Chapter 3

LinkEHR: A Platform for the Normalization of Legacy Clinical Data Based on Archetypes

José Alberto Maldonado  
_Instituto ITACA, Universitat Politècnica de València, Spain_

Diego Boscá  
_Instituto ITACA, Universitat Politècnica de València, Spain_

David Moner  
_Instituto ITACA, Universitat Politècnica de València, Spain_

Montserrat Robles  
_Instituto ITACA, Universitat Politècnica de València, Spain_

**ABSTRACT**

Normalization of data is a prerequisite to achieve semantic interoperability in any domain. This is even more important in the healthcare sector due to the special sensitivity of medical data: data exchange must be done in a meaningful way, avoiding any possibility of misunderstanding or misinterpretation. In this chapter, we present the LinkEHR system for clinical data standardization and exchange. The LinkEHR platform provides tools that simplify meaningful sharing of electronic health records between different systems and organizations. Key contributions of LinkEHR are the development of a powerful medical concept, expressed in the form of archetypes, editing framework based on formal semantics capable of handling multiple electronic health record architectures, the definition of high-level non-procedural mappings to describe the relationship between archetype and legacy clinical data and the semi-automatic generation of XQuery scripts that transform legacy data into XML documents compliant with the underlying electronic health record data architecture and at the same time satisfy the constraints imposed by the archetype.

DOI: 10.4018/978-1-4666-3000-0.ch003
1. INTRODUCTION

Nowadays, patient health data is scattered among many information systems. This leads to distributed and heterogeneous data resources creating a large gap between the potential and actual value of the information content of Electronic Health Record (EHR) systems. Closing this gap by making efficient use of the health data held by these systems, could improve significantly patient care, clinical efficiency, patient safety and empower research activities.

To achieve a complete EHR that encompasses all existing health information about a patient independently of where it was generated or where it is stored and maintained crucially depends on a high level of semantic interoperability among EHR systems. Semantic interoperability is the ability of computer-based systems to seamlessly communicate, incorporate and use information generated by external systems. Data standardization (normalization) is a prerequisite to achieve semantic interoperability in any domain. The use of standards enables clear, complete, and efficient communication of data by means of agreements on language and procedures. This is even more important in the healthcare sector due to the special sensitivity of medical data. Data exchange must be done in a meaningful way, avoiding all possibility of misunderstanding or misinterpretation. As a consequence any project involving the sharing or communication of EHR must face the challenge of how to model faithfully the data to be communicated. The first and basic requirement is the definition of a common information architecture for communicating EHR extracts. The good news are that considerable effort has been invested over the years on the definition of EHR architectures (Kalra, 2006) and currently a mature body of EHR standards is available.

In this chapter we describe the LinkEHR platform along with some use cases in real settings. LinkEHR is a software platform whose main purpose is the creation a virtual, standardized and federated EHR view from data distributed among heterogeneous and autonomous health information systems. LinkEHR uses archetypes as a means to achieve semantic enrichment and normalization of existing health data. The main objectives of LinkEHR are then twofold. Firstly, in the context of data integration, the use of archetypes as a semantic layer over the data repositories, whose contents need to be integrated, described and exchanged, associating them with formal semantics. Secondly, the use of archetypes for making public existing clinical information in the form of XML EHR extracts compliant with the underlying EHR architecture.

2. DUAL MODEL EHR ARCHITECTURES

During the last decade different organizations have been working on the definition of an EHR architecture for the faithful communication of health data. The Technical Committee 251 (health informatics) of the European Committee for Standardization (CEN/TC251) completed in 2007 a European Standard for the communication of the EHR called EN13606 whose part 1 (reference model) (International Organization for Standardization, 2008a) became an ISO standard in February 2008. One of the main contributions of CEN/ISO 13606 is the utilization of an EHR architecture based on the dual model methodology (Beale, 2002) for the description of the structure and