic health record (EHR) system uses "hypertension" and another one may call it "elevated blood pressure". To solve this problem, the standards such as ICD-9/10, SNOMED CT, LOINC and many more are used [4]. The advantage of the standardization is to facilitate the research query process.

The analysis of big data is of high priority but the processing of such data is a major challenge. Here, one of the major problems is a data lookup in a legacy EHR system. Physicians and research staff need that patient information is accessible

---

<sup>1</sup> Corresponding Author.

regardless its way of keeping records in current EHR system. Mostly, the data are left within the legacy system, requiring users to login to another system in order to access all relevant patient information. This process is very expensive and time-consuming.

Here, we present the bridge methodology, between legacy EHR and a CRE used at the University Hospitals of Geneva (Hôpitaux Universitaires de Genève, HUG) that closely conjoins with the hospital clinical information systems.

## **1. Data and Methods**

The association of 8 hospitals of Geneva and 40 ambulatories spread throughout the territory of the canton merged into the first university hospital in Switzerland, the HUG.

HUG is comprised of more than 2,000 beds. According to annual statistics, approximately 60,000 admissions and over 900,000 outpatient visits took place. There were over 6,000,000 laboratory testing done in 2013. The total number of full-time employees is about 9,000, where the number of physicians is about 1.300.

### *1.1. Design objectives*

The design of the bridge methodology between legacy EHR and a CRE was focused around several goals. The primary goal is to provide a secure and clear representation of patient information for research purposes. Moreover, in healthcare domain, the cross-border integration of data is a challenge. Due to the ethical reasons, data providers are not allowed to store patient data outside of their hospital infrastructure (Intranet).

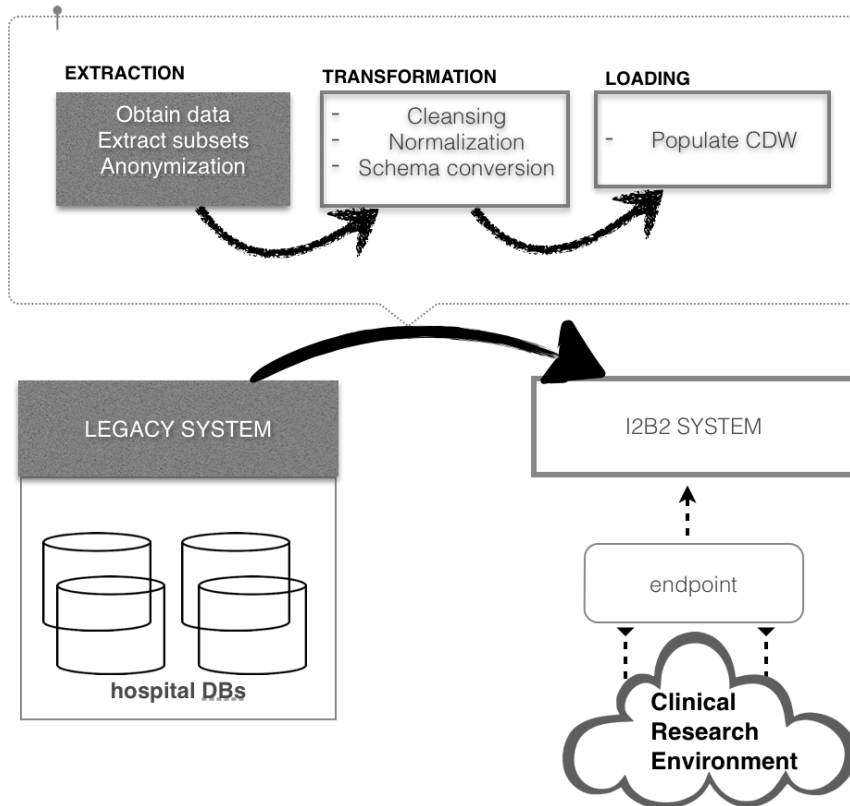
Furthermore, due to a high-level protection to preserve system stability the real time analysis/visualisation of data cannot be applied. Such data representation needed to be based on a software framework that is secure and could be easily extended.

### *1.2. Implementation tools*

Respecting all constrains, described above, it was decided to use Open Source software and the i2b2 (Informatics for Integrating Biology & the Bedside) framework. The design of i2b2 is based on web services that provide securable remote access to its various parts [5]. Moreover, researchers can use i2b2 to perform research queries of de-identified data.

Clinical Information System (CIS) of HUG include different database management systems and access protocols. In order to provide a homogenous access to the information, the i2b2 server placed between CIS and the i2b2 client or a so-called query endpoint, see Fig. 1. The endpoint is responsible for the resulting information representation extracted from the i2b2 server. The i2b2 server is periodically fulfilled by pre-constructed ETL processes in the following order: 1) retrieve the content of CIS 2) perform model transformation 3) load data into the i2b2 CDW.

The first principle step at ETL stage is to pseudonymise sensitive patients' data [8] such as patient ID number, Episodes of care, date of birth. Despite the ability of i2b2 to restrict/grant access to identified data within its security policies, a pseudonymisation step took place during the extraction phase. In all the circumstances, by doing so, the restitution of patient identifiers is reversible only by means of authorities that are responsible for the data extraction phase. In our case, a hash encryption function is



**Figure 1.** Information Workflow. The Intranet components are shown in colour.

used at the Extraction phase. The mapping table with keys to identified data never leaves the Intranet location of HUG. Such data as international standardization codes are available without constraints.

