Enabling Clinical Data Reuse with openEHR Data Warehouse Environments

Luis Marco-Ruiz, Pablo Pazos Gutiérrez, Koray Atalag, Johan Gustav Bellika, Kassaye Yitbarek Yigzaw

Abstract

Modern medicine needs methods to enable access to data, captured during health care, for research, surveillance, decision support and other reuse purposes. Initiatives like the National Patient Centered Clinical Research Network in the US and the Electronic Health Records for Clinical Research in the EU are facilitating the reuse of Electronic Health Record (EHR) data for clinical research. One of the barriers for data reuse is the integration and interoperability of different Healthcare Information Systems (HIS). The reason is the differences among the HIS information and terminology models. The use of EHR standards like openEHR can alleviate these barriers providing a standard, unambiguous, semantically enriched representation of clinical data to enable semantic interoperability and data integration. Few works have been published describing how to drive proprietary data stored in EHRs into standard openEHR repositories. This tutorial provides an overview of the key concepts, tools and techniques necessary to implement an openEHR-based Data Warehouse environment to reuse clinical data. We aim to provide insights into data extraction from proprietary sources, transformation into openEHR compliant instances to populate a standard repository and enable access to it using standard query languages and services.

Keywords:
Data Reuse; openEHR; Data Warehousing; Electronic Health Records; Clinical Information Systems.

Tutorial Description

This tutorial will cover the stages involved in the design and implementation of an openEHR-based DW environment for clinical data reuse. First, we will present the challenges involved in modeling clinical information structures for data reuse. Specifically, we will provide some orientation in the design of archetypes intended to represent the information structure of the generated openEHR instances. Second, we will continue presenting the best practices for data extraction to create an integrated canonical view of the data before standardizing it. Third, we will provide insights into data transformation from proprietary canonical views into openEHR archetype compliant instances. We will cover the transformation and aggregation functions needed to carry out the standardization process. Fourth, we will explain the openEHR repository load and show several use cases using standard and nonstandard query languages to create the data sets needed for different reuse scenarios. We will explore the advantages and drawbacks of each query language. Finally, we will explore the limitations of current approaches and the ongoing developments to overcome them.

Tutorial Speakers

Luis Marco-Ruiz, M.Sc.; Norwegian Centre for Integrated Care and Telemedicine, University Hospital of North Norway; Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway; Qualified member of the openEHR Foundation; Tromsø, Norway

Pablo Pazos Gutiérrez, Ingeniero en Computación, UdelaR, Montevideo, Uruguay; Clinical Informatics Consultant at CaboLabs Medical Informatics and Standards; Lecturer at Asociación Chilena de Informática en Salud (ACHISA); Coordinator at the openEHR-ES community; Qualified member of the openEHR Foundation - Localization Programme Member for Latin America

Koray Atalag, MD, PhD, FACHI: Senior Research Fellow in Biomedical Informatics at Auckland Bioengineering Institute and National Institute for Health Innovation at the University of Auckland, New Zealand; openEHR Localisation Programme Leader

Johan Gustav Bellika, PhD; Chief Researcher at the Norwegian Center for Integrated Care and Telemedicine, University Hospital of North Norway; Professor at the Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway; Tromsø, Norway

Kassaye Yitbarek Yigzaw, M.Sc.; Department of Computer Science Faculty of Science and Technology UiT The Arctic University of Norway; Tromsø, Norway

General Topics

The tutorial general topics will include:

- Data Modeling for clinical data reuse
- Data Extraction from proprietary sources
- Privacy-preserving distributed clinical data reuse
Specific Educational Goals

The main goal is to provide insights into the DW development for clinical data reuse with openEHR. We will present the challenges in data reuse that arise from the heterogeneity of the proprietary data formats used to store clinical data in Health Information Systems. Specifically, the tutorial will cover how to prepare data for reuse through its extraction from proprietary data repositories, transformation into openEHR compliant standards, load into an openEHR repository and query via openEHR query languages.

Expected Outcomes

We expect the attendee to have a general view of: (a) the stages involved in data preparation for reuse with openEHR; (b) available technologies to implement each of the stages; (c) limitations of current approaches and open research fields of clinical data reuse.

Expected Attendees

Software Developers, Software Architects, Database Managers, Consultants or Decision Makers involved in projects related to clinical data integration, normalization or reuse.

Tutorial Structure and Arguments

1. Data Modeling for Reuse

The tutorial will cover the edition of low-level abstract archetype-based data views used to represent openEHR data instances to feed the DW. The design of those archetypes, as opposed to archetypes used to represent clinical concepts, focuses in allowing an easy to map fine grained data structure. The aim of this stage is to generate minimally aggregated archetypes flexible enough to allow the definition of a wide range of queries for different scenarios.

2. Data Extraction from Proprietary Sources

Patient’s health records are often distributed across multiple institutions where they received care. Patient’s privacy concerns are key factors inhibiting access to health data. The tutorial will review different privacy-preserving techniques for distributed health data access and the operations involved in data extraction (integration, data cleansing, record linkage and terminological annotation). We will present how using those techniques a canonical view that integrates data from heterogeneous data sources can be produced.

3. Data Transformation Into openEHR-compliant Instances

We will present the transformation and aggregation functions to be applied over the canonical view previously defined to transform from canonical data instances into openEHR-compliant archetype instances. We will review available tools for data transformation into OpenEHR instances. We will present experiences in data normalization with both commercial [1] and open-source tools [2]. Besides, the tutorial will cover how to leverage the data transformation systems with Representational State Transfer (REST) web services APIs to enable on demand access to openEHR standardization services.

4. Data Load of openEHR Repositories

The tutorial will present the challenges involved in loading openEHR data repositories with the instances created in the transformation stage. We will present some of the available APIs for data submission to some of the available repositories and the integration among the transformation stage and the repository using REST architectures.

5. Data Query and Reuse

After exposing the steps needed to populate the openEHR repository. We will cover how to query data for reuse through RESTful APIs and the Archetype Query Language. The tutorial will briefly comment other alternatives for data query like triplets-based systems (SPARQL) [3].

6. Limitations of Current Approaches and Ongoing Developments to Overcome Them

Finally, the tutorial will review the current challenges presented in clinical data reuse using openEHR, the ongoing works to overcome them and open research fields in the area of data reuse and openEHR DW.

References