### 1.3. Data Mapping

The i2b2 schema is represented as a five-axis star: PATIENT, VISIT, OBSERVATION, CONCEPT and PROVIDER.

The HUG CIS data corresponding to PATIENT, VISIT and PROVIDER can be easily integrated directly to the i2b2 schema while to fulfil CONCEPT and OBSERVATION is a challenging task.

The mapping of CONCEPT dimension, which consists of lab tests, medications and etc., is challenging. HUG data are not all standardized in accordance with the international standardization codes. Moreover, i2b2 keeps a tree structure of concepts in its schema it is necessary to return data records containing the name of concept, the id for the concept and ids of concept's ancestors. Nevertheless, it is not exceptional to find a free form text with the optional standardisation codes.

In order to avoid missing important concepts, it was necessary to add additional values to standard dictionaries. This required a separate concept table that stores new repository codes for data that have no standard dictionary code. In some cases of free text mapping, specific algorithms such as described in [4,6,7] based on biomedical terminologies (SNOMED CT, NEWT and etc.) were implemented to perform term normalization of data retrieved from CIS.

Most data are stored in the OBSERVATION table. In this table every row is an observation on a patient (Diagnosis, Procedures, Laboratory data and etc.). Each row correlates with one observation about a patient made during a visit. There are several requirements to the patient observation such as: to be related to a patient, to contain temporal information (e.g. start and end date), to be related to concepts defined in CONCEPT dimension.

## 2. Discussion and Conclusion

To date, there is no optimal solution for the legacy system replacement. In many research cases the incompatibility of data in such systems becomes a time-consuming challenge for the research communities. On other hand the simplicity of data representation by such system as i2b2 opens the horizons for the effective clinical data management. By building the bridge between the LS and research environment the achievements in tasks such as clinical investigations/studies, quality management and etc. become feasible. This type of bridging omits the need in global data mapping strategies, costly extraction of relevant clinical data from source systems and etc.

Currently, this way of bridging is involved in the framework of one of the largest European public/private partnership projects in the eHealth area - EHR4CR<sup>2</sup> (Electronic Health Records for Clinical Research). This project aims to provide adaptable, scalable and interoperable solutions for reusing data from EHR systems for the clinical research. The CRE of EHR4CR avoids a global data model representation, which makes possible to integrate different sources of data and to ensure the scalability. The i2b2 CDW settled by HUG is only required to have the corresponding endpoint and the corresponding mapping of added sources to EHR4CR data codes.

By bridging data the quality of information can also be improved as the data from different EHR systems is integrated through a standardized process of data extraction, normalization, transformation, and loading (ETL). The resulting consistency of data may increase research confidence as the bridging can eliminate the conflicting results that are often reported due to the different data sources or different ETL processes.

We will continue to deploy services to update the HUG CDW continuously and to evaluate its usefulness, advantages and disadvantages.

## Acknowledgements

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement n° [115489], resources of

---

<sup>2</sup> The IMI EHR4CR Project runs over 5 years (2011-2015). This project involves 33 partners (academic and industrial). More information at <http://www.ehr4cr.eu>

which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013) and EFPIA companies' in kind contribution.

## References

- [1] H.U. Prokosch, T. Ganslandt, Perspectives for medical informatics. Reusing the electronic medical record for clinical research, *Methods Inf Med.* **48**(2009), 38–44.
- [2] C.Ohmann, W. Kuchinke, Future developments of medical informatics from the viewpoint of networked clinical research. Interoperability and Integration, *Methods Inf. Med.* **48** (2009), 45-54
- [3] P. Knaup, O. Bott, C. Kohl, C. Lovis, S. Garde, Electronic patient records: moving from islands and bridges towards electronic health records for continuity of care, *Yearb Med Inform*, 2007, 34–46.
- [4] P. Ruch, J. Gobeill, C. Lovis, A. Geissbühler, Automatic medical encoding with SNOMED categories, *BMC medical informatics and decision making*, **8** (2008), S6.
- [5] S.N. Murphy, M. Mendis, K.Hackett, R. Kuttan, W. Pan, L.C. Phillips, H.C. Chueh, Architecture of the open-source clinical research chart from Informatics for Integrating Biology and the Bedside, *In AMIA Annual Symposium Proceedings (2007). American Medical Informatics Association.*, 548
- [6] D. Vishnyakova, E. Pasche, C. Lovis, P. Ruch Utilization of ontology look-up services in information retrieval for biomedical literature, *Studies in health technology and informatics* **186** (2012), 155-159.
- [7] Y. Vishnyakov, R. Vishnyakov Semantic proximity in information retrieval and documents classification In Computational Intelligence and Informatics, *IEEE 14th International Symposium on* (2013), 131-134.
- [8] P. Ruch, R. H. Baud, A. M. Rassinoux, P. Bouillon, G. Robert, Medical document anonymization with a semantic lexicon, *In Proceedings of the AMIA Symposium, American Medical Informatics Association* (2000), 729.